Comparative Analysis of Collocations and Semantic Preference of "Risk" in Risk Management Reports from Selected Malaysian and American Banks

Khairul Firhan Yusob[1,2], Afida Mohamad Ali[1*], Mei Yuit Chan[1], Geok Imm Lee[1]

[1]Faculty of Modern Languages and Communication, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

[2]Akademi Pengajian Bahasa, Universiti Teknologi MARA Pahang Kampus Jengka, 26400 Bandar Jengka, Pahang, Malaysia

***Corresponsing Author**

DOI: https://doi.org/10.47772/IJRISS.2026.10100510

Received: 28 January 2026; Accepted: 02 February 2026; Published: 15 February 2026

ABSTRACT

Banks rely on risk management reports (RMRs) to detect, analyse, and manage potential threats to their operations. These reports provide insights into a bank's exposure to risk and support informed decision-making to minimise losses. Previous corpus studies have shown that the word risk is often perceived negatively. This study investigates how risk is represented in RMRs, which are intended to assure stakeholders of banks' ability to manage risks effectively. A self-compiled corpus comprising RMRs from Malaysian and American banks was utilised in this study. Collocational patterns were examined using the Graph Coll feature in Lancs box 6.0, and collocate strength was measured through Mutual Information (MI) scores. Semantic preference was further analysed using UCREL Semantic Analysis System (USAS). The findings reveal that although the collocates of risk in both corpora differ in frequency and MI scores across the two corpora, ency the word is consistenly paired with domain-specific terms denoting risk types. Additionally, risk in Malaysian RMRs exhibits a broader range of semantic associations than in American reports. The study enhances understanding of how the concept of risk is linguistically represented across banking contexts and provides insights into the persuasive construction of expertise in financial reporting.

Keywords: American banks, collocations, Malaysian banks, risk management reports, semantic preference

INTRODUCTION

A risk management report (hereafter, RMR) is a document that provides an overview of an organisation's risk management activities. It plays a crucial role in identifying, assessing and managing risks in organisations. These reports offer a comprehensive analysis of different types of risks, such as credit risk, market risk, operational risk and liquidity risk (Al-Tamimi & Al-Mazrooei, 2007). They also help businesses make informed decisions and take appropriate actions to mitigate risks and protect their financial stability. In the business world, financial institutions, especially banks, rely heavily on RMRs to support effective decision-making and demonstrate sound risk governance.

One of the key reasons RMRs are important for banks is that they help in complying with regulatory requirements (Gurendrawati et al., 2021). As banks are subject to various regulations and guidelines that require the implementation of effective risk management practises, RMRs provide evidence that the bank is complying with these regulations and that adequate measures are in place. In addition, RMRs facilitate transparency and accountability. Banks also have a responsibility to their stakeholders, including shareholders, investors and regulators to provide accurate and timely information about their risk exposure and management strategies

Through RMRs, banks can communicate this information effectively, enhance transparency and build trust with stakeholders (Muigai & Maina, 2018).

Despite their importance, the RMRs have several limitations. The Association of Chartered Certified Accountants (ACCA, 2014), reported that RMRs are often too general, overly positive, too bland, too and excessively lengthy, which undermines their informativeness. One possible reason for this is the use of the word risk, which often carries negative connotations. The lexical item risk tends to create an unfavourable impression among readers, and reporting on it may be perceived as revealing weaknesses or potential problems, which can be counterproductive (Mazumder & Hossain, 2018). As a result, banks may frame RMRs in ways that sound more positive or reassuring. The negative perception of risk is also reflected in other studies. For example, Hardy and Colobini (2011) found that the use of 'risk' is typically used with negative meanings in medical discourse, while Müller and Mell (2021), in a diachronic study of German parliamentary debates noted that that risk remained largely negative despite evolving across domains such as health, economics, and technology.

While negativity surrounding the word risk has been observed in various contexts, limited research has explored how it is represented in RMRs. Furthermore, cross-national comparisons of RMRs remain scarce, particularly between Malaysian and American banking contexts. Addressing this gap, the present study investigates how risk is linguistically represented in RMRs by analysing its collocational patterns and semantic preferences across the two contexts.The data used in this study are RMRs from Malaysian and American banks. The corpus of RMRs from Malaysian banks is referred to as CORDMAB (Corpus of Risk Disclosure by Malaysian Banks), while the corpus consisting of reports from American banks is referred to as CORDAB (Corpus of Risk Disclosure by American Banks). Further details on corpus compilation and procedures are provided in the Materials and Methods section. In order to shed light on the issue of the use of risk in RMRs, this study seeks to address the following research questions:

i. What are the collocational patterns of risk in CORDMAB and CORDAB?

ii. How do these patterns compare across the two corpora?

iii. How is risk represented in terms of semantic preference in each corpus?

LITERATURE REVIEW

Collocation

Corpus linguistics is an approach to the study of language that focuses on the systematic investigation of large collections of texts in order to uncover patterns of language use that might otherwise be inaccessible through intuition or small scale investigation (Römer, 2011). Since the advent of corpus linguistics in the late twentieth century, it has been at the core of the study of language in applied linguistics, discourse analysis, language teaching, and computational linguistics due to its potential for delivering empirical support for linguistic theories (Baker et al., 2007).

In the study of language, collocation analysis is a core concept in the study of language. It is a concept in language that refers to the phenomenon in which words tend to co-occur within a certain range more frequently than what is statistically expected. The concept of collocation was originally coined by Firth (1957) and later developed by Sinclair (1991), who showed that the study of collocation can reveal syntactic, semantic, and pragmatic tendencies. Collocation usually takes the form of adjective-noun, verb-noun, or preposition-noun phrases. It has been identified through the application of statistical methods such as mutual information or log likelihood (Sinclair & Carter, 1994).

In the study of language, it has been demonstrated in corpus linguistics that the study of collocation is more than just a description of language; it has the potential for interpreting the construction of meaning, stance, and social realities. Collocation is therefore a statistical phenomenon in language in which words tend to co-occur, and the interpretation of the meaning of the words is achieved through the study of concordance and the qualitative investigation of the words (Erdentug & Musayeva Vefalı, 2018; Spicksley & Franklin, 2023). It has been

demonstrated in the study of language that meaning is not located in individual words; instead, meaning is located in groups of words. Semantic preference is therefore the tendency of a word to prefer the company of semantically related groups of words, while semantic or discourse prosody is the positive or negative value that emerges through the repeated patterns of collocation (Bednarek, 2008; Hunston, 2007; Partington, 2004; Stubbs, 1995).

The initial research in the field of semantic prosody is indicative of the potential of collocation analysis in unearthing underlying evaluations that may not always be obvious to the casual reader. For instance, in the initial research in the field of semantic prosody, Louw (1993) points out the diagnostic potential of semantic prosody in the identification of pragmatic effects such as irony. In subsequent research, Stubbs (1995) points out the potential of collocational patterns in systematically orienting words towards either positive or negative meaning. The research in the field of semantic prosody conducted by Partington (2004) and Hunston (2007) is indicative of the potential of collocation in unearthing underlying evaluations.

The empirical research in the field of public and institutional discourse points to the analytical potential of collocation in identifying underlying evaluations. In the empirical research conducted by Baker et al. (2007), collocation is used to identify its contribution to the portrayal of minorities in the UK press. In the research conducted by Gabrielatos and Baker (2008), collocation is used to identify underlying phraseologies such as movement, threat, and burden, used to represent refugees and asylum seekers.

Collocation has also been used in the analysis of identity construction. Abdullah et al. (2010) illustrate this by showing how collocation is used in representations of Malay identity in Hikayat Abdullah. Elewa (2025) illustrates the use of collocation in gendered representations of religion. Both studies show how collocation analysis can be used in revealing connections between micro-level lexis and macro-level sociocultural representations of identity.

In more specialized institutional settings, collocation has been used in the analysis of issues of legitimacy and authority. Lukin and García Marrugo (2024) show how collocations of war in legal discourse contribute to prosodies of normalization and legitimation of warfare. Fuoli and Beelitz (2025) use collocation analysis in their analysis of the term net zero in corporate discourse and show how collocational context signals either reputational positioning or commitment. These studies show that in institutional settings, collocation not only reflects institutional assumptions but also works to reinforce them.

Finally, diachronic and event-sensitive studies show that collocation is responsive to social change. Spicksley and Franklin (2023) show how collocation in representations of frontline shifted during the COVID-19 pandemic. Palacios Martínez (2010) illustrates that the non-standard ain't shows collocation patterns associated with pragmatic effects. These studies show that collocation analysis can be used in revealing social meaning in non-standard varieties of English.

Collocation has also been shown to be useful in semantic differentiation and near-synonymy studies. Phoocharoensil (2020) demonstrates that words like consequence, result, and outcome can be differentiated from one another in terms of collocation, while Prihantoro (2022) demonstrates that the word energy shows diachronic semantic shift in its collocates. Finally, Kazmaly (2024) demonstrates that collocation, and word sketch in particular, can provide a window into culturally shared concepts, like the concept of personality in words like shy and outgoing.

The above studies, therefore, show the methodological robustness of collocation analysis in linking frequency and interpretation, as proposed by semantic preference, prosody, and contextualisation in concordance (Bednarek, 2008; Hunston, 2007; Louw, 1993; Partington, 2004; Stubbs, 1995). Despite its obvious potential in the analysis of stance, legitimacy, and semantic shift in institutional and corporate communication (Fuoli & Beelitz, 2025; Lukin & García Marrugo, 2024; Spicksley & Franklin, 2023), collocation has been less commonly used in the analysis of risk accountability in audited corporate communication, which forms the rationale for the current study, which seeks to utilize collocation and lexical patterning in the analysis of RMRs.

Semantic Preferences

In corpus linguistics, semantic preference refers to the tendency for a word to be found with other words from a particular semantic set, and so shed light on some aspect of its semantic potential (Sinclair, 1991; Stubbs, 2001). Semantic preference has been closely associated with semantic prosody, which refers to the evaluative connotations which tend to build up in the collocational company of a word, and discourse prosody, which extends semantic prosody into larger discourse chunks (Louw, 1993; Partington, 2004; Sinclair, 1991). This approach has challenged the idea that meaning is contained in individual words and has emphasized the importance of phraseological units in meaning construction. One of the most important findings in these studies is that semantic preference has been found to correlate with grammatical rather than topical variation. In her study, Flowerdew (2012) has found a strong link between semantic preference and syntactic environment, while Partington (2004) has found a link between a change in complementation and a change in semantic sets and contextual meaning. These studies reinforce the idea that semantic preference is a linguistic rather than a topical phenomenon. In addition, studies on semantic preference across different genres have further supported its explanatory potential. Begagić (2013) has found a semantic preference for make sense, which often occurs with semantic sets related to difficulty, e.g., try, attempt, and struggle. Similarly, Lertcharoenwanich (2023) has found differences in collocates for near-synonyms such as blank, empty, and vacant.

Current corpus-assisted discourse analysis makes use of the concept of semantic preference to establish a connection between lexical patterning and institutional and social meaning. Spicksley and Franklin (2023) show how the concept of frontline develops its own set of semantic associations through its connection with occupational groups, which reflect moral and material value allocation in public discourse. Elewa (2025) also shows how the feminine form in spoken religious discourse can be related to different semantic sets depending on grammatical number, which establishes the genre-specific gender meaning. Similarly, Erdentug and Musayeva Vefalı (2018) show how the meaning of old and past develops negative ideological meaning through discourse prosody in New Age discourse. In legal discourse, Lukin and García Marrugo (2024) show how the meaning of the concept of war is developed through its immediate lexical context, even in highly formal discourse. Kazmaly (2024) also shows how distributional analysis can identify the subtle semantic differences between personality adjectives, which supports the analytical value of the concept of semantic preference.

At the same time, researchers have also emphasized the importance of methodological rigor in the analysis of the concept of semantic preference and prosody. Bednarek emphasizes the importance of distinguishing between the concept of semantic preference as the co-occurrence of semantic sets and the concept of semantic prosody as the evaluative coloration of meaning. At the same time, Whitsitt emphasizes the importance of not over-interpreting the categories when they are not well-defined. Cross-linguistic research also shows the difficulties in the analysis of the concept of semantic preference, as Wei and Li (2014) show how the lexical equivalents may not show strong semantic alignment. However, the problem of reliability in the analysis of the concept of semantic preference and prosody is also a concern, as Russnes (2025) shows how low inter-rater reliability was found for the analysis of prosody, despite the shared training, because of the fine-grained nature of the categories involved.

In response to this challenge, recent research has sought to use more transparent methods. Fuoli & Beelitz (2025), for instance, use collocation and concordance analysis coupled with computational methods to investigate semantic fields associated with net zero, thus providing a clearer connection between data and interpretation. Overall, previous research on semantic preference appears to offer a viable perspective on understanding the relationship between collocation, evaluative meaning, and genre, albeit with a need to pay closer attention to grammatical context, semantic categorisation, and coding to avoid impressionistic analysis.

MATERIALS AND METHODS

This study utilizes a comparative corpus-based research design to examine the repeated occurrence of the word risk in regulatory risk materials (RMRs) developed by banks in Malaysia and the United States. The corpus-based research design relies on the repeated occurrence of linguistic patterns as a basis for the study, while the corpus-assisted research design utilizes concordance techniques to interpret the findings (Sinclair, 1991;

Hunston, 2002). This design enables the clear linkage of statistical associations to contextualized meanings, thereby increasing transparency and interpretability.

The research design also utilizes collocation and semantic preference as a methodological framework. Collocation examines the strength of association between risk and its associated words, while semantic preference examines whether these associated words form a semantic set (Stubbs, 2001). These two concepts provide a clear framework for the development of an empirical model of the patterning of institutional meanings (Sinclair, 1991, 1996; Stubbs, 2001; Partington, 2004; Bednarek, 2008). The comparative research design utilizes two functionally equivalent corpora as a basis for the study, as these corpora have a similar genre and time frame (2016-2020).

Corpus Data

Analytical Procedures

The data consist of two corpora compiled from audited risk management reports issued by banks in Malaysia and the USA. The corpora qualify as corpora by virtue of their systematic compilation, electronic storage, and representation of a clearly defined genre in a limited institutional milieu. The corpora are termed CORDMAB—Corpus of Risk Disclosure by Malaysian Banks—and CORDAB—Corpus of Risk Disclosure by American Banks. In terms of comparability, eight banks from each country have been selected and consolidated RMRs collected for five years from 2016 to 2020. This gives 40 RMRs from each country and a total of 80 RMRs. RMRs from Malaysia have been collected from Bursa Malaysia's official website, while RMRs from American banks have been collected from the investor relations or corporate governance sections of their official websites. The RMRs have been selected on the basis that they have a clearly demarcated risk management section.

Finally, the resulting corpus contains 155,162 word tokens for CORDMAB and 717,676 word tokens for CORDAB. Although the American corpus is significantly larger, as is typical of U.S. institutional disclosures due to their level of detail and length (Ren & Lu, 2021), this does not have any negative effect on the analysis. This study is focused on measuring collocation strength using Mutual Information (MI) scores, as opposed to frequency counts (Cheng, 2012; Hunston & Oakey, 2009). An MI measure of ≥ 3.0 was used as a threshold to establish significant collocations (Hunston, 2022). In accordance with established corpus linguistics principles, corpus representativeness and functional comparability have been given greater importance than corpus size comparability (Biber et al., 1998). In both cases, RMRs are compliant with regulations and have been prepared following the Basel III system. In addition, CORDMAB exceeds the minimum corpus size required for lexical analysis of homogeneous genres (Biber, 1993). In order to maintain anonymity of the institutions involved, a coded identifier was given to each report. A summary of corpus composition is given in Table 1.

Table 1. Summary of corpus data

Criteria	**CORDMAB**	**CORDAB**
Source (Annual Reports)	Bursa Malaysia	Banks' websites
No. of Texts	40	40
Word Token	155,162	717,676
Average length	3971	17942
Word Type	4465	13148
Year	2016 – 2020	

Analysis

Two tools of corpus analysis were used: LancsBox 6.0 and the UCREL Semantic Analysis System (USAS). The first tool, LancsBox 6.0, is a tool developed at Lancaster University, UK, and it has functionalities for

concordance, collocation, dispersion, word lists, and n-grams (Brezina et al., 2020). Using this tool, it is possible to identify the different collocation patterns related to the node word risk. Figure 1 presents the user interface of the tool.

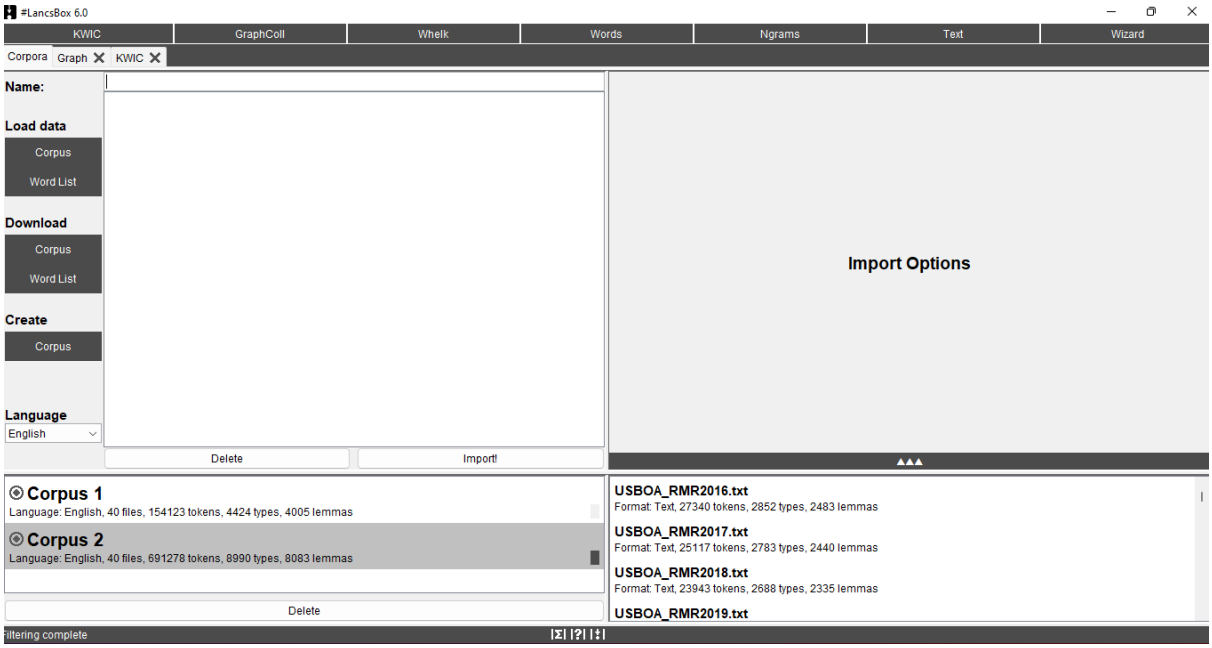

Figure 1. The interface of Lancsbox 6.0

The second tool used for the analysis is the UCREL Semantic Analysis System (USAS), an automatic semantic tagging tool developed by the UCREL research center at Lancaster University. The tool assigns a semantic tag to a word based on a semantic taxonomy with 21 main categories and 232 subcategories (Archer et al., 2002). The tool is used for the analysis of the different collocation patterns related to the node word risk, and it is used here due to the fact that it allows the classification of the different collocates into semantic groups without having to classify them as topics. Previous evaluations of the tool have demonstrated a very high level of precision for automatic semantic tagging (Rayson, 2003; McEnery et al., 2006). Figure 2 presents the user interface of the tool.

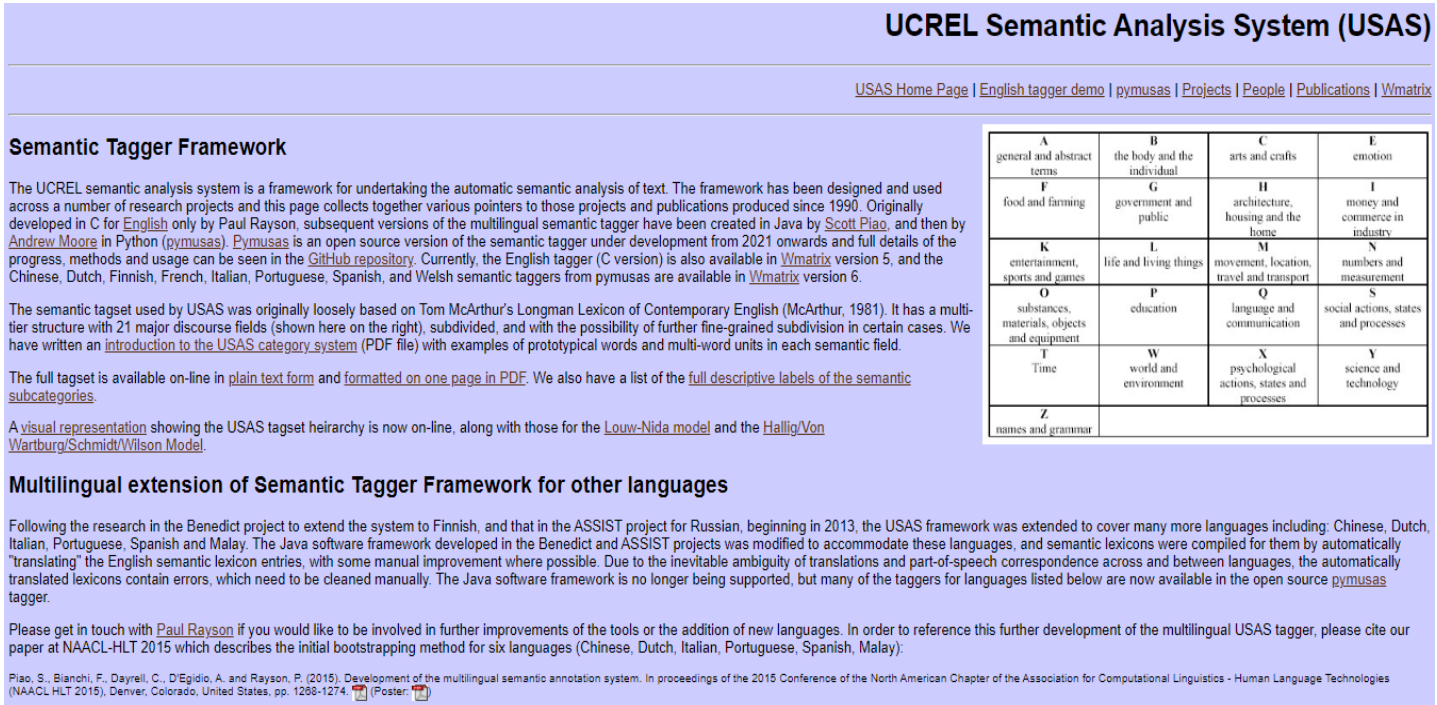

Figure 2. The interface of USAS

Prior to analysis, all RMRs were converted into text file format (.txt), and any diagrams or images that could not be properly converted were excluded from the dataset. Later, they were uploaded into Lancsbox to obtain the collocational patterns. This was done to answer Research Questions 1 and 2. However, only content words were selected in the results. The collocation span was set at five words to the left (L5) and five words to the right (R5) of the node, following Sinclair's (1991) argument that meaningful collocates typically occur within this range. Moreover, only collocates with a Mutual Information (MI) score greater than 3 and log likelihood ≥ 15.13 were included in the analysis, as this threshold indicates a statistically significant and meaningful association between the node word and its collocates (Cheng, 2011; Hunston, 2002, Rayson et al., 2004).

After the significant collocates of risk in both corpora were obtained, the list was uploaded to USAS. This was done to answer Research Question 3 by classifying the collocates into their respective semantic sets. USAS assigned a part-of-speech (POS) tag to each collocate using the CLAWS and later generated the output through its semantic tagging component, SEMTAG. Through the analysis, the semantic fields of the collocates of risk were revealed. In addition, it identified whether the semantics of the collocates carried positive or negative connotations.

FINDINGS

Table 2 shows the 10 most frequently used words in CORDMAB and CORDAB. As expected, the word risk has the highest frequency, as the corpora focuses on risk management. In addition, the word list for both corpora contains more nouns than any other parts of speech. This is probably due to the fact that the RMR outlines the bank's risk management framework and policies, where terms and areas related to risk and risk management are introduced, defined and explained. Furthermore, all of the top 10 words in both corpora were shown to possess very high dispersion rates-a measure that shows how evenly a word is distributed across different texts within a corpus. A high dispersion score (typically close to 1.0 or above 70 on a 0–100 scale, as used in LancsBox) indicates that a word has a consistent appearance throughout most or all the texts, rather than being concentrated in one or two (Brezina et al., 2020; Gries, 2008). This is typical of homogeneous, genre-restricted corpora where core lexical items have regular recurrences as a result of common communicative purposes and textual conventions (Biber et al., 1998; Adolphs, 2006). The fact that these key terms, in particular the ones of risk, management, credit, and board, possess high dispersion confirms that they are not idiosyncratic to either individual banks or years but are central to the discourse on risk management in the whole dataset.

Table 2 Top 10 words used in CORDMAB and CORDAB

	CORDMAB			**CORDAB**		
No.	Type	Frequency	Dispersion	Type	Frequency	Dispersion
1	risk	3984	69.362585	risk	13574	72.413722
2	management	2732	41.69037	our	7164	111.94209
3	group	2710	52.489372	credit	5828	15.420798
4	board	1346	27.915275	loans	5411	38.934198
5	internal	1343	40.745162	management	4683	24.074631
6	business	1184	17.007241	we	3959	55.284998
7	compliance	1139	25.353135	market	3193	14.359594
8	control	1114	29.670129	december	3017	17.425727

9	shariah	920	50.297917	interest	2644	21.628805
10	group's	912	24.708928	portfolio	2561	18.44575

Collocations

In the collocation analysis, only words that naturally co-occurred with risk in both corpora were examined. The collocates were then sorted by frequency of occurrence, and only those with MI scores above 3 were retained, indicating a statistically significant association with risk.As can be seen in Table 3, there are five collocates of risk in both corpora – board, credit, operational, market and liquidity. This is most likely due to the requirements of Basel III, a set of international banking regulations developed by the Basel Committee on Banking Supervision (BCBS). As the regulations are applied globally, it is not surprising that some similar collocates of risk are used.

Table 3 The left collocates of risk in CORDMAB and CORDAB

	CORDMAB			**CORDAB**		
No.	Collocate	MI Score	Frequency	Collocate	MI Score	Frequency
1	group	3.39601	757	credit	4.195	2138
2	group's	4.0251	394	market	4.9661	1999
3	board	3.07029	300	our	3.78574	1979
4	credit	4.94385	254	operational	5.89968	1698
5	shariah	3.20423	225	liquidity	4.26199	900
6	operational	5.07547	221	compliance	5.30701	808
7	market	5.47168	199	model	4.90918	656
8	statement	4.55272	175	interest	3.57929	633
9	subsidiaries	6	170	rate	4.29189	563
10	integrity	7	165	firmwide	5.28872	505

In addition, the left collocates of risk in CORDAB include the personal pronoun our. Further investigation found that the pronouns were used to refer to the banks. This is illustrated in the following excerpt, "We employ our risk management process, referred to as Identify, Measure, Monitor and Control, as part of our daily activities." (USBOA_RMR2020). This suggests that American banks tend to personalise their RMRs, projecting a sense of involvement and responsibility in managing risks. In contrast, Malaysian banks commonly use board/group or group's to indicate possession. For example, "The Board has also established the Board Risk Committee, whose responsibilities, amongst others, include overseeing the effective implementation of the EWRM framework." (MCIB_RMR2016). This finding echoes Ren and Lu (2021), who observed that companies in China likewise avoided the use of personal pronouns when referring to themselves. Instead, they preferred to use Company/Group.

As for the right collocates of risk, they predominantly consist of nouns and adjectives. The collocate management occurs with the highest frequency in the corpus which is unsurprising since the section is typically called 'Risk Management' or 'Risk Management Report'. Similar to the left collocates, several right collocates overlap between the two corpora, including appetite, committee, framework, and management. In addition, the right collocates also include words that help mitigate risks, such as management, control, compliance, system, policy,

framework, boundary and governance. These words are used to convey that the risks undertaken by banks are carefully monitored and kept under control. (Davis & Tama-Sweet, 2012; Tan & Yeo, 2023). The high number of noun collocates of risk reflects an effort to make the RMRs appear highly technical in nature. As a result, they can easily convince readers who may not be able to grasp the actual gist of the reports (Bhatia, 2008). Table 4 summarises the right collocates of risk identified in the corpus.

Table 4. The right collocates of risk in CORDMAB and CORDAB

	CORDMAB			**CORDAB**		
No.	Collocate	MI Score	Frequency	Collocate	MI Score	Frequency
1	management	4.843257	2081	management	5.457299	4121
2	control	4.740113	790	committee	5.116137	1199
3	internal	4.173369	643	appetite	6.081779	776
4	compliance	4.04817	500	framework	5.247718	726
5	appetite	5.740481	471	oversight	5.16257	594
6	framework	4.596287	371	limits	4.962098	575
7	committee	3.851007	345	including	3.671528	526
8	system	4.441148	313	governance	5.135726	507
9	policies	3.800924	260	exposures	4.434231	501
10	external	6	240	exposure	3.811861	454

Semantic Preferences of risk

The analysis of semantic preference determines the link between the collocates of risk with other words that share similar semantic features. The analysis also shows how risk is used in RMRs even though it is known to have negative connotations. The results of the semantic preferences of risk in both corpora were sorted according to their log-likelihood scores as presented in Table 5.

Table 5. Summary of semantic preferences of risk

	CORDMAB				**CORDAB**			
No.	Collocate	Tagset	Likelihood	Category	Collocate	Tagset	Likelihood	Category
1	management	S7.1+	1496.33	In power	management	S7.1+	5540.98	In power
2	appetite	X7+	789.52	Wanted	operational	A1.1.1	3921.08	General actions / making
3	control	S7.1+	525.55	In power	appetite	X7+	1607.01	Wanted
4	operational	A1.1.1	296.82	General actions / making	compliance	S7.1-	1214.21	No power

5	market	I2.2	255.59	Business: Selling	framework	X4.2	840.42	Mental object: Means, method
6	credit	I1.1	197.11	Money and pay	oversight	X3.4-	693.18	Unseen
7	internal	M6	152.12	Location and direction	firmwide	Z99	597.41	Unmatched
8	compliance	S7.1-	113.28	No power	limits	A1.7+	560.68	Constraint
9	system	Y2	104.43	Information technology and computing	credit	I1.1	558.82	Money and pay
10	external	M6	84.61	Location and direction	model	A4.1	541.73	Generally kinds, groups, examples
11	committee	S5+	73.55	Belonging to a group	governance	G1.1	506.06	Government
12	subsidiaries	S5+	69.1	Belonging to a group	liquidity	I1	398.6	Money generally
13	integrity	G2.2+	63.19	Ethical	exposures	A15-	223.54	Danger
14	group	S5+	54.46	Belonging to a group	our	Z8	168.69	Pronouns
15	groups	S5+	50.12	Belonging to a group	rate	I2.2	149.11	Business: Selling
16	shariah	Z99	41.99	Unmatched	market	I2.2	110.85	Business: Selling
17	statement	I1	25.15	Money generally	committee	S7.1+	109.82	In power
18	board	A1.1.1	21.17	General actions / making	including	A1.8+	46.63	Inclusion
19	policies	Q1.2	17.45	Paper documents and writing	exposure	A15-	25.11	Danger
20	framework	X4.2	16.85	Mental object: Means, method	interest	I1.1	18.99	Money and pay

As shown in Table 5, the collocates of risk for each subcorpora belong to five major semantic fields. For CORDMAB, they relate to A: general and abstract terms, G: government and public, I: money and commerce, M: movement, location, location, travel and transport, Q: language and communication, S: social actions, states

and processes, X: psychological actions, states and processes and Y: science and technology and Z: names and grammar. The collocates in CORDAB, on the other hand, show limited diversity, as they are not associated with three domains namely M, Q and Y. The differences in the themes were most likely due to the different focus of the RMRs in the two countries, even though they were prepared according to a similar guideline, Basel III. The focuses include the governance and public-regulatory environment in which risk is managed (such as references to regulators, governance structures, public policy or state-linked initiatives), geographical footprint and cross-border exposure (for example location-based or movement-related descriptions of operations and markets), risk communication and internal reporting processes (for instance disclosure, communication, reporting, review and monitoring routines), and lastly technology-related risk and operational infrastructure like systems, digital channels, technology controls.

In CORDMAB, the collocates of risk mostly belong to S: social actions, states and processes. In this context, risk is presented as systematically managed, as there are entities responsible for mitigating it. Hence the co-occurrence of risk with group, groups, subsidiaries, management, control and compliance. These collocates also show that despite the uncertainty in the business, the banks are well prepared and ready to face the challenges ahead. Another noticeable pattern is the association with the semantic field of money, which is expected given that RMRs are a key genre within banking institutions. The collocates in this field are related to payments activities, such as credit and business and to commercial transactions, such as market.

The semantic preferences of risk in CORDAB differ noticeably from those in CORDMAB, as they are linked to a narrower range of semantic fields. In contrast to the other corpus, the collocates in CORDAB are predominantly associated with general and abstract terms. The node word risk co-occurs with operational and compliance, which describe the risk categories. In addition, risk also collocates with terms denoting inclusion, (i.e. including), constraint, (e.g. limits), and danger (e.g. exposure and exposures). Another emerging semantic theme is money. Similar to CORDMAB, which includes the collocates credit and market, CORDAB features three additional collocates within this theme, one of which is liquidity, a term that broadly represents money or financial resources. The other two collocates, interest and rate, belong to similar semantic sets of credit and market respectively.

One of the features of USAS is the identification of words that are perceived as positive or negative on a semantic scale. Although this does not apply to all words, as it is only found in some semantic sets, it helps to clarify how positive and negative words are used as collocations, especially for words that are known to have a negative meaning, such as risk. In CORDMAB, eight collocates were found to be semantically positive, which included group, groups, subsidiaries, integrity, management, control, internal, appetite, framework, committee and only one collocate was negative, compliance. It was regarded as negative since it belonged to the category of no power. In RMRs, compliance often collocates with “risk assessment” (compliance risk assessment) which shows banks’ effort to meet the governance expectations set by the regulatory authorities (Securities Commission Malaysia, 2021). Figure 3 shows concordance lines of management and compliance in CORDMAB.

Left	Node	Right
all important areas for the	Management	and Board to ensure continuity
reviewed this Statement on Risk	Management	and Internal Control. Their limited
and effectiveness of the risk	management	and internal control systems of
the effectiveness of the Risk	Management	and Internal Control Framework and
review based on the Risk	Management	and Internal Control system adopted
consideration the assurance from the	management	and input from the relevant
is viewed that Group's Risk	Management	and Internal Control system are
STATEMENT ON RISK	MANAGEMENT	AND INTERNAL CONTROL BOARD RESPONSIBILITY
of internal control and risk	management,	and is constantly keeping abreast
responsibilities on the Group's risk	management	and internal control system. The

Left	Node	Right
cover inter alia financial, operational,	compliance	controls and risk management procedures.
Group Board Risk Management and	Compliance	Committee (GBRMC) to include the
of oversight on internal controls,	compliance	and risk management strategies, policies
operating statistics, legal and regulatory	compliance,	breach of law or regulations
that the Group's risk management,	compliance	and system of internal controls
based on the risk management,	compliance	and internal control system. 2.
awareness programs/training on controls and	compliance	including controls certification programs to
creating a robust control and	compliance	environment. g. The management of
risk management policy and risk	compliance.	GRMF is governed by a
the principle that Risk and	Compliance	is Everyone's Responsibility and hence,

Figure 3. Concordance lines of management (+) and compliance (-) in CORDMAB

In contrast, CORDAB has less collocates with positive semantic sets and more collocates with negative sets. The positive ones included management, committee, appetite, limits and including while the negative ones were compliance, oversight, exposure and exposures. Figure 4 shows concordance lines of management and compliance.

Left	Node	Right
from the management-level Enterprise Risk	Management	Committee regarding current or emerging
current or emerging risk matters.	Management	Oversight of Risk In addition
that oversee the Company's risk	management	framework, the Company has established
in carrying out their risk	management	responsibilities. Each risk-focused governance committee
of executives. The Enterprise Risk	Management	Committee, chaired by the Company's
Risk Officer (CRO), oversees the	management	of all risk types across
the Company. The Enterprise Risk	Management	Committee reports to the Board's
governance and oversight at the	management	level. Corporate Risk develops our
the context of our risk	management	framework described above. As part
reported to the Enterprise Risk	Management	Committee on a quarterly basis

Left	Node	Right
we incur when acting in	compliance	with the specified standard. For
of historical market data in	compliance	with regulatory requirements. Total Stressed
and after successful completion and	compliance	with terms during this period,
non-financial, such as operational including	compliance	and model risks, strategic and
its oversight of the Company's	compliance	with legal and regulatory requirements.
a result of cyber attacks.	Compliance	Risk Management Compliance risk (a
cyber attacks. Compliance Risk Management	Compliance	risk (a type of operational
associated impacts, including to customers.	Compliance	risk encompasses violations of applicable
responsibility for all aspects of	compliance	risk, including financial crimes risk.
Board's Risk Committee approves the	compliance	risk and financial crimes risk

CONCLUSION

The current study aims to examine the collocational profile of the word risk, including its semantic associations, found in the RMRs of selected Malaysian and American banking institutions. The study shows that, overall, there is a degree of similarity and difference in the collocational profile of the word risk across the reporting contexts of the selected banking institutions. Firstly, it is found that there is a degree of overlap in the high-salience collocates of CORDMAB and CORDAB, indicating that there is a common core of words that are used to name and describe different types of risks, exposures, and governance practices. This similarity is likely a result of internationally agreed upon disclosure practices, including those stipulated under the Basel III accord, which promotes the adoption of similar reporting practices to provide greater market discipline and transparency, including the reporting of prudential information and risks (Basel Committee on Banking Supervision, 2017). However, it is important to note that the similarity should not be attributed solely to the adoption of the Basel III accord. The similarity could be attributed to a number of factors, including the constraints of reporting genres, institutional pressures, and even the adoption of international risk taxonomies.

Second, the analysis indicates that the rhetorical and phraseological choices are systematic and vary across the corpora. In the case of the CORDAB corpus, the use of the possessive pronoun our indicates a stronger tendency towards self-inclusive reference and reader orientation, which can be considered an engagement or alignment strategy in institutional communication, rather than a referential choice (Hyland, 2005; Asay et al., 2018). Finally, the profile of the most frequent collocates in both corpora indicates that the most frequent word class is nouns, with a lower occurrence of adjectives. This profile can be considered an indication of the classificatory and definitional orientation of regulated reporting, where risks are named, classified, and explained through technical labels and terminology that are considered constant and predictable (Hunston, 2002; Sinclair, 1991). Overall, these findings indicate a discourse style that orients towards taxonomic description and managerial control, while still allowing some interpersonal positioning.

With respect to semantic preference, the analysis indicates that the semantic field of the word risk in the CORDMAB corpus extends to a wider range of semantic fields compared to the CORDAB corpus. However,

the semantic fields found in the American corpus are found to fall within the range of those found in the Malaysian corpus. This indicates that, while the purpose of the reports in both contexts is the same, the range of thematic focus or emphasis of the reporting may vary. The findings are therefore in accordance with a cautious stance that similarities in the use of words and rhetorical and phraseological choices across contexts are indicative of shared institutional purposes, while the range of semantic fields found in the American corpus may indicate the presence of contextually or locally salient reporting priorities or preferred reporting templates or the distribution of the content of the report sections. Finally, the semantic profiling indicates that the co-text of the word risk is often oriented towards mitigation, management, and governance. This can be considered an indication of the rhetorical tendency of the word risk to be often presented in a reassuring manner, which may become overly positive at times (ACCA, 2014). This, however, does not imply that the negative potential of the word risk is negated; it indicates that RMRs often present a new context of the word risk that orients towards the notion of control.

This study makes a contribution to the growing body of linguistic research on RMRs by illustrating the way in which a regulated disclosure genre can create risk accountability through lexico-semantic routines. The linguistic research on risk disclosures is relatively underdeveloped and considered an under-explored area in associated business and accounting research (Aziz et al., 2022; Yusob et al., 2022). This current study offers a replicable linguistic analysis of the way in which risk is constructed as meaningful through co-selection rather than the selection of individual words (Aziz et al., 2022; Sinclair, 1991; Stubbs, 2001; Yusob et al., 2022;). The current study offers a contribution to the linguistic understanding of risk discourse in that it illustrates a phraseological approach to meaning, where collocation and semantic preference can be seen as evidence of the way in which meaning is created at the institutional level (Bednarek, 2008; Partington, 2004). The methodological approach to the study, which combines collocation and semantic tagging, offers a way in which the two approaches can be linked to provide a more complete discourse analysis, provided that the results are checked and validated to ensure that the discourse claims can be made (Hunston, 2002). The current study offers a number of avenues that can be followed in further research, such as (i) the examination of more sectors than the banking sector, (ii) the examination of a wider time period to see whether the phraseological pattern changes over time, and (iii) the examination of more nodes and multiword units such as risk appetite, risk exposure, and emerging risk to gain a more comprehensive phraseological pattern in the way in which risk is constructed in discourse.

REFERENCES

1. Abdullah, I. H., Hashim, R. S., & Mohamed Husin, N. (2011). Lexical associations of Malayness in Hikayat Abdullah: A collocational analysis. Research Journal of Applied Sciences, 5(6), 429-433.
2. ACCA (2014). Reporting Risk, The Association of Chartered Certified Accountants, London.
3. Adolphs, S. (2006). Introducing electronic text analysis: A practical guide for language and literary studies. Routledge.
4. Al-Tamimi, H. A. H. and Al-Mazrooei, F. M. (2007). Banks' risk management: a comparison study of UAE national and foreign banks. The Journal of Risk Finance, 8(4), 394-409. https://doi.org/10.1108/15265940710777333
5. Aziz, R. A., Abas, N., Yusob, K. F., & Yusof, N. M. (2022). Genre Analysis of Risk Management Reports by an Islamic Financial Institution in Malaysia. GEMA Online Journal of Language Studies, 22(2).
6. Baker, P., McEnery, T. and Gabrielatos, C. (2007) 'Using Collocation Analysis to Reveal the Construction of Minority Groups: The Case of Refugees, Asylum Seekers and Immigrants in the UK Press', paper given at Corpus Linguistics 2007, University of Birmingham, 28–30 July 2007. Available at: http://eprints.lancs.ac.uk/602/.
7. Begagić, M. (2013). Semantic preference and semantic prosody of the collocation make sense. Jezikoslovlje, 14(2–3), 403–416. Bhatia, V. K. (2008). Genre analysis, ESP and professional practice. English for specific purposes, 27(2), 161-174.
8. Biber, D. (1993). Representativeness in corpus design. Literary and Linguistic Computing, 8(4), 243-257.
9. Biber, D., Conrad, S., & Reppen, R. (1998). Corpus linguistics: Investigating language structure and use. Cambridge University Press.
10. Cheng, W. (2011). Exploring corpus linguistics: Language in action. Routledge.

11. Davis, A. K., & Tama-Sweet, I. (2012). Managers' use of language across alternative disclosure outlets: earnings press releases versus MD&A. Contemporary Accounting Research, 29(3), 804-837.
12. Erdentuğ, S., & Musayeva Vefalı, G. (2018). What is "old" and "past" in New Age discourse: A qualitative analysis of corpus evidence. Discourse, Context & Media, 24, 85–91.
13. Firth, J. R. (1957). Papers in linguistics 1934-1951. Oxford University Press.
14. Flowerdew, L. (2012). Corpora and language education. Palgrave Macmillan.
15. Fuoli, M., & Beelitz, A. (2025). Corporate buzzword or genuine commitment? A corpus-assisted analysis of corporate 'net-zero' pledges by major global corporations. Applied Corpus Linguistics, 5(3), 100142.
16. Gurendrawati, E., Khairunnisa, H., Ulupui, I. G. K. A., Suryarini, T., & Zakaria, A. (2021). Bank risk profile and credit growth in Indonesia. Jurnal Ilmiah Akuntansi Dan Bisnis, 16(1), 84. https://doi.org/10.24843/jiab.2021.v16.i01.p06
17. Gries, S. T. (2013). Statistics for linguistics with R: A practical introduction. Walter de Gruyter.
18. Hardy, D. E., & Colombini, C. B. (2011). A genre, collocational, and constructional analysis of RISK. International Journal of Corpus Linguistics, 16(4), 462-485.
19. Hunston, S. (2002). Corpora in applied linguistics. Cambridge University Press.
20. Hunston, S. (2007). Semantic prosody revisited. International Journal of Corpus Linguistics, 12(2), 249–268.
21. Kazmaly, A. (2024). The lexical semantics of personality words: On 'shy' and 'outgoing'. Lingua, 302, 103695.
22. Lertcharoenwanich, P. (2023). Semantic preferences of the adjectives blank, empty and vacant: A corpus-based study. LEARN Journal: Language Education and Acquisition Research Network, 16(1), 365–383.
23. Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), Text and technology: In honour of John Sinclair (pp. 157–176). John Benjamins.
24. Lukin, A., & Marrugo, A. G. (2024). War in law: A corpus linguistic study of the lexical item war in the laws of war. Applied Corpus Linguistics, 4(1), 100088.
25. Mazumder, M. M. M., & Hossain, D. M. (2018). Exploring the nature of risk disclosure in the annual report narratives of Bangladeshi pharmaceutical companies: an impression management perspective. International Journal of Comparative Management, 2(3-4), 273-296.
26. McEnery, T., Xiao, R., & Tono, Y. (2006). Corpus-based language studies: An advanced resource book. Routledge.
27. Muigai, R. G. and Maina, M. W. (2018). Effect of credit risk management practices on performance of commercial banks in Kenya. International Journal of Finance and Banking Research, 4(3), 57. https://doi.org/10.11648/j.ijfbr.20180403.12
28. Müller, M., & Mell, R. M. (2022). 'Risk' in political discourse. A corpus approach to semantic change in German Bundestag debates. Journal of Risk Research, 25(3), 347-362.
29. Palacios Martínez, I. M. (2010). "It ain't nothing to do with my school." Variation and pragmatic uses of ain't in the language of British English teenagers. English Studies, 91(5), 548–566.
30. Phoocharoensil, S. (2020). A Genre and Collocational Analysis of Consequence, Result, and Outcome. 3L: Southeast Asian Journal of English Language Studies, 26(3).
31. Prihantoro, P. (2022). A Collocation Analysis of 'energy'in Brown Family Corpus. In E3S Web of Conferences (Vol. 359, p. 03012). EDP Sciences.
32. Ren, C., & Lu, X. (2021). A multi-dimensional analysis of the management's discussion and analysis narratives in Chinese and American corporate annual reports. English for Specific Purposes, 62, 84-99.
33. Römer, U. (2011). Corpus research applications in second language teaching. Annual Review of Applied Linguistics, 31, 205-225. https://doi.org/10.1017/s0267190511000055
34. Rayson, P. (2003). Matrix: A Statistical Method and Software Tool for Linguistic Analysis through Corpus Comparison. [Doctoral thesis, Lancaster University].
35. Rayson, P., Berridge, D. and Francis, B. (2004) 'Extending the cochran rule or the comparison of word frequencies between corpora', in G. Purnelle, C. Fairon and A. Dister (eds) Le Poids des Mots: Proceedings of the 7th International Conference on Statistical Analysis of Textual Data (JADT 2004) (Louvain-la-Neuve, Belgium: Presses universitaires de Louvain), pp. 926–36.
36. Russnes, M. (2025). Semantic prosody, categorisation and inter-rater reliability. Research Methods in Applied Linguistics, 4(3), 100264.

37. Sinclair, J. (1966). Beginning the study of lexis. In C. E. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robins (Eds.). In memory of J. R. Firth (pp. 410-430). Longman.
38. Sinclair, J. (1991). Corpus, concordance, collocation. Oxford University Press.
39. Sinclair, J. (1996). The search for units of meaning. Textus, 9(1), 75-106.
40. Sinclair, J. and Carter, R. (1994). Corpus, concordance, collocation. The Modern Language Journal, 78(3), 407. https://doi.org/10.2307/330144
41. Spicksley, K., & Franklin, E. (2023). Who works on the 'frontline'? comparing constructions of 'frontline'work before and during the COVID-19 pandemic. Applied Corpus Linguistics, 3(3), 100059.
42. Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. Functions of language, 2(1), 23-55.
43. Stubbs, M. (2001). Words and phrases: Corpus studies of lexical semantics. John Wiley & Sons.
44. Tan, H. T., & Yeo, F. (2023). You have been forewarned! The effects of risk management disclosures and disclosure tone on investors' judgments. Accounting, Organizations and Society, 105, 101400.
45. Wei, N., & Li, X. (2014). Exploring semantic preference and semantic prosody across English and Chinese: Their roles for cross-linguistic equivalence. Corpus Linguistics and Linguistic Theory, 10(1), 103–138.
46. Yusob, K. F., Mohamad Ali, AChan, M. Y., & Lee, G. I. (2022). Rhetorical Organisation of Risk Management Reports by Malaysian Banks. Proceedings, 82(1), 80-88.
47. Zethsen, K. K. (2021). Corpus–based cognitive semantics: Extended units of meaning and their implications for Translation Studies. Linguistica Antverpiensia, New Series – Themes in Translation Studies 7, 249–262. https://doi.org/10.52034/lanstts.v7i.218.