Asia Pacific Journal of Corpus Research
Vol. 3, No. 2, 17-33

http://www.icr.or.kr/apjcr
e-ISSN 2733-8096
https://doi.org/10.22925/apjcr.2022.3.2.17

Clustering Keywords to Define Cybersecurity: An Analysis of Malaysian and ASEAN Countries' Cyber Laws

Siti Aeisha Joharry
(Universiti Teknologi MARA, UiTM)
Syamimi Turiman
(Universiti Teknologi MARA, UiTM)
Nor Fariza Mohd Nor
(Universiti Kebangsaan Malaysia, UKM)

Joharry, S. A., Turiman, S., & Mohd Nor, N. F. (2022). Clustering keywords to define cybersecurity: An analysis of Malaysian and ASEAN countries' cyber laws. *Asia Pacific Journal of Corpus Research, 3*(2), 17-33.

While the term is nothing new, 'cybersecurity' still seems to be defined quite loosely and subjectively depending on context. This is problematic especially to legal writers for prosecuting cybercrimes that do not fit a particular clause/act. In fact, what is more difficult is the non-existent single 'cybersecurity law' in Malaysia, rather than the current implementation of 10-related cyber security acts. In this paper, the 10 acts are compiled into a corpus to analyse the language used in these acts via a corpus linguistics approach. A list of frequent words is firstly investigated to see whether the so-called related laws do talk about cybersecurity followed by close inspection of the concordance lines and habitually associated phrases (clusters) to explore use of these words in context. The 'compare 2 wordlist' feature is used to identify similarities or differences between the 10 Malaysian cybersecurity related laws against a corpus of cyber laws from other ASEAN countries. Findings revealed that ASEAN cyber laws refer mostly to three cybersecurity dominant themes identified in the literature: technological solutions, events, and strategies, processes, and methods, whereas Malaysian cybersecurity-related laws revolved around themes like human engagement, and referent objects (of security). Although these so-called cyber related policies and laws in Malaysia are highlighted in the National Cyber Security Agency (NACSA), their practical applications to combat cybercrimes remain uncertain.

Keywords: Corpus Linguistics, Cyber Security, Cybercrimes, Information Security, National Cyber Policy

1. Introduction

At the current speed of technology and advancement, 'cybersecurity' is no longer an uncommon term. To many, cybersecurity covers a very broad topic. The concept of cybersecurity, computer ethics, cybersafety, cybercrime, and cyberspace are interrelated with each other (Zukarnain et al., 2020). In the past, Mat and colleagues opine that "[c]ybersecurity is one of the recent areas of concern for national and global security in the 21st century" (2019, p. 214). There is numerous research on defining 'cybersecurity' but one relevant study investigated literature on the term 'cybersecurity' through lexical semantic analysis and proposed a new definition:

> "The approach and actions associated with security risk management processes followed by organizations and states to protect confidentiality, integrity and availability of data and assets used in cyber space. The concept includes guidelines, policies and collections of safeguards, technologies, tools and training to provide the best protection for the state of the cyber environment and its users" (Schatz et al, 2017, p. 66).

Cybersecurity is argued by Mat et al. (2019) as difficult to handle because "[c]ybersecurity is a new area of security study which is not fully understood in terms of its nature, dimension, trends of occurrence and other issues related to it" (p. 215). In fact, cybercrimes are fast changing due to rapid advance in technology while law, as Brenner (2010) admits, "changes slowly" (p. 14). Researchers add that "[t]here is still the existence of [a] gap in cybercrime and cyber law as cyber laws are still poorly construed or simply do not apply to the types of crimes to be investigated" (Mat et al., 2019, p. 216).

In recent years, cybersecurity has become a critical global problem across the world leading to an increased reliance on the Internet that has created a large number of security threats which can become problematic. ASEAN, therefore, has stepped up and taken action to address the cybersecurity problem as it is growing at a fast pace with regard to digital connectivity (Mohamed Mizan et al., 2019). According to Mohamed Mizan et al. (2019), literature reports on cybersecurity in ASEAN countries mainly focused on cybersecurity challenges and issues. They found that ASEAN studies on cybersecurity focuses on several themes, such as defense against innovative cyber-attacks (DCA), strategies against cybersecurity threats (SCS), government policies and protection against privacy (GPP), protection of computer infrastructures in the government (PCG), and legal and ethical issues on cyberspace (LEC). Although research in the area has mushroomed over the years, types of research on cybersecurity from a linguistic perspective and through the use of corpus linguistics approach are rather scarce (except one study on cybersecurity using computational linguistics methods that is Ghasiya and Okumura, 2022).

1.1. Cybersecurity in Malaysia

According to statistics collected from Cyber Security Malaysia's website, a total number of 10,699 incidents were reported in 2018, representing a 34% increase compared to the previous year. The report further notes that in 2018, the most reported incident is online fraud with a total of 5,123 incidents coming from organizations, home users, private sectors, industries as well as from foreign entities and was predicted that the number would continue to escalate in the coming years. This is hard evidence that shows cybercrimes are increasing at an alarming rate (Supayah & Ibrahim, 2016). In fact, it was argued that "[c]yber attacks are also becoming sophisticated in their ability to evade detection by security appliances and Law Enforcement Agencies" (same report above). Given the advent of the Multimedia Super Corridor (MSC) and other developing technologies, Malaysia firstly outlined several cyber laws that were discussed in Arowosaiye (2013, p. 125) as part of the country's initiative to assure and attract foreign investors in realising the Malaysian Vision 2020 at the time. Arowosaiye further stated that the related laws pertain to two sets of information technology laws: 1) commerce enabling cyber laws and 2) societal cyber laws (see Zaiton Hamin, 2004, p. 211). Table 1 presents 10 cybersecurity-related laws as described in the official portal of Malaysia's National Cyber Security Agency (NACSA).

Table 1. Malaysian Cyber Laws according to National Cyber Security Agency (NACSA)

No.	Laws	About
1.	Act 332: Copyright (Amendment) Act 1987	• An Act to make better provisions in the law relating to copyright and for other matters connected therewith.
2.	Act 562: Digital Signature Act 1997	• An Act to make provision for, and to regulate the use of, digital signatures and to provide for matters connected therewith.
3.	Act 563: Computer Crimes Act 1997	• An Act to provide for offences relating to the misuse of computers.
4.	Act 564: Telemedicine Act 1997	• An Act to provide for the regulation and control of the practice of telemedicine; and for matters connected therewith.
5.	Act 574: Penal Code 2015	• An Act relating to criminal offences.

6.	Act 588: Communications and Multimedia Act 1	• An Act to provide for and to regulate the converging communications and multimedia industries, and for incidental matters.
7.	Act 658: Electronic Commerce Act 2006	• An Act to provide for legal recognition of electronic messages in commercial transactions, the use of the electronic messages to fulfill legal requirements and to enable and facilitate commercial transactions through the use of electronic means and other matters connected therewith.
8.	Act 680: Electronic Government Activities Act 2	• An Act to provide for legal recognition of electronic messages in dealings between the Government and the public, the use of electronic messages to fulfill legal requirements and to enable and facilitate the dealings through the use of electronic means and other matters connected therewith.
9.	Act 709: Personal Data Protection Act 2010	• An Act to regulate the processing of personal data in commercial transactions and to provide for matters connected therewith and incidental thereto.
10.	Act 825: Anti-Fake News (Repeal) Act 2020	• The Anti-Fake News Act 2018 [Act 803] is repealed.

Source: https://www.nacsa.gov.my/legal.php

As noted in Arowosaiye (2013), six primary cyber laws or information technology laws were enacted in response to the need of the Malaysian Multimedia Super Corridor (MSC) and the Malaysian national ambition to become a developed nation by the year 2020 (Arowosaiye, 2013) and they were the Copyright (Amendment) Act 1997, Telemedicine Act 1997, Communications and Multimedia Act 1998, Digital Signature Act 1997, and Computer Crimes Act 1997. In keeping with the latest Malaysian cyber laws as shown on their website, NACSA added another five more: Electronic Commerce Act 2006, Electronic Government Activities Act 2007, Personal Data Protection Act 2010, Penal Code, and Anti-Fake News (Repeal) Act 2020 (Malaysian Cyber Laws).

Since there have been changes to the set of cyber related laws in Malaysia, it would be timely to compare and contrast cybersecurity legal discourses between Malaysia and its neighbouring ASEAN countries. In essence, the purpose of this study is to examine two research questions:

1) What are the similarities and differences between the Malaysian cybersecurity-related laws and cyber laws in other ASEAN countries?
2) To what extent are these laws reflecting the dominant themes of an interdisciplinary definition of 'cybersecurity'?

What follows suit are key insights from the literature that would underpin the methodology of this study. Again, as to reiterate, the goal of this paper is to compare two corpora of cyber laws: one in Malaysia and another from its neighbouring ASEAN region in terms of how the legal discourse is described or written similarly or differently. This is then continued with an in-depth analysis of Craigen et al.'s (2014) cybersecurity dominant themes that are discussed together with the keywords extracted from the corpora. This would hopefully tie us back to how interdisciplinary cybersecurity law-making or legal discourse in ASEAN countries are as well as shedding light about how ASEAN countries can exchange information on ways to strengthen descriptions of cybersecurity in legal texts.

2. Key Insights from Literature

2.1. Defining Cybersecurity

Schatz et al. (2017) argued that there seems to be a "lack of a uniformly accepted definition of cyber security as described in previous studies" (see p. 55). They go on in their paper to group possible definitions of cybersecurity into three categories: industry-based, government-based, and academic-based. These three categories make up the "authoritative" sources that would be considered a more

generic and non-biased perspective of the definition. While this may be the case, Mat et al. (2019) mentioned earlier that the description of cybercrime and cyber laws is problematic and differences between applications of these laws are poorly construed and generally irrelevant to the types of crimes investigated. Additionally, Reyes, O'Shea, Steele, Hansen, Jean and Ralph (2007) noted that "cyber crime laws are still poorly worded or simply don't apply to the types of crimes being investigated" (p. 4).

But first, the term 'cybersecurity' is hard to pin down. One study attempts to redefine the term cybersecurity based on an in-depth analysis of the relevant literature on the topic.[1] Craigen et al. (2014) have proposed that "[c]ybersecurity is the organization and collection of resources, processes, and structures used to protect cyberspace and cyberspace-enabled systems from occurrences that misalign de jure from de facto property rights" (p. 17). They go on to argue that:

> "[a]rticulating a concise, inclusive, meaningful, and unifying definition will enable an enhanced and enriched focus on interdisciplinary cybersecurity dialectics and thereby will influence the approaches of academia, industry, and government and non-governmental organizations to cybersecurity challenges" (2014, p. 13).

2.2. Unpacking 'Cybersecurity' and Its Dominant Themes

To unpack Craigen et al.'s (2014) definition of 'cybersecurity', we can look at the term into three parts: what it is (usually an organization that has a collection of resources, processes, and structures); what it does (that protects cyberspace and cyberspace-enabled systems from…); and the specified area of concern (occurrences that misalign standards established by law from standards that are based on facts but not formally recognised). Redefining 'cybersecurity' to Craigen et al. (2014) is important because in their review of the literature, they did not find a definition that is inclusive, impactful, and unifying. They believe that the concept "is a complex challenge requiring interdisciplinary reasoning; hence, any resulting definition must attract currently disparate cybersecurity stakeholders, while being unbiased, meaningful, and fundamentally useful" (p. 15).

Furthermore, they discovered that there are five typical dominant themes surrounding 'cybersecurity', namely technological solutions; events; strategies, processes, and methods; human engagement; and referent object (of security) (Craigen et al., 2014, p. 15). While they did not elaborate on these further, it could be seen that the themes are quite straightforward. 'Technological solutions' would mean to refer to solutions that are technology-based (e.g., computer software, specific machine learning) while 'events' would refer to cyber occurrences or activities such as cybercrime, phishing, scams and others. 'Strategies, processes and methods' may seem to be interlinked with technological solutions, but perhaps could be identified as something more specific like a particular method or technique that is used to combat the events. 'Human engagement' would be significant as it is also relevant to see who does what and how things are done unto, whereas 'referent object (of security)' could be subjective insofar how different countries identify an object as a referent to/of security.

In this paper, an attempt is made to collect descriptions of cyber laws from ASEAN countries and to compare and contrast them against Malaysia's 10 cybersecurity-related laws in order to see any variabilities, particularly in the way the laws are written as well as evaluating for their interdisciplinary quality through comparison with the cybersecurity dominant themes described in Craigen et al. (2014). Through use of the corpus linguistics approach, these legal documents are firstly compiled into two separate corpora and using a corpus linguistic software, will we be able to linguistically analyse how cybersecurity is framed further.

---

[1] To the best of the authors' knowledge, there is no such study within the discipline that defines cybersecurity explicitly and in turn, Craigen et al.'s (2014) was used as the theoretical framework underpinning the analysis of corpus findings in the study.

3. Methods

3.1. What is Corpus Linguistics?

Corpus linguistics (CL) is an area of study that focuses upon a set of procedures, or methods, for studying language and deals with "some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions" (McEnery & Hardie, 2012, p. 1). In other words, corpus linguistics involves the use of computers to study and analyse naturally-occurring language that is collected with some purpose in mind and stored in what is known as a 'corpus' (*corpora* for plural), or a "large collection of authentic texts that have been selected and organised following precise linguistic criteria" (Leech, 1991, p. 8). Following Biber, Reppen and Friginal (2010), corpus linguistics is a research approach to examine spoken and written discourse, which provides empirical, frequency-based investigations of naturally occurring language-in-use. In the context of our paper, we will be using computer-based techniques (through a corpus tool/software) to extract and automatically generate linguistic findings that could be analysed in terms of Craigen et al.'s (2014) cybersecurity dominant themes.

Since cyber laws are a legal type of written discourse, there is an opportunity to collect these legal documents that are freely available to the public on the web and create a corpus that serves to understand just that: language used to frame or describe cyber laws in Malaysia. Through simple conversion of the typical PDF documents to TXT, each cyber law is labelled and compiled into its respective corpus. Corpus analysis software allows the users to process and organise the textual data, and calculates statistical information about the data in the corpus that offers the language analysis to be more empirical. Generally, the classic techniques involved in corpus linguistics are frequency lists through generating wordlists, collocational analysis and concordancing.

3.2. How Does CL Help to Analyse Language?

The present study employs WordSmith Tools 8.0[2] to extract linguistic information from the corpus built among the 10 cyber security-related laws in Malaysia, henceforth: Malaysia CyberSecurity-related Laws (MCSL). Since we were interested in examining how cyber laws are written, we extended our collection of cyber laws to other ASEAN countries, mainly taken from cyberlaws.net where users are provided with a list of cyber laws from different countries. This allowed us to make a separate reference corpus on the same topic, but from other countries within the ASEAN region.[3] More specifically, keyness of items generated automatically from the 'Compare 2 wordlists' feature from WordSmith Tools were maintained using the default calculation (focus on Log Ratio minimum of 1.500) with significance p-value not more than 0.01 (this means that we had more words to work with) and no specific criteria was excluded.[4] This is similar to the keyword analysis (obtained from the 'keyword' feature of WordSmith Tools), where the frequency of all items in the two corpora is compared followed by concordance analysis of the key items to establish patterns of differences, absence and similarity in the two corpora (Gries, 2010, p. 285). The normalised frequency of particular items (of which two different size corpora can benefit from the 'Compare 2 wordlists' feature) is firstly compared as a starting point in corpus studies adopting the bottom-up approach (corpus-driven analysis). Word clusters that were identified using the same corpus tool were also carried out without making any

[2] Scott, M. (2020). WordSmith Tools version 8. Stroud: Lexical Analysis Software.

[3] https://cyberlaws.net/cyber-law-repository/cyber-laws-different-countries/

[4] A word is said to be "key" if it occurs in the text at least as many times as the user has specified as a minimum frequency; its frequency in the text when compared with its frequency in a reference corpus is such that the statistical probability as computed by an appropriate procedure is smaller than or equal to a p value specified by the user; and in addition, the strength of keyness must be at least as great as the minimum log ratio set by the user (WordSmith Tools manual, 2020).

changes to the default setting; the choice to investigate 3-word sequences proved reliable as similar and different clusters cropped up more compared to generic short 2-word sequences or longer less frequent types (4/5-word sequences).

Corpus linguistic studies using keyword analysis have been employed in many disciplines. Health care research has benefitted from using corpus linguistics approach, for instance, consultation between health care professionals and patients in the UK, provide insights into how language is used to indicate and identify the different phases and purposes of consultation between the health care professionals and patients, and the relationship between the health care professionals and their patients (Adolphs et al., 2004). Meanwhile, Kim and Park (2019) focused on identifying research trends in arts psychotherapy through extensive keyword network analysis. The results revealed the evolution of arts psychotheraphy from a longitudinal perspective, beginning in the 1990s through to the new millennium. Banaji et al. (2017) explored young people's conceptualization of active citizenship with eight participating researchers examining issues related to youth, active citizenship, and Europe. Climate change research is another area that employs corpus linguistics using data from sources such as newspaper and social media (e.g., Koteyko, 2010; Jaspal & Nerlich, 2014; Fløttum et al. 2014; Skovgaard, 2014; Willis, 2017). Human rights issue has also been examined (migrant workers in Gabrielatos & Baker, 2008) as well as (mis)representations of Muslims in the newspaper (Baker et al. 2013). Although this review is not exhaustive, it shows the applicability of corpus linguistics approach using keyword analysis or keyness that can provide insights into language use, trends and research implications for further research in a particular field or discipline.

Since Craigen et al. (2014, p. 13) note that the term [cybersecurity] is used broadly and its definitions are highly variable, context-bound, often subjective, and, at times, uninformative, they call on for a more multidisciplinary approach. In turn, the present study combines two sets of cyber(security) laws within the ASEAN region (including Malaysia) to compare and contrast between how they are described in relation to cybersecurity, particularly with use of corpus linguistic methods.

Table 2 presents the two corpora used in this study. Although there were more (15) laws/acts in the ASEAN Cyber Laws corpus (henceforth, ACLC), the Malaysian CyberSecurity-related Laws corpus (MCSLC) still outnumbered them in terms of total number of words (171,882). This allowed for the 'Compare 2 Wordlists' feature to be used in discerning whether there are any stylistic differences, i.e. in the way in which the laws were written, between the two corpora (example is shown in Appendix 1). More importantly, significance testing between BIC and LogR provided in the tool, helps determine whether items chosen as key words are statistically significant between two different size corpora.[5] This would eventually lead to answering Research Question 1) What are the similarities and differences between the Malaysian cybersecurity-related laws and cyber laws in other ASEAN countries? Meanwhile, to address the second Research Question, 2) To what extent these laws reflect the dominant themes of an interdisciplinary definition of cybersecurity, close inspection of the concordance lines will be analysed in terms of Craigen et al.'s (2014) cybersecurity dominant themes.

Table 2. Description of the Corpora

Corpus	Number of Laws/Acts	Number of Words
Malaysian CyberSecurity-related Laws	10	171,882
ASEAN Cyber Laws (Brunei, Cambodia, Indonesia, Laos, Myanmar, Philippines, Singapore, Thailand, and Vietnam)	15	105,644

[5] It is important to note here that ASEAN keywords were inspected against their occurrences in the Malaysian corpus in order to see which words were deemed highly unusually more frequent in ASEAN, indicating that these words are less frequent in Malaysian legal texts and therefore not universally shared among/highlighted in ASEAN countries.

4. Summary of Key Findings

4.1. Similarities and Differences between the Malaysian Cybersecurity-related Laws and Cyber Laws in other ASEAN Countries

As explained earlier, WordSmith Tool's 'Compare 2 wordlists' feature was used to answer the first research question: wordlists from both corpora were extracted and compared to each other automatically, which revealed frequently occurring words in both corpora as compared to the other[6] and are presented in Table 3. For the purpose of this study, only words that appear no less than 10% of the corpora are included for analysis.

Table 3. Frequently Occurring Words in Both Corpora Using the Compare 2 Wordlist Feature (Normalised Frequencies in Brackets Show Occurrences in the Corpus No Less Than 10%)

ASEAN Cyber Laws Corpus (ACLC)
Nouns/adjectives (35): *electronic* (1.45), *information* (0.73), *computer* (0.43), *law* (0.41), *message** (0.40), *signature* (0.35), *article* (0.32), *E* (0.30), *transactions* (0.30), *record** (0.27), *organisation* (0.25), *system* (0.24), *use* (0.24), *individual* (0.21), *originator* (0.19), *ED* (0.18), *must* (0.17), *documents* (0.15), *security* (0.14), *addressee* (0.14), *legal* (0.14), *records* (0.14), *state** (0.13), *paragraph* (0.12), *business* (0.12), *key* (0.12), *transferable* (0.11), *Singapore* (0.11), *general* (0.11), *signatures* (0.11), *secure** (0.11), *technology* (0.11), *number** (0.10), *transaction* (0.10), *sent* (0.10)
Malaysian Cyber Security-related Laws Corpus (MCSLC)
Nouns/adjectives (33): *he* (0.41), *imprisonment* (0.36), *Malaysia* (0.32), *term* (0.32), *his* (0.32), *fine** (0.30), *extend* (0.28), *punished* (0.26), *whoever* (0.24), *liable* (0.22), *minister* (0.21), *Z* (0.21), *code* (0.20), *licence* (0.17), *cause** (0.17), *committed* (0.16), *copyright* (0.16), *network* (0.16), *work** (0.16), *property* (0.16), *tribunal* (0.14), *commissioner* (0.14), *user* (0.14), *death* (0.14), *also* (0.13), *servant* (0.13), *licensed** (0.11), *ringgit* (0.11), *commits* (0.11), *likely* (0.11), *facilities* (0.10), *thousand* (0.10), *him* (0.10)

Upon first inspection, the Compare 2 Wordlists feature (as shown in Table 3) has revealed that both corpora do not share similar frequent words. In ACLC, words are mostly referring to technology like *electronic*, *information*, *computer*, *message*, [...] *technology* whereas frequently occurring words in MCSLC show words that refer to people or groups of people (*he*, *his*, *whoever*, *minister*, *tribunal*, *commissioner*, *user*, *servant*, *him*). Following Craigen et al.'s (2014) five dominant themes of cybersecurity, it can firstly be argued that for the ASEAN corpus, cyber laws appear more on technological solutions (*information*, *computer*), events (*transactions*, *records*), and strategies, processes, and methods (*system*, *security*). Meanwhile, cybersecurity-related laws in Malaysia seem to revolve around themes like human engagement (*he*, *his*, *whoever*), and referent objects (of security) (*property*, *death*). As Mat et al. (2019) rightfully pointed out, countries do not share the same cybercrime laws so these early findings simply show the differences how ASEAN countries (including Malaysia) frame or describe their national cyber (security) laws and that these laws are subjective in nature. It is interesting to note that for the Malaysian corpus, more attribution is given to human engagement and how that may influence the other frequent words deemed to express strategies, processes, and methods (e.g., *imprisonment*, *fine*, *death*). It was also found that the pronouns *he*, *his* and *him* occurred in MCSLC, but was not so obvious in the ASEAN corpus. Surprisingly, there were more verbs and adverbs in the Malaysian corpus (*extend*, *punished*, *committed*, *also*, *commits*, *likely*) compared to 'use', 'must', and 'sent' that were higher in the ASEAN corpus. While this first observation seems to argue that the former set of laws are more stringent in that words depicting "effects" like punishments can be seen more strikingly, it could be safe to say why ASEAN laws seem less intense.

Although highly frequent words in both corpora pointed to different words related to cybersecurity, some words relating to legal discourse can still be found (e.g., *law*, *legal*, *imprisonment*, *fine*) and these could be referred to the language of legality from the type of register collected. Since these documents

[6] The key word lists can be found in the fourth tab at the bottom of each generated list on the screen.

are related to cyber laws, legal jargon like *originator*, *tribunal*, *paragraph*, *liable* or even letters like *E* and *Z* did not come as a surprise. While this might have been picked up as a difference in stylistic, there is reason to believe in the subjectivity of the way legal documents are written across different ASEAN nations and in turn, not examined further. Also, both corpora showed more noun or adjective-like words compared to verbs (or auxiliaries) such as *use*, *must*, *sent*, *extend*, *punished*, *committed*, *commits*, and adverbs like *also* and *likely*. Where nouns that could also function as verbs – as in, they can be regarded as polysemous – they were given an asterisk (*) until proven contextually.

Ultimately, MCSLC seems to show less technical words related to cybersecurity and technology whereas in ACLC, there is an extant number of words depicting cybersecurity and information technology (e.g., *electronic*, *information*, *computer*, *security*). Through use of the 'Compare 2 Wordlists' feature on WordSmith Tools, we were able to briefly compare and contrast between different words that appear frequently in both corpora. Upon constructing a more comprehensive definition to 'cybersecurity', Craigen et al. (2014) add that "[a]lthough some of these definitions include references to non-technical activities and human interactions, they demonstrate the predominance of the technical perspective within the literature" (p. 15). As a result, ASEAN cyber laws seem to be more technical in terms of cybersecurity than Malaysia's and perhaps this could be mainly due to the mixture of "related" acts that were combined in NASCA's definition of Malaysia's cybersecurity law.

4.2. How Do These Laws Reflect the Dominant Themes of an Interdisciplinary Definition of Cybersecurity?

Upon extracting the keywords from the ASEAN cyber law corpus against the Malaysian one, a total of 104 statistically significant keywords were firstly identified. These are then grouped into the five dominant themes as suggested by Craigen et al. (2014). As far as inter-rater reliability is concerned, only a few words were disagreed upon and reclassified again after meeting a consensus. In fact, several words that could be classified twice were marked with an asterisk (*) as exact meanings could only be distinguished by close readings of the concordance (within context).

Table 4. Keywords Classified by Cybersecurity Dominant Themes (Adopted and Adapted from Craigen et al., 2014)

Theme	ASEAN Cyber Laws Corpus
1. Technological Solutions (5)	**electronic (1535)**, computer (451), system (253), technology (114), software (27)
2. Events (12)	**activities (89)**, contained (49), penalty (58), receiving (36), transaction (104), secure (114), anti (32), sending (56), stored (53), treated (25), retained (39), retention (29)
3. Strategies, processes, and methods (22)	**law (434)**, general (117), documents (155), contract (80), procedure (94), procedures (79), rules (54), acknowledgement (53), certificates (56), appropriate (51), collection (56), validity (47), subsections (33), method (50), storage (62), division (34), generated (50), measures (34), recognition (33), schedule (62), regulation (32), resolution (34)
4. Human engagement (26)	**individual (249)**, originator (197), addressee (152), party (97), their (86), parties (96), duties (74), sent (103), providing (68), provides (57), management (44), employee (40), providers (62), organization (62),* secretary (38), obligation (48), council (33), created (41), agency (91),* follows (49), responsible (33), agencies (77),* association (64),* board (32), organizations (92),* reliability (49)
5. Referent objects (of security) (24)	**information (775)**, law (434), signature (372), article (341), legal (147), state (141), key (122), security (153), business (124), national (100), communication (90), record (288), number (105), identity (36), financial (54), international (74), entity (39), goods (31), integrity (59), commerce (29), record (288), sensitive (29), transactions (320), identification (25)
6. Others (miscellaneous) (15)	use (255), e (320), paragraph (128), about (71), related (83), using (77), applicable (71), can (77), pursuant (40), third (56), ensure (48), must (175), p (50), prior (31), specific (26)

Table 4 presents 104 statistically significant keywords that appear more in ACLC compared to MCSLC. For each of the keywords, we decided to group them into relevant themes as stated by Craigen et al. (2014). In turn, the five dominant themes as they suggest above act as a guide to determine whether linguistic items within legal documents of our corpora describe cybersecurity from a holistic and interdisciplinary perspective. Therefore, for the purpose of this paper, keywords from the ASEAN corpus were examined to reflect the bigger population of countries that have cybersecurity enforcements (and where the 10 cybersecurity-related laws in Malaysia could benefit).

Overall, 26 keywords were grouped under human engagement where words like *individual*, *originator*, *addressee*, *party* and others refer to the actors (doers or ones receiving action) in the discourse. There were 24 referent objects (of security) such as *information*, *message*, *signature*, *article*, *legal* and so on where entities listed in this category describe the type of things related to security. The next frequent type of words involves the theme on "strategies, processes, and methods" (22), which seems to include words that refer to the ways in which organisations plan and take action of the events happening in their respective countries: *general*, *documents*, *contract*, *procedure(s)*, *rules* and so on. The dominant theme of "events" is a refinement of "occurrences" (Craigen et al., 2014, p. 17) and therefore included words like *contained*, *penalty*, *receiving*, *transaction*, *secure*, *sending*, *stored*, *treated*, *retained*, and *retention* (12). Meanwhile, for the dominant theme on "technological solutions", only five keywords were considered and they were *electronic*, *computer*, *system*, *technology*, and *software*. However, there were 15 keywords that we listed as miscellaneous because they do not particularly belong in the earlier dominant themes. These words include *use*, *E*, *paragraph*, *about*, *related*, *using*, *applicable*, *can*, *pursuant*, *third*, *ensure*, *must*, *P*, *prior*, *anti*, and *specific*. To a certain extent, these remaining 15 words could be reflecting the genre/register of legal documents and therefore can be identified as register markers or jargons. It is also important to note that certain words are classified twice where they may function to represent two dominant themes rather than just one (for instance, *organization(s)*, *agency(ies)* and *association* could also be classified as referent objects of security, but are grouped under human engagement).

For the purpose of this study, only the highest frequent word in three selected dominant themes (they are *activities*, *individual*, and *information*) from Craigen et al. (2014) were analysed further in terms of clustering and concordancing. The other two themes (Technological solutions and Strategies, processes and methods) were not chosen as 'electronic' and 'law' appeared to be used quite similarly in both corpora and therefore did not provide much interesting findings.

Under the second dominant theme; 'Events', we chose to further analyse *activities* (89) occurring in 12 out of 15 texts of ACLC, while occurring in 6 out of 10 of the acts in MSCLC. Interestingly, it can be seen in Tables 5 and 6 that the top frequent clusters are strikingly different -activities refer to 'electronic *transaction* activities' in the ASEAN corpus (23 times) compared to 'electronic *government* activities' in the Malaysian one (9 times).

Table 5. Typical Combination Phrase of 'Activities' in ACLC

No.	Cluster	Frequency
1.	electronic transaction activities	23
2.	the activities of	9
3.	of electronic transaction	8
4.	e signature certification	6
5.	the development of	5

Table 6. Typical Combination Phrase of 'Activities' in MCSLC

No.	Cluster	Frequency
1.	electronic government activities	9
2.	the activities of	6
3.	Activities of a	6
4.	Criminal activities of	5

Further inspection on the concordance lines (lines 11-19 in Figure 1) reveals that 'electronic government activities' is referring to a specific act, in particular Act 680: ELECTRONIC GOVERNMENT ACTIVITIES ACT 2007 that is an act:

> "to provide for legal recognition of electronic messages in dealings between the Government and the public, the use of electronic messages to fulfill legal requirements and to enable and facilitate the dealings through the use of electronic means and other matters connected therewith" (Act 680: ELECTRONIC GOVERNMENT ACTIVITIES ACT, 2007).

This implies that for Malaysia, there is a specific act within the cyberlaw that stipulates for the acknowledgement of electronic messages between the Government and the public as legitimate. Meanwhile, for other ASEAN countries, it could be seen from lines 1-10 in the same figure that there is neither specific mention of electronic messages nor government-related ones, rather these could be interpreted as part of 'electronic transaction activities' in general.

1	unit within their villages on the implementation of **electronic transaction activities**; 5. Collect statistics and information on electronic
2	Consider and mediate proposals on **electronic transaction activities** under their jurisdictions; 29 4. Coordinate with relevant units within
3	Implement projects, programs, plans and activities on the development of **electronic transaction activities** from higher levels
4	electronic transactions within their jurisdictions; 8. Regularly summarize and report **electronic transaction activities** to the District
5	dictions to manage **electronic transaction activities**; 7. Collect statistics and information on electronic transactions within their
6	Coordinate with relevant sectors and parties within their jurisdictions to manage **electronic transaction activities**; 7. Collect statistics
7	electronic transactions; 5. Consider and resolve proposals on **electronic transaction activities** under their responsibilities.
8	operations of electronic transactions; 5. Consider and resolve proposals on **electronic transaction activities** under the
9	monitor the Village Units of Science and Technology on the implementation of **electronic transaction activities**; 4. Apply standards and
10	programs and plans with regard to the development of **electronic transaction activities** of higher levels; 28 2. Disseminate laws and
11	IMINARY Short title and commencement 1. (1) This Act may be cited as the **Electronic Government Activities** Act 2007. (2) This Act
12	by whom or on whose behalf, the electronic message is generated or sent; **Electronic Government Activities** 9 addressee means a
13	electronic message; (j) guidelines for the payment and receipt of money; and **Electronic Government Activities** 11 (j) any other matters
14	to the electronic signature after the time of signing is detectable; and **Electronic Government Activities** 13 (c) any alteration made to
15	I continue to apply to any digital signature used as an **electronic** signature in any **Government activities**. Seal 14. (1) Where any law
16	such document shall be in accordance with such specified form. **Electronic Government Activities** 15 Prescribed form 20. Where any law
17	personally, the requirement of the law is fulfilled if the document is **Electronic Government Activities** 17 submitted in accordance with
18	used any agreed procedure, that the electronic message was a duplicate. **Electronic Government Activities** 19 Time of dispatch
19	specified or agreed, within a reasonable time, the originator **Electronic Government Activities** 21 (a) give notice to the addressee stating

Figure 1. Concordance Lines for 'Electronic Transaction Activities' in ACLC and 'Electronic Government Activities' in MSCLC

In terms of the 'Human engagement' theme, we chose to examine *individual* (249 times in ACLC, 120 times in MSCLC) that could represent the doer/agent and/or person, which the act/law acts upon. In ACLC, it was found that the word mostly revolves around personal data of an individual, as Table 7 shows. Interestingly, the typical combination phrase of 'individual' in the MCSLC on the other hand displays a recurring pattern of an "individual licence" (see Table 8) that suggests a license for a

specified person to conduct a specified activity.

Table 7. Typical Combination Phrase of 'Individual' in the ACLC

No.	Cluster	Frequency
1	personal data about	42
2.	about an individual	22
3.	about the individual	21
4.	Frequency	19
5.	data about the individual	19
6.	personal data about the	18
7.	personal data about an	16
8.	data about an individual	16
9.	personal data about an	16
10.	data about an	16
11.	of an individual	14
12.	of the individual	14
13.	of personal data	13
14.	of personal data about	13
15.	without the individual's	11
16.	the individual's consent	11
17.	of personal data about the	11
18.	an individual who	10
19.	the individual's personal	10

Table 8. Typical Combination Phrase of 'Individual' in the MCSLC

No.	Cluster	Frequency
1	an individual licence	45
2.	the individual licence	26
3.	of an individual	21
4.	individual licence granted	9
5.	licence granted under	9
6.	or cancellation of	8
7.	individual licence under	8
8.	of the individual	8
9.	for an individual	8
10.	cancellation of an	8
11.	suspension or	8
12.	granted under this	7
13.	the other individual	6
14.	individual licence and	5
15.	in the individual	5
16.	application for an	5
17.	licence under this	5
18.	individual licence shall	5
19.	individual licence or	5

Upon closer inspection of the concordance lines, "individual licence" occurred 73 times in Act 588: COMMUNICATIONS AND MULTIMEDIA ACT 1998 in MSCLC where this term "means a licence for a specified person to conduct a specified activity and may include conditions to which the conduct of that activity shall be subject" (Act 588: COMMUNICATIONS AND MULTIMEDIA ACT, 1998). More specifically, this act is an act "to provide for and to regulate the converging communications and multimedia industries, and for incidental matters" (https://www.mcmc.gov.my/en/legal/acts/communications-and-multimedia-act-1998-reprint-200). As can be seen in Figure 2, individual licenses

are described in terms of their conditions regulated by the government to allow for users/service providers to conduct activities online (also related to licenses granted under the Telecommunications Act 1950). While this may not be foreign to other ASEAN countries, the collocation "individual license" appeared to be more statistically significant in MSCLC than in ACLC. However, more mention of personal data of the individual/about the individual was seen in ACLC than in MSCLC, perhaps indicating that there are more laws representing individual rights in other ASEAN countries besides Malaysia.

1	which are declared by the Minister under section 13. (4) Notwithstanding subsection (2), all **individual licences** shall be deemed to include the relevant
2	under, or across any land. (6) If the Minister refuses to grant an **individual licence** to an applicant, the Minister shall give the applicant a written notice
3	refuse the application. (8) If the Minister neither grants, nor refuses to grant, an **individual licence** within thirty days from the receipt of the
4	the Commission, the Minister is deemed, at the end of the period, to have refused to grant the **individual licence** unless the applicant receives a written
5	shall be payable upon the approval of the application. Restriction on the grant of an **individual licence** 31. The Minister may not grant an individual
6	additional conditions of the **individual licence** as declared by the Minister and included in the individual licence. Modification, variation or revocation of
7	the Minister and included in the **individual licence**. Modification, variation or revocation of individual licence conditions 33. (1) The Minister may, at any
8	with section 13 (a) modify or vary the special or additional conditions of an existing **individual licence**; (b) revoke the special or additional conditions of
9	(b) revoke the special or additional conditions of an existing **individual licence**; or (c) impose further special or additional conditions to an existing
10	licence; or (c) impose further special or additional conditions to an existing **individual licence**. Communications and Multimedia 43 (2) The procedures

Figure 2. Concordance Lines for 'Individual Licence' in the MCSLC

For the final dominant theme, we chose the word 'information' that occurred 775 in ACLC and 295 times in MSCLC as a salient referent object of security. It could be seen in Table 9 that most occurrences of 'information' include "personal/electronic information" whereas in MSCLC, the word collocates with "processing/rights management" (see Table 10). This shows that the Malaysian legal discourse on 'personal information' is not as frequent compared to in ASEAN cyber laws.

Table 9. Typical Combination Phrase of 'Information' in the ACLC

No.	Cluster	Frequency
1.	the personal information	48
2.	electronic information and	43
3.	electronic information and or	37
4.	information and or electronic	37
5.	information and or	37
6.	and or electronic	37
7.	information and or electronic	36
8.	and or electronic documents	36
9.	or electronic documents	36
10.	of personal information	35
11.	sensitive personal information	33
12.	an information system	32
13.	personal information controller	32
14.	of the information	23
15.	information in the	23
16.	information contained in	22
17.	the information contained	21
18.	processing of personal	19
19.	the personal information	19
20.	personal information and	18
21.	processing of personal information	18

Table 10. Typical Combination Phrase of 'Information' in the MCSLC

No.	Cluster	Frequency
1.	information processing system	14
2.	of the information	12
3.	an information processing	11
4.	to give information	11
5.	information relating to	11
6.	information contained in	10
7.	the information contained	9
8.	information which may	9
9.	that the information	9
10.	contained in the	9
11.	rights management information	8
12.	any information or	7
13.	any information respecting	7
14.	omits to give	7
15.	in the certificate	7
16.	to give any	7
17.	such information as	6
18.	which may lead	6
19.	may lead to	6
20.	any information which	6
21.	give any information	6
22.	information respecting that	6

1	the data subject agrees to the collection and processing of **personal information** about and/or relating to him or her. Consent shall be evidenced by
2	by the data subject to do so. (c) Data subject refers to an individual whose **personal information** is processed. (d) Direct marketing refers to
3	or storage of electronic data, electronic message, or electronic document. (g) **Personal information** refers to any information whether recorded in a
4	with other information would directly and certainly identify an individual. (h) **Personal information** controller refers to a person or organization who
5	a person or organization who controls the collection, holding, processing or use of **personal information**, including a person or organization who
6	another person or organization to collect, hold, process, use, transfer or disclose **personal information** on his or her behalf. The term excludes: (1) A
7	person or organization; and (2) An individual who collects, holds, processes or uses **personal information** in connection with the individual personal,
8	to any natural or juridical person qualified to act as such under this Act to whom a **personal information** controller may outsource the processing of
9	(j) Processing refers to any operation or any set of operations performed upon **personal information** including, but not limited to, the collection
10	of Court and other pertinent laws constitute privileged communication. (l) Sensitive **personal information** refers to personal information: (1) About an

Figure 3. Concordance Lines for 'Personal Information' in the ACLC

1	to prejudice the fair trial of a person; (c) which would involve the unreasonable disclosure of **personal information** about any individual (including a
2	likely to prejudice the fair trial of a person; or (c) involve the unreasonable disclosure of **personal information** about any individual (including a deceased

Figure 4. Concordance Lines for 'Personal Information' in the MSCLC

Figures 3 and 4 show 'personal information' in context between ACLC and MSCLC. Although this collocation appears more frequently in ACLC and concordance lines indicate reference to the sensitivity of personal information, its definition as well as other related details about it (e.g., processing of personal information, personal information controller), in MSCLC, there were only two mentions of 'personal information' and when they do occur, both instances refer to the clause in Act 588 COMMUNICATIONS AND MULTIMEDIA ACT (1998). These two separate lines refer to the issue of publishing: publication of information by a commission or to another party involving "[an unreasonable disclosure of personal information about any individual (including a deceased person)]" (Act 588 COMMUNICATIONS AND MULTIMEDIA ACT, 1998), and in turn, would not safeguard an

individual's personal information on the web necessarily.

5. Discussion and Conclusion

This paper aimed at comparing and contrasting language used in cyber(security) laws from ASEAN countries, including Malaysia using the corpus linguistics approach. We have shown that salient keywords occurring in both corpora show characteristics of a legal discourse that is subjective in describing cybersecurity within the respective countries, and that use of the salient keywords point to (a certain extent) elements of the dominant themes suggested by Craigen et al. (2014). More specifically, initial findings differentiate Malaysia from other ASEAN countries in that the latter identified technological solutions (*information*, *computer*), events (*transactions*, *records*), and strategies, processes and methods (*system*, *security*) as dominant themes in cyber security whereas for Malaysia specifically, themes like human engagement (*he/his*, *whoever*) and referent objects (of security) (*property*, *death*) were more dominant. One explanation for this could be that MCSLC uses less technical words related to cybersecurity compared to ACLC where most cybersecurity-related words were deemed technical, which resonated with Craigen et al.'s (2014) interdisciplinary definition of 'cybersecurity'. Arguably, Malaysia has not come up with a unified single cybersecurity law or act, rather issues related to cybercrime for instance, are arbitrated on the basis of 10 cybersecurity-related acts as mentioned in NACSA.

However, upon close inspection of 104 statistically significant keywords that appeared more in ACLC compared to MCSLC, there were more keywords (26) grouped under human engagement (e.g., *individual*, *originator*, *addressee*, *party*) and referent objects of security (24) such as *information*, *message*, *signature*, *article*, and so on where entities listed in this category described the type of things related to security. It was interesting to find that some distinct clusters occurred in ACLC, but were not found in MSCLC such as 'electronic transaction activities' in the former compared to 'electronic *government* activities' in the latter corpus, thus suggesting that the laws in Malaysia are more government-specific. In terms of human engagement as a theme, there appeared to be a mention of an "individual licence" in MCSLC that was not found in ACLC. Even though permission to own a license may be present within ASEAN cyber laws, it paled in comparison to MCSLC that highlighted the phrase more strikingly, perhaps further emphasizing the gap between governmental and individual laws in the country. Finally, when it came to referent objects of security between the two corpora, it was found that legal discourse on 'personal information' was not as frequent in MSCLC than in ACLC. When it did appear, concordance lines showed that they occurred only twice in the Communications and Multimedia Act, referring to the clause under publications and therefore 'personal information' could be argued as not being the center of cybersecurity-related laws in Malaysia, but merely highlighted in issues related to publications.

All in all, cyber laws depicted in ACLC and MSCLC showed a variety in terms of how legal discourse is framed (linguistically) between ASEAN countries, including Malaysia about how cyber activities are described using language, and in turn, how they are possibly used to protect or penalise actions related to cyber or electronic communication networks. It can be argued that in constructing legal discourse on cyber (security), we first need to have a universal working definition of 'cybersecurity' whose earlier definitions are highly variable, often subjective, and at times, uninformative (Craigen et al., 2014). The call for a more standardised or universal working definition of cybersecurity will facilitate effective laws and policies – addressing specific cyber activities that may require legal actions such as identified in the two corpora, so as not to cause systemic problems in the countries' legislations. Additionally, policymakers will have clearer goals and guidance as they debate and produce new laws related to cybersecurity compared to the drawback of the existing cybersecurity related laws in Malaysia that hinder the efforts to curb the growing prevalence of cyber threats. Future research may

already include 'virtual reality' as a term to be defined operationally as it seems to be the next avenue for legal practitioners to think about, for it has not been listed in any (as far as our knowledge on this topic is concerned) ASEAN (including Malaysia) country's cyber (security) law.

Acknowledgments

This work was supported by the Ministry of Education, Malaysia under the Fundamental Research Grant Scheme for Research Acculturation of Early Career Researchers (FRGS-RACER) [grant number 600-IRMI/FRGS-RACER 5/3 (111/2019)].

References

Adolphs, S., Brown, B., Carter, R., Crawford, P., & Sahota, O. (2004). Applying corpus linguistics in a health care context. *Journal of Applied Linguistics*, *1*(1), 9-28.

Arowosaiye, Y. I. (2013). Evolution of Malaysian cyber laws and mechanism for secured online transactions. *Pandecta Research Law Journal*, *8*(2), 117-126.

Baker, P., Gabrielatos, C., & McEnery, T. (2013). Sketching Muslims: A corpus driven analysis of representations around the word 'Muslim' in the British press 1998–2009. *Applied Linguistics*, *34*(3), 255-278.

Banaji, S., Mejias, S., Kouts, R., Piedade, F., Pavlopoulos, V., Tzankova, I., Mackova, A., & Amnå, E. (2018). Citizenship's tangled web: Associations, gaps and tensions in formulations of European youth active citizenship across disciplines. *European Journal of Developmental Psychology*, *15*(3), 250-269.

Biber, D., Reppen, R., & Friginal, E. (2010). Research in corpus linguistics. In Kaplan, R. B. (Ed.), *The Oxford Handbook of Applied Linguistics* (pp. 548-570). New York, NY: Oxford University Press.

Brenner, S. W., (2010). *Cybercrime: Criminal Threats from Cyberspace*. Santa Barbara: Praeger.

Craigen, D., Diakun-Thibault, N., & Purse, R. (2014). Defining cybersecurity. *Technology Innovation Management Review*, *4*(10), 13-21.

Fløttum, K., Gjesdal, A. M., Gjerstad, Ø., Koteyko, N., & Salway, A. (2014). Representations of the future in English language blogs on climate change. *Global Environmental Change*, *29*, 213-222.

Gabrielatos, C., & Baker, P. (2008). Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK press, 1996-2005. *Journal of English Linguistics*, *36*(1), 5-38.

Ghasiya, P., & Okamura, K. (2022). A hybrid approach to analyze cybersecurity news articles by utilizing information extraction & sentiment analysis methods. *International Journal of Semantic Computing*, *16*(01), 135-160.

Hamin, Z. (2004). The legal response to computer misuse in Malaysia: The computer crimes act 1997. *UiTM Law Review, 2*, 210-234.

Jaspal, R., & Nerlich, B. (2014). When climate science became climate politics: British media representations of climate change in 1988. *Public Understanding of Science*, *23*(2), 122-141.

Kim, D., & Park, J. (2019). Using keyword network analysis. *Journal of Arts Psychotherapy*, *15*(1), 1-20.

Koteyko, N. (2010). Mining the internet for linguistic and social data: An analysis of 'carbon compounds' in web feeds. *Discourse & Society*, *21*(6), 655-674.

Leech, G. (1991). The state of the art in corpus linguistics. In Ajimer, K., & Altenberg, B. (Eds.), *English Corpus Linguistics* (pp. 8-29). London: Longman.

Mat, B., Pero, S., Wahid, R., & Sule, B. (2019). Cybersecurity and digital economy in Malaysia: Trusted

law for customer and enterprise protection. *International Journal of Innovative Technology and Exploring Engineering*, *8*(3), 214-220.

McEnery, T., & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Aractice*. Cambridge University Press.

Mohamed Mizan, N. S., Ma'arif, M. Y., Mohd Satar, N. S., & Shahar, S. M. (2019). CNDS-cybersecurity: issues and challenges in ASEAN countries. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(1.4), 113-119.

Reyes, A., Brittson, R., O'Shea, K., & Steele, J. (2011). The problem at hand. In Reyes, A., Brittson, R., O'Shea, K., & Steele, J. (Eds.), *Cyber Crime Investigations: Bridging the Gaps between Security Professionals, Law Enforcement, and Prosecutors* (pp. 1-22). Massachusetts: Elsevier.

Schatz, D., Bashroush, R., & Wall, J. (2017). Towards a more representative definition of cyber security. *Journal of Digital Forensics, Security and Law*, *12*(2), 52-74.

Skovgaard, J. (2014). EU climate policy after the crisis. *Environmental Politics*, *23*(1), 1-17.

Supayah, G., & Ibrahim, J. (2016). An overview of cyber security in Malaysia. *Kuwait Chapter of Arabian Journal of Business and Management Review, 6*(4), 12-20.

Zukarnain, Z. A., Hashim, M. Z., Muhammad, N., Mansor, F. A., & Azib, W. N. H. W. (2020). Impact of training on cybersecurity awareness. *Gading Journal of Science and Technology, 3*(01), 114-120.

Appendix

Appendix 1. Example of a 'Compare 2 Wordlists' feature on WordSmith Tools

N	Key word	freq. in Malaysian	%	Texts	freq. in ASEAN	Rc. %	BIC	Log_L	Log_R
1	MALAYSIA	558	0.32%	10	1		509.43	521.96	8.42
2	HE	712	0.41%	9	37	0.04%	446.47	459.00	3.56
3	MINISTER	361	0.21%	9	67	0.06%	91.39	103.93	1.73
4	CAUSE	289	0.17%	8	33	0.03%	115.29	127.82	2.43
5	HIMSELF	79	0.05%	8	9		22.46	35.00	2.43
6	NOTWITHSTANDING	86	0.05%	8	13	0.01%	17.99	30.52	2.02
7	CHARGED	49	0.03%	7	5		10.76	23.29	2.59
8	CONTINUE	38	0.02%	7	5		2.62	15.16	2.22
9	COPY	153	0.09%	7	20	0.02%	48.81	61.34	2.23
10	EITHER	89	0.05%	7	19	0.02%	8.97	21.51	1.53
11	FINE	511	0.30%	7	108	0.10%	112.65	125.18	1.54
12	HIM	165	0.10%	7	20	0.02%	57.46	70.00	2.34
13	HIS	548	0.32%	7	66	0.06%	221.02	233.55	2.35
14	IMPRISONMENT	613	0.36%	7	96	0.09%	198.02	210.55	1.97
15	LIABLE	383	0.22%	7	59	0.06%	121.06	133.59	2.00
16	PREMISES	64	0.04%	7	0		48.79	61.33	1,062.61
17	PROCEEDING	50	0.03%	7	5		11.53	24.06	2.62
18	RINGGIT	193	0.11%	7	0		172.40	184.94	1,064.20
19	TERM	557	0.32%	7	69	0.07%	220.05	232.59	2.31
20	THOUSAND	169	0.10%	7	32	0.03%	35.00	47.54	1.70
21	AGONG	26	0.02%	6	0		12.38	24.91	1,061.31
22	APPEAR	25	0.01%	6	1		4.88	17.41	3.94
23	APPEARS	32	0.02%	6	1		11.10	23.63	4.30
24	AUTHORIZED	163	0.09%	6	32	0.03%	31.37	43.90	1.65
25	BELIEVE	105	0.06%	6	14	0.01%	28.92	41.45	2.20

THE AUTHORS

Siti Aeisha Joharry is a senior lecturer in the Department of English Language and Linguistics at Akademi Pengajian Bahasa, Universiti Teknologi MARA, Shah Alam. Her research interests include Corpus Linguistics, Corpus-assisted discourse analysis, and English for Professional Communication.

Syamimi Turiman is a senior lecturer in the Department of English Language and Linguistics at Akademi Pengajian Bahasa, Universiti Teknologi MARA, Shah Alam. Her research interests include Corpus Linguistics, Discourse Analysis, and English for Professional Communication.

Nor Fariza Mohd Nor is an Associate Professor at the Center for Research in Language and Linguistics, UKM. Her areas of research interests are critical discourse analysis, corpus linguistics and digital humanities.

THE AUTHORS' ADDRESSES

First and Corresponding Author
Siti Aeisha Joharry
Lecturer
Department of English Language and Linguistics
Akademi Pengajian Bahasa
Universiti Teknologi MARA
40450 Shah Alam, Selangor, MALAYSIA
E-mail: aeisha@uitm.edu.my

Co-author
Syamimi Turiman
Lecturer
Department of English Language and Linguistics
Akademi Pengajian Bahasa
Universiti Teknologi MARA
40450 Shah Alam, Selangor, MALAYSIA
E-mail: syamimituriman@uitm.edu.my

Co-author
Nor Fariza Mohd Nor
Associate Professor
Center for Research in Language and Linguistics
Pusat Kajian Bahasa & Linguistik
Universiti Kebangsaan Malaysia
43600 Bangi, Selangor, MALAYSIA
E-mail: fariza@ukm.edu.my

Received: 2 November 2022
Received in Revised Form: 30 November 2022
Accepted: 15 December 2022