3L: Language, Linguistics, Literature® The Southeast Asian Journal of English Language Studies
Vol 29(3), September 2023 http://doi.org/10.17576/3L-2023-2903-21

A Corpus-based Analysis of Frequently Occurring Noun Collocations in Geographical Information System (GIS) Research Articles

ANG LENG HONG
School of Humanities
Universiti Sains Malaysia, Malaysia
lenghong@usm.my

ABSTRACT

The importance of collocations in language learning cannot be overstated, particularly with successful academic writing. Previous research on collocation compilation has focused on identifying and compiling lists of general multiword expressions (MWEs) found in academic genres for pedagogical purposes. However, the wide-angle approach employed by these studies may be insufficient for English for Specific Purposes (ESP) courses, given certain formalities and conventions associated with academic texts, which are often demonstrated through MWEs unique to each field or domain of study. This paper presents a corpus-based investigation to develop a list of commonly and pedagogically useful noun collocations specific to the Geographical Information System (GIS). The research used a corpus containing 1.5 million words from GIS research articles. An innovative four-step selection process was used to compile this list. The outcomes of this study are twofold. First, the collocations associated with the 20 most frequently encountered nouns (including noun + noun, noun + verb, verb + noun, and adjective + noun combinations) within the GIS research articles were identified. Second, this investigation presented fresh insights into the different categories of noun collocations. The research provides important pedagogical implications for ESP instruction, particularly on how collocations might be better understood and taught. The findings emphasise the necessity of focusing on domain-specific collocations, underscoring the role of context in language learning and teaching. The results are expected to help shape effective instructional strategies and material development in ESP classrooms, particularly in GIS-related contexts.

Keywords: domain-specific; noun; collocation; academic writing; corpus-based

INTRODUCTION

Collocations, fundamental to language fluency and mastery, are pillars of linguistic proficiency. Their central role in language acquisition is undeniable, with proficiency in their use being a hallmark of linguistic adeptness (Lei & Liu, 2018; Nizonkiza & Van de Poel, 2019). This proficiency directly influences the accuracy and fluency of language output and comprehension (Basal, 2019; Durrant, 2014). However, despite their significance, mastering collocations remains a daunting task for many, even among advanced learners. This challenge often arises from the inherent complexity of collocations, interference from one's native language, and the nuances introduced by certain teaching methodologies (Nesselhauf, 2003).

Collocations, often semantically transparent (like 'serve the purpose', 'dark night', and 'make a decision'), seem straightforward due to their individual word relationships (Durrant, 2014). As a result, learners and teachers may overlook collocation errors. Moreover, learners may gravitate towards collocations that align with their first language (L1), as seen in the Malaysian context, where native speakers of the Malay language might perform a literal translation from their mother tongue (Ang et al., 2011). For example, in English, 'bake a cake' is '*buat satu kek*' (make a cake) in Malay, and 'home country' is '*negara asal*' (original country). Learners' reliance on L1 transfer may suggest that they express themselves by constructing the expressions from individual

words rather than from semi-prefabricated units (Lei & Liu, 2018; Wray, 2002). Such tendencies imply their struggle to master collocation knowledge.

The issue is compounded by inadequate explicit teaching and limited exposure to commonly used collocations in non-native language learning contexts. While some believe that phraseology should be implicitly taught, it raises concerns if learners lack appropriate instruction and input on collocations. According to Wray (2002), learners are likely to view and learn words independently, ignoring the concept of word chunks. Moreover, even with sufficient input, choosing specific collocations to focus on remains challenging for learners (Henriksen, 2013).

To address these difficulties in English for Specific Purposes (ESP) settings, researchers have compiled lists of collocations and multiword expressions (MWEs), for instance, the cross-disciplinary collocation lists from seven academic subject areas (Durrant, 2009), general academic written English MWEs (Liu, 2012), academic collocation lists (Ackermann & Chen, 2013), and academic English collocation lists (Lei & Liu, 2018). Despite appreciating the usefulness of general academic collocation lists in aiding ESP students, the wide-angle and cross-disciplinary approach to identifying academic collocations does not meet writing requirements in a specific discipline.

The research underscores MWEs exhibit distinct variations across disciplines (Hyland, 2008). Such variations are particularly evident in scientific and academic writing, where the language is frequently laden with specialised and technical collocations (Durrant, 2014). This trend towards discipline-specific phrasings arises from the constant evolution and advancement in fields of study. As new knowledge is generated, it necessitates the creation and acceptance of novel phrases to describe emerging concepts and methodologies. Consequently, the inherent diversity and specificity of these collocations across academic fields or domains underscore the importance of domain-specific resources. Generic lists of academic collocations may fall short of addressing the nuanced linguistic requirements of each discipline or domain. Therefore, developing discipline or domain-specific academic collocation lists becomes pivotal to ensure effective and accurate academic writing. These tailored lists would give writers a refined toolset, enabling them to convey complex ideas within their respective domains precisely.

Nouns, the most frequently occurring lexical word category in English, are often found in collocations in academic writing (Biber et al., 1999; Halliday, 1993; Nizonkiza & Van de Poel, 2019). Due to their higher prevalence in academic texts than other registers, nouns and their collocates are significant in academic writing (Biber et al., 1999). Consequently, teaching collocations should place emphasis on nouns and their collocates (Lei & Liu, 2018; Nizonkiza & Van de Poel, 2019). Nouns and their collocates, which form collocations, are essential elements of academic writing (Biber et al., 1999), contributing to the conciseness demanded in scholarly works (Parkinson, 2015). However, novice writers and learners often struggle with effectively using nouns and noun constructions, particularly in scientific writing, due to their lexical density and genericity (Halliday, 1993; Liu & Lu, 2020).

To advocate a domain-specific approach in collocation studies, this paper focuses on identifying and compiling noun collocations associated with the concept and fundamental attributes of the Geographical Information System (GIS) language, a key element in understanding, researching, and representing geographical data. GIS language, influenced by changes in geographical language, can be seen as a comprehensive and digital symbol system that accurately depicts geographical entities, their spatial dispersion, and dynamic processes (Hu et al., 2014). It aids in creating a universal understanding of geographical space by illustrating geographical

scenarios or symbols with specific geometric, status, processes, and temporal-spatial relationships, among other features (Miller & Wentz, 2003).

In domains like GIS, students and emerging writers are required to produce a broad array of written genres indicative of the diverse contexts in which they operate (e.g., private sector, academic institutions) and the various audiences they communicate with (e.g., research colleagues, the public, governmental bodies). This highlights the significance of ESP writing in GIS. Thus, this study aims to examine and compile the most common and pedagogically useful noun collocations in GIS research articles, which can serve as a beneficial reference and instructional resource in the relevant ESP setting.

LITERATURE REVIEW

DEFINITIONS AND APPROACHES TO COLLOCATION

Defining 'collocation' is a crucial task in collocation research, given the differing interpretations of the term across various approaches. The two main perspectives are the frequency-based and phraseological approaches. The frequency-based approach, supported by scholars such as Sinclair (1987), views collocation as a word's association with others appearing more frequently than by chance in its context. The higher the occurrence probability, the stronger the collocation. This approach employs statistical formulae and uses tools like *AntConc* (Anthony, 2020), *Wordsmith Tools* (Scott, 2020), and *Lancsbox* (Brezina et al., 2020) to identify collocations. Terms like 'collocate', 'node' and 'span' are used to pinpoint the focus word (node) and the adjacent words (collocates) that form potential collocations. The frequency-based approach emphasises the tendencies of words to co-occur, distinguishing between common and significant associations, where the significance is measured by the exclusivity of the relationship between the node and the collocates (Brezina et al., 2020). This exclusivity is gauged using statistical measures like mutual information, log-likelihood, or logDice values.

On the other hand, the phraseological approach classifies word combinations based on their degrees of fixedness, ranging from fixed and semantically opaque combinations like idioms to partially fixed ones like collocations and eventually to free combinations (Cowie, 1998). Here, collocation is seen as a word combination distinct from idioms and free combinations, with components that are transparent but also "characterised by arbitrary limitations of choice" (Gyllstad, 2007, p. 11). Advocates of this approach tend to dismiss frequency and statistics as decisive factors in identifying strong collocations.

However, both approaches have limitations. The phraseological approach struggles with categorising different word combinations, while the frequency-based approach, which relies solely on frequency, may extract frequent but less meaningful combinations. Some scholars, such as Kjellmer (1987), Stubbs (1995), and Nesselhauf (2003), have proposed integrating both approaches. Kjellmer (1987, p.133) defined collocation as a "sequence of words that occurs more than once in identical form and which is grammatically well-structured." Stubbs (1995) used frequency to identify word combinations with grammatical relations, while Nesselhauf (2003) used frequency to complement her analysis of learner corpora, proposing three categories of word combinations: free combinations, restricted collocations and idioms.

In this study, given its relevance, Kjellmer's (1987) definition of collocation was adopted, and frequent noun collocations, i.e., noun + noun, adjective + noun, verb + noun, and noun + verb collocations in GIS research articles were identified, extracted and compiled following the relevant procedures.

COLLOCATION LIST COMPILATION

Recent years have seen efforts towards creating useful academic MWE lists, such as Liu's (2012) compilation from the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). In general academic writing, Liu (2012) identified frequent MWEs like idioms, lexical bundles, and phrasal verbs. However, his methodology relied on pre-established lists of potential MWEs, which might have overlooked less frequent but valuable MWEs. Ackermann and Chen (2013) focused on various collocations, employing statistics, part-of-speech tagging, manual review, and expert opinions to develop the Academic Collocation List (ACL). Despite its value, the ACL might overlook some useful collocations due to its ± 3 extraction method, and it faced criticism over the extensive use of human judgement that removed a considerable number of English-specific collocations, as pointed out by Lei and Liu (2018). This removal might have limited the ACL's potential as an idiomatic English collocation resource crucial for language proficiency (Nesselhauf, 2003).

Lei and Liu (2018) went a step further, incorporating syntactic dependency relation analysis in creating the academic English collocation list (AECL) using the Stanford CoreNLP program (Toutanova et al., 2003). This list minimised human judgement to reduce the possibility of excluding qualified collocations. Despite its broad interdisciplinary scope, the AECL may not fully address the needs of learners requiring discipline-specific collocational knowledge. Lei and Liu (2018) stressed the importance of discipline or domain-specific academic English collocation lists. This stance aligns with the findings of Hyland (2008) and Reppen and Olson (2020), who both noted the variability of phraseological use across distinct disciplines or domains. This variability underscores that genres and disciplines often convey meaning uniquely through lexicogrammar and are influenced heavily by discipline-specific vocabulary. Supporting this perspective, Durrant (2014) highlighted that academic discourse is not a monolithic entity but varies significantly across disciplines in its lexical and grammatical structures. Moreover, Crosthwaite et al. (2017) found that novice writers often adopt a generalised writing style. They theorised that this tendency might arise from a deficiency in the collocational knowledge essential for discipline-specific discourse, emphasising the need for more targeted lexical instruction in academic writing courses.

GEOGRAPHICAL INFORMATION SYSTEM (GIS)

A Geographical Information System (GIS) is a powerful tool that allows users to capture, store, manipulate, analyse, manage, and present different geographical or spatial data (Miller & Wentz, 2003). Fundamentally, GIS bridges the gap between what we see in the real world and how we can visually interpret and understand it digitally. It incorporates spatially referenced data to understand patterns, relationships, and trends in landscapes, demographics, environmental change, and much more (Hu et al., 2014). GIS is widely applicable, from urban planning to environmental conservation, public health, and business analytics.

Linguistically, GIS language pertains to the terminology, concepts, and expressions associated with the domain of GIS. When discussing GIS language, it revolves around the unique terms vital for understanding and working within the GIS framework. Central to GIS language includes datasets superimposed cohesively to form a comprehensive map or model. The datasets can be of various types, like vectors (points, lines, and polygons) or raster (grid-based, like satellite imagery). Examples of MWEs related to GIS language include 'spatial analysis', which refers to the process of examining the locations, attributes, and relationships of features in spatial data, 'layer stacking', denoting the overlaying of multiple data layers to create a composite image or map, and 'attribute table', which is a spreadsheet or database linked to spatial features, providing more information about each feature in the map. Through these systems and terminologies, GIS offers a dynamic way to interpret and model the world around us (Hu et al., 2014).

Despite the increasing significance of GIS in various professional domains, there remains a noticeable gap in comprehensive studies focusing on GIS-specific language. In particular, understanding MWEs, which often encapsulate discipline-specific nuances, is paramount for clarity, consistency, and effective communication within the field. This linguistic aspect becomes even more crucial given the interdisciplinary nature of GIS, which sees its application ranging from urban planning and environmental science to healthcare and transportation, among others. Each domain might subtly or significantly shape how GIS language is employed. Within this context, noun collocations—combinations of nouns or nouns with other words that frequently co-occur—emerge as particularly telling indicators of the discourse's evolving nature. These collocations can serve as gateways to deeper thematic insights, revealing prevalent trends, research foci, and the general direction of academic discourse. Addressing this oversight, the current study seeks to identify the most commonly used and pedagogically useful noun collocations within academic articles related to GIS. By doing so, it aims not only to enrich the understanding of GIS-specific lexical patterns but also to provide a linguistic foundation that can inform and guide both novice researchers and seasoned practitioners in the field. Therefore, in this study, the following question is explored:

What are the most frequent noun collocations in GIS research articles?

METHODOLOGY

THE CORPUS

This study exclusively focuses on research articles due to their pivotal role in academic engagement. Their central position in academic writing (Durrant, 2009) makes them a rich source for identifying valuable academic collocations. Research articles are the primary forums for presenting and debating new knowledge (Groom, 2007), and they represent a standard of good writing that learners strive to achieve (Hyland, 2008). In line with the principles of corpus development, which emphasises representativeness (Sinclair, 2004), four top-ranked GIS journals in the Geography domain were selected. These journals, chosen based on the Journal Citation Reports by the Web of Science group, were further confirmed for representativeness by consulting two professors in GIS.

To reflect recent trends in GIS writing, 312 highly cited empirical research articles published between 2016 and 2021 in these four journals were selected. All selected papers were converted from PDF to text files. Then, a data cleaning process was carried out to remove irrelevant

information, such as tables, figures, numerical data, footnotes, and bibliographical references, to avoid skewing the analysis results. This process left us with the actual contents of the research articles, forming a corpus of 1,500,276 words.

THE PROCEDURE FOR COLLOCATION EXTRACTION

THE MOST COMMON NOUNS AS NODE WORDS FOR EXTRACTING COLLOCATIONS

The collocation extraction process was divided into four steps: (i) selection of the most frequent nouns, (ii) identification and extraction of noun collocation patterns, (iii) validation of collocations, and (iv) manual checking. These steps are further explained in the following paragraphs.

In this corpus-based study, I employed the online corpus analysis tool and resource *Sketch Engine* developed by Kilgarriff et al. (2014) to generate and extract the relevant corpus data. The initial step was to select a list of node words, all of which were nouns, to serve as entries for potential collocations (Ackermann & Chen, 2013).

This study aims to establish a list of academic noun collocations that are prevalent and useful in GIS. To achieve this, the *Word List* function in Sketch Engine was used to select the top 20 nouns as node words for collocation extraction. Table 1 shows the top 20 nouns chosen for this research. These top 20 nouns comprise 55,312 words, or 3.7% of the total corpus.

TABLE 1. Top 20 nouns in GIS research articles corpus

	Top 20 nouns in alphabetical order	Frequency
1	Area	6502
2	Change	5896
3	Density	5232
4	Distribution	4821
5	Image	4622
6	Information	3652
7	Land	3522
8	Line	3367
9	Location	2811
10	Map	2312
11	Model	1751
12	Movement	1622
13	Network	1302
14	Point	1258
15	Population	1159
16	Road	1136
17	Sensor	1121
18	Space	1110
19	Technology	1066
20	Visualisation	1050

DETERMINING AND EXTRACTING NOUN COLLOCATION PATTERNS

The next step in this study involved identifying and extracting potential noun collocations. I focused on four noun collocation patterns: adjective + noun, noun + noun, verb + noun, and noun + verb. Biber et al.'s (1999) classification of noun pre-modifiers was adopted in this study to classify the pre-modifiers of the head nouns. According to Biber et al. (1999), there are four categories of noun pre-modification: general adjectives (e.g., 'thematic map', 'temporal distribution'), *ed*-participial modifiers (e.g., 'undeveloped land', 'observed location'), *ing*-

participial modifiers (e.g., 'low-lying land', 'existing sensor'), and noun modifiers (e.g., 'location point", 'sensor network').

Each noun from our top 20 noun list was used as the node word in extracting the relevant MWEs. Given that collocations were defined as statistically recurring sequences of grammatically correct items (Kjellmer, 1987), lemma search and Corpus Query Language (CQL) were used to extract the four types of MWEs that were statistically significant in the corpus. Using CQL and lemma search enabled the extraction of word combinations with both singular and plural forms of the nouns. MWEs that appeared at least ten times per million words with a mutual information (MI) score above 3.0 were extracted using *Sketch Engine*'s *Concordance* tool.

In the case of adjective + noun and noun + noun patterns, the adjective in the former pattern and the first noun in the latter were treated as pre-modifiers in the respective MWEs. The head nouns were subsequently listed as entries in the collocation list. Building on Lei and Liu's (2018) approach, for verb + noun patterns, sequences such as verb + noun, verb + article + noun, verb + adjective + noun, and verb + article + adjective + noun were extracted and classified under verb + noun collocations. An example would be 'develop network', 'develop [the] network', 'develop [spatial] network', and 'develop [the] [spatial] network'. The collocation can be found under the noun entry 'network'. Similarly, for noun + verb pattern, a lemma search was performed to extract both singular and plural forms of the nouns and all variants of the verbs, including simple present and past tense verbs. The list of collocations, presented in lemma form, can be found in the Appendix.

VALIDATION OF THE NOUN COLLOCATIONS

The next step involved data validation. This step was to ascertain whether the noun collocations extracted were specific to GIS writing. The standardised frequency of each noun collocation was compared to its counterpart in the British Academic Written English Corpus (BAWE), an academic English corpus comprised of about 6.5 million words spanning 30 different disciplines. The standardised frequency was set as per million words. In line with Cunningham's (2017) methodology, noun collocations whose normalised frequencies were higher in the BAWE reference corpus were removed. The Symmetric Mean Absolute Percentage Error (sMAPE) was utilised to compare these frequencies. This method calculates "the difference of the two values over the average of the two values" (Cunningham, 2017, p. 75). In practical terms, this equation is represented as (GIS - BAWE)/[(GIS + BAWE)/2]. The maximum sMAPE value is 2, assuming the noun collocation frequency in the BAWE reference corpus is not zero. In this study, I used a threshold sMAPE value of 1.95 to choose only those noun collocations that occurred about 100 times more often in the GIS research articles corpus than in BAWE.

MANUAL CHECKING

The final stage of collocation extraction was a manual inspection. Following the validation study, a thorough examination of all noun collocations that met the threshold sMAPE value of 1.95 was conducted to ensure that the noun structures extracted were indeed the types of collocations relevant to the study. For example, the noun + noun collocations were manually checked to ensure they belong to the same syntactic group and if the first noun acts as a modifier to the second noun. This manual checking differs from the human judgement used by Ackermann and Chen (2013). Its primary objective was to confirm the syntactic relationship between the words within the noun

combination structures and reduce errors in automatic part-of-speech tagging. The analysis results and a discussion of the noun collocation list in GIS are presented in the following section.

RESULTS AND DISCUSSION

Table 2 provides detailed information about the noun collocations in GIS, including the patterns of these collocations, their corresponding type and token frequencies, and individual percentages of the total list. The classification of noun pre-modification, as proposed by Biber et al. (1999), is only relatable to studies that classify the noun pre-modifier in noun + noun collocations as a noun rather than an adjective, such as the works of Ackermann and Chen (2013), and Lei and Liu (2018). Conversely, some studies, like Nizonkiza and Van de Poel (2019), referenced the Oxford Collocations Dictionary to classify the noun pre-modifier as an adjective and categorise combinations like 'investment/family business' as adjective + noun collocations. Different classifications of noun pre-modification types may lead to varying, or even conflicting, results. Therefore, it is essential to specify the classification scheme applied in studies of a similar nature.

TABLE 2. Distribution of noun collocations by type and token

	Type	Percentage	Token	Percentage
noun + noun	350	32	4810	32
noun + verb	289	26	3105	21
verb + noun	250	23	3677	25
adjective + noun	211	19	3301	22
	1100	100	14893	100

The research findings presented above indicate that the combination of noun + noun and adjective + noun configurations constitute more than half, precisely 51%, of the total noun collocation types found. These combinations play a pivotal role in constructing precise academic prose, hence substantiating the claim made by Biber et al. (1999) that collocations are the most commonly found type of noun construction in academic written English.

Of these collocations, a substantial majority, about 32%, involve a combination of two nouns. For example, 'image rectification', 'information visualisation', and 'line density' are commonly found phrases. This pattern effectively encapsulates complex ideas prevalent in GIS academic discourse into succinct, clear phrases. The first noun, or the pre-modifier, serves a vital purpose. It provides specific information about the second noun or the head noun, defining or determining its purpose, identity, source, or content. This form of construction allows for detailed conceptualisation of GIS concepts or notions. This mechanism is, thus, very useful in GIS academic writing where there is a need for both precision and brevity.

On the other hand, the adjective + noun pattern, while less common at 19% prevalence, is nonetheless significant. Instances like 'residential location', 'digital model', and 'focal point' exemplify this pattern. This type of collocation typically uses an adjective to modify and specify the meaning of the noun, adding an extra layer of detail or context in GIS writing. It enables the author to deliver more nuanced and detailed information about the head noun, which enhances the clarity of the argument or point being made. Although less common than the noun + noun construct, adjective + noun collocations are essential in forming cogent and succinct academic prose in GIS.

The findings above present a contrasting view to the previously discussed research on noun collocations. In the studies by Ackermann and Chen (2013) and Lei and Liu (2018), adjective + noun combinations were more prevalent than noun + noun combinations, differing significantly from Biber et al.'s assertion. These alternative research findings further enrich our understanding of the use and distribution of collocations in academic writing. In Ackermann and Chen's (2013) study, which used both corpus-driven and human-judged methodologies to establish a cross-disciplinary collocation list, only a minor fraction, 2.5% of the collocations were noun + noun combinations such as 'background knowledge', 'data set', and 'assessment process'. In stark contrast, 71.8% were adjective + noun combinations, which included general collocations like 'small proportion', 'significant impact', and 'qualitative analysis'. This prevalence of adjective + noun combinations suggests that these collocation types may be more versatile across different academic disciplines, serving to qualify or specify the noun in diverse contexts effectively. Lei and Liu's (2018) study corroborated this pattern to a certain extent. In their self-created general academic corpus, noun + noun combinations represented a slightly more significant fraction of 6.5%, with examples like 'case study', 'leisure time', and 'study period'. However, almost half, 49% of all collocations, were still comprised of adjective + noun combinations such as 'important aspect', 'recent study', and 'experimental group'. Moreover, an intriguing discovery was that there was no overlap between the collocations found in the GIS corpus and those from Ackermann and Chen's (2013) and Lei and Liu's (2018) studies. This divergence implies that domain-specific collocations might not be detected in cross-disciplinary collocation studies, suggesting the need to consider the impact of disciplinary language norms when examining academic collocations. These divergent results highlight that while certain forms of collocations might be prevalent in specific academic domains, their distribution can vary widely across different disciplines. Hence, it reinforces the necessity to understand the conventions of language usage in particular fields of study when examining or employing collocations.

Noun + verb combinations are the second most common, representing 26% of the collocations examined. This understanding of collocations benefits English for Specific Purposes (ESP) students, especially those focused on GIS topics, as they frequently need to discuss specific entities or processes within this discipline. In these noun + verb collocations, the noun typically represents an entity or concept within the GIS field, while the verb usually specifies the action or process related to the noun. For instance, in the phrase 'movement facilitates', 'movement' is a GIS entity, and 'facilitates' is its associated action. Similarly, in the combination '[The] map produces', 'map' is the entity, and 'produces' describes the action. Such combinations are common in GIS-related discourse, for instance, '[The] movement facilitates/affects/enhances/increases' and '[The] map produces/creates/exhibits/generates'. These combinations are integral to expressing complex relationships or processes in the GIS field concisely and accurately. The significance of this pattern becomes more evident when considering the practical applications for ESP students. For instance, when ESP students need to articulate changes in GIS-related processes or specific entities, their familiarity with these collocations can be highly beneficial. By employing these noun + verb combinations, students can precisely and succinctly describe complex GIS concepts or processes, aiding in clearly and effectively communicating ideas.

Verb + noun collocations form the third largest category of collocations, accounting for 23% of all noun collocation patterns identified in the study. This pattern is also prevalent in the research conducted by Ackermann and Chen (2013) and Lei and Liu (2018), ranking as the second most common type. Examples of this kind of collocation within the GIS context include combinations such as 'urbanise [the] area', 'estimate [the] population', and 'deploy [the] sensor'.

These combinations essentially reflect an action (verb) performed on an object (noun), effectively communicating processes, methods, or actions within a specific academic or professional context. In addition, this pattern has been the subject of many studies in learner writing and collocation learning, as highlighted in the work of Basal (2019) and Nesselhauf (2003). Interestingly, Lei and Liu (2018) noted that this type of collocation frequently presents the highest number of individual pairings that language learners find more challenging to grasp than other collocation types. This observation underscores the complexity and importance of the verb + noun pattern in academic language learning and usage. However, the current study also stresses the significance of studying other collocation patterns. Indeed, noun-based combinations such as noun + noun collocations, despite being somewhat overshadowed by verb + noun collocations in some studies, have proven to be prevalent in academic registers, as Biber et al. (1999) and Nizonkiza and Van de Poel (2019) argued. The findings of this study echo these arguments, demonstrating that noun + noun combinations are also frequently encountered in academic writing, particularly within the GIS discipline.

The findings of this study point to the significant role that noun collocations play in academic writing, especially within specialised domains like GIS. Research has shown, as highlighted by Durrant (2009) and others, that while mastering these collocations can be difficult for language learners, it is critical to their capacity to properly convey complicated concepts and engage in discourse within their subject of study. The primary objective of this study, which is to identify and compile the most frequently used noun collocations, is particularly beneficial for ESP learners. This is because noun collocations can offer concise and precise expressions for encapsulating complex concepts, making them valuable tools for academic writing, for instance, the node word, 'model', a term often used in GIS-related discourse. ESP learners can benefit greatly from understanding the various adjectival modifiers that can describe different types of 'models'. Examples include 'spatial model', 'digital model', 'linear model', 'statistical model', and 'dynamic model'. Each combination conveys a specific type of model, allowing for precise and succinct expression of complex concepts.

The findings from various studies, including those by Durrant (2014) and Hyland (2008), highlight the considerable variation in vocabulary and phrase structures across different genres, registers, and academic disciplines. This variation is particularly pronounced in noun collocations, where certain combinations are more prevalent in certain academic fields than others. Taking the example of the node word 'population', we see a variety of combinations such as 'population flow', 'population density', 'population count', 'population size', 'population growth', 'population distribution', and 'population ratio'. These noun + noun collocations are markedly more common in the GIS corpus than in the BAWE corpus. This pattern exemplifies the specialised language use within academic disciplines and underscores the value of specialised collocation lists for different fields. This disciplinary preference for certain noun collocations aligns with Hoey's (2005) lexical priming theory, which suggests that language users are 'primed' to use certain collocations more frequently in specific contexts. According to this theory, our prior experiences with language subtly influence our lexical choices and sentence structures over time, a phenomenon evident in academic writing. For instance, certain noun collocations might be 'primed' for frequent use in GIS-related discourse due to their relevance and commonality within the field, validated through sMAPE. Such priming of collocations in specific contexts further underscores the importance of adopting a discipline-specific approach when compiling collocation lists for language instruction, as Hyland (2008) advocated. This ensures the vocabulary lists align closely with the specialised language used in various academic disciplines, enhancing their effectiveness in ESP teaching. By

tailoring vocabulary instruction to meet the unique linguistic requirements of each discipline or domain, ESP learners might acquire the necessary language skills for effective communication within their fields.

CONCLUSION

This study advocates for the need to move beyond a broad, one-size-fits-all approach to the pedagogical listings of collocations. As argued by Hyland (2008), expecting a single vocabulary inventory to serve all students across various academic disciplines is unrealistic. In response, the paper adopts a discipline-specific approach for a more valuable and practical collocation profiling, focusing specifically on noun collocation patterns in GIS academic writing. This research's innovative approach lies in its methodology, which combines statistical analyses, CQL, and sMAPE test to identify noun collocations unique to academic writing in GIS. This study emphasises the need to expose ESP learners to these frequently used but discipline-specific collocations. Given that noun collocations can have various types of collocates, they serve as valid learning targets. Therefore, a reference tool documenting noun collocations' prevalence and patterns could be valuable for ESP learners and novice writers. This tool could serve as a helpful guide when uncertainty arises regarding the appropriateness of specific noun collocations. Teaching strategies could include providing ESP learners with authentic concordance exercises featuring selected noun collocations. As Byrd and Coxhead (2010, p. 56) pointed out, this can allow students to gain adequate exposure to common noun collocations through "multiple focused encounters in context and the classroom". Showcasing examples of noun collocations from real-world texts helps underscore their centrality to language use and the potential for enhanced fluidity and native-like quality in writing when used effectively.

Besides exploring noun collocations using concordance data, instructors could employ more specific teaching methods. Examples include using vocabulary notebooks and a "class vocabulary box," as Byrd and Coxhead (2010, p. 57) suggested. This approach can foster active and continuous learning, with students regularly recording, reviewing, and recalling key collocations. Moreover, the GIS collocation list could be employed as testing material to assess learners' use of noun collocations and their vocabulary knowledge (Lei & Liu, 2018). Regular testing using these materials can highlight areas for improvement and inform instruction, ensuring that learners gradually progress in their understanding and usage of these crucial language elements. Lastly, digital tools like online corpora can be integrated into teaching to enhance students' familiarity with collocations further. Interactive activities such as collocation quizzes, word-matching games, or sentence creation tasks using specific collocations can make learning engaging and memorable. This approach to teaching collocations emphasises active learning, deep understanding, and the practical application of knowledge in real-world contexts.

While this research provides valuable insights into noun collocations within GIS academic writing, it has certain limitations that must be acknowledged. First, the focus of the investigation was solely on the most prevalent nouns and their corresponding collocates, forming noun collocations. The study did not delve into other types of collocations, such as those involving adverbs, which can also hold considerable pedagogical importance. Expanding the focus to include these could enrich our understanding of the complexities of collocation use within the discipline. The second limitation is that the research only concentrated on academic articles, potentially excluding insights from other genres that might also be relevant to GIS. A wider scope

encompassing diverse genres could provide a more comprehensive view of the subject matter. Despite these limitations, the study contributes to understanding noun collocations within GIS. It opens the path for more extensive research in the area, encouraging future investigations to broaden their scope to other genres and academic disciplines.

Moving forward, the subsequent research will strive to delve deeper into the realm of noun collocations in GIS by adopting a part-genre approach. This involves scrutinising and contrasting noun collocations across various sections of GIS research articles, which can provide a more granular understanding of collocation use within specific contexts. A more detailed and complex comprehension of noun collocation usage in academic writing is hoped to be cultivated. This will, in turn, contribute to more effective and targeted language instruction in ESP and GIS.

ACKNOWLEDGEMENTS

The author wishes to sincerely thank the reviewers and the editor for their insightful comments and valuable suggestions.

REFERENCES

Ackermann, K., & Chen, Y. H. (2013). Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, *12*(4), 235–247. https://doi.org/10.1016/j.jeap.2013.08.002

Ang, L. H., Rahim, H. A., Hua, T., & Salehuddin, K. (2011). Collocations in Malaysian English learners' writing: A corpus-based error analysis. *3L: The Southeast Asian Journal of English Language Studies*, *24*(2), 82–94.

Anthony, L. (2020). *AntConc Version 3.5.9*. [Software]. Tokyo, Japan: Waseda University. Available at https://www.laurenceanthony.net/software

Basal, A. (2019). Learning collocations: Effects of online tools on teaching English adjective-noun collocations. *British Journal of Educational Technology*, *50*(1), 342–356. https://doi.org/10.1111/bjet.12562

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). *LancsBox Version 5*. [software]. Available at http://corpora.lancs.ac.uk/lancsbox

Byrd, P., & Coxhead, A. (2010). 'On the other hand': Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, 5, 31–64.

Cowie, A. P. (1998). *Phraseology: theory, analysis, and applications*. Oxford: Oxford University Press.

Crosthwaite, P., Cheung, L., & Jiang, F. K. (2017). Writing with attitude: Stance expression in learner and professional dentistry research reports. *English for Specific Purposes, 46*, 107–123. https://doi.org/10.1016/j.esp.2017.02.001

Cunningham, K. J. (2017). A phraseological exploration of recent mathematics research articles through key phrase frames. *Journal of English for Academic Purposes, 25*, 71–83. http://dx.doi.org/10.1016/j.jeap.2016.11.005

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for Academic Purposes. *English for Specific Purposes, 28*(3), 157–169. https://doi.org/10.1016/j.esp.2009.02.002

Durrant, P. (2014). Discipline and level specificity in university students' written vocabulary. *Applied Linguistics*, *35*(3), 328–356. https://doi.org/10.1093/applin/amt016

Groom, N. W. (2007). *Phraseology and epistemology in humanities writing: a corpus-driven study*. Unpublished PhD thesis, University of Birmingham.

Gyllstad, H. (2007). *Testing English collocations: developing receptive tests for use with advanced Swedish learners*. Unpublished PhD thesis, Lund University.

Halliday, M. A. K. (1993). *Writing science literacy and discursive power*. London: Falmer Press.

Henriksen, B. (2013). Research on L2 learners' collocational competence and development – a progress report. *Eurosla Monographs Series*, *2*, 29–56.

Hoey, M. (2005). *Lexical priming: a new theory of words and language*. London: Routledge.

Hu, Z., Tang, G., & Lu, G. (2014). A new geographical language: A perspective of GIS. *Journal of Geographical Sciences*, *24*, 560–576. https://doi.org/10.1007/s11442-014-1106-2

Hyland, K. (2008). As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4–21. https://doi.org/10.1016/j.esp.2007.06.001

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, *1*, 7–36. https://doi.org/10.1007/s40607-014-0009-9

Kjellmer, G. (1987). Aspects of English collocations. In W. Meijs (Ed.), *Corpus linguistics and beyond* (pp. 133–140). Amsterdam: Rodopi.

Lei, L., & Liu, D. (2018). The academic English collocation list: A corpus-driven study. *International Journal of Corpus Linguistics*, *23*(2), 216–243. https://doi.org/10.1075/ijcl.16135.lei

Liu, D. (2012). The most frequently-used multiword constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, *31*(1), 25–35. https://doi.org/10.1016/j.esp.2011.07.002

Liu, Y., & Lu, X. (2020). Chinese EFL learners' misconceptions of noun countability and article use. *System*, *90*, 102222. https://doi.org/10.1016/j.system.2020.102222

Miller, H. J., & Wentz, E. A. (2003). Representation and spatial analysis in geographic information system. *Annals of the Association of American Geographers*, *93*(3), 175–200.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, *24*(2), 223–242. https://doi.org/10.1093/applin/24.2.223

Nizonkiza, D., & Van de Poel, K. (2019). Mind the gap: Towards determining which collocations to teach. *Stellenbosch Papers in Linguistics Plus*, *56*, 13–31. http://dx.doi.org/10.5842/56-0-775

Parkinson, J. (2015). Noun-noun collocations in learner writing. *Journal of English for Academic Purposes*, *20*, 103–113. https://doi.org/10.1016/j.jeap.2015.08.003

Reppen, R., & Olson, S. (2020). Lexical bundles across disciplines: A look at consistency and variability. In U. Römer, V. Cortes & E. Friginal (Eds.), *Advances in corpus-based research on academic writing* (pp. 169–182). John Benjamins.

Scott, M. (2020). *WordSmith Tools version 8*. Stroud: Lexical Analysis Software.
Available at https://lexically.net/wordsmith/

Sinclair, J. (1987). Collocation: a progress report. In R. Steele & T. Thomas (Eds.), *Language topics: Essays in honor of Michael Halliday II* (pp. 319–331). John Benjamins.

Sinclair, J. (2004). Corpus and texts: Basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice*. Retrieved 10 May 2023 at https://users.ox.ac.uk/~martinw/dlc/chapter1.htm

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, *2*(1), 23–55. https://doi.org/10.1075/fol.2.1.03stu

Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. *The 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (Volume 1)* (pp. 173–180). Canada: Association for Computational Linguistics. https://doi.org/10.3115/1073445.1073478

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

APPENDIX

GIS NOUN COLLOCATION LIST IN ALPHABETICAL ORDER

1	**area**	
	area + noun	ratio, size, difference, distribution
	noun + area	path, service, treatment, core, land, catchment, downtown, forest, impact, target, buffer
	area + verb	include, cover, become, remain, grow, develop, contain, surround
	verb + area	cover, expand, manage, identify, delineate, locate, affect, calculate, urbanise, develop, construct
	adjective + area	urban, suburban, coastal, high, risky, geographic, potential, residential, metropolitan, rural
2	**change**	
	change + noun	detection, vector, pattern, analysis, rate
	noun + change	climate, sea-level, temperature, class, pattern, water-level, area, neighbourhood, landscape, surface, density, population
	change + verb	occur, result, vary, cause, remain, become
	verb + change	monitor, characterise, quantify, measure, assess, affect, simulate, detect, observe, model
	adjective + change	global, environmental, temporal, urban, seasonal, climatic, large, subtle, local, sharp
3	**density**	
	density +noun	surface, function, map, value, calculation, zone, pattern, level, point, area
	noun + density	population, road, housing, building, pipe
	density + verb	influence, use, have
	verb + density	normalise, predict, calculate, compare, distribute, map, model, estimate
	adjective + density	strong, high, low, maximal, urban, residential
4	**distribution**	
	distribution + noun	process, dynamic, pattern, function, map, information, model
	noun + distribution	population, income, frequency, area, space, rainfall
	distribution + verb	resemble, model, use
	verb + distribution	project, report, compare, expect, estimate, follow, influence, model, quantify, reflect, depict
	adjective + distribution	normal, geographical, cumulative, random, temporal, spatial, seasonal, local

5	**image**	
	image + noun	size, segmentation, classification, sequence, interpretation, pixel, rectification, feature, resolution
	noun + image	satellite, landsat, reference, input
	image + verb	provide, show, exhibit, become, serve, contain
	verb + image	identify, acquire, collect, use, obtain, transform, capture, display, process
	adjective + image	optical, panoramic, large, digital, temporal
6	**information**	
	information + noun	system, technology, retrieval, visualisation, exchange, extraction, overload, space, service
	noun + information	location, real-time, attribute, accessibility, change
	information + verb	become, include, contain, show
	verb + information	provide, contain, extract, obtain, collect, store, acquire, disseminate, utilise
	adjective + information	geographic, geospatial, spatial, temporal, locational, demographic, contextual, dynamic
7	**land**	
	land + noun	use, cover, value, classification
	noun + land	forest, city
	land + verb	become, cover, represent
	verb + land	characterise, locate, develop, cover, use
	Adjective + land	agricultural, private, low-lying, dry, undeveloped, rural, urban
8	**line**	
	line + noun	segment, width, string, direction, error, feature, density
	noun + line	zebra, centre, power, transition, railway, regression
	line + verb	represent, cross, pass, indicate
	verb + line	draw, detect, use, handle, cross
	adjective + line	solid, parallel, gradient, vertical, dotted
9	**location**	
	location + noun	information, score, point, accuracy, factor
	noun + location	point, event, activity, server, workplace, sample
	location + verb	influence, affect, remain, provide, differ, become
	verb + location	represent, share, predict, identify, record
	adjective + location	residential, geographic, relative, spatial, potential, observed

10	**map**	
	map + noun	layer, symbol, element, projection, overlay
	noun + map	density, vector, reference, road, base, distribution, displacement
	map + verb	match, become, represent, reflect, indicate, reveal, exhibit, contain
	verb + map	produce, create, exhibit, explore, generate, show, watermark, overlay
	adjective + map	thematic, digital, historic, topographic, analogue, interactive
11	**model**	
	model + noun	output, repository, performance, estimates, parameter
	noun + model	regression, simulation, gravity, migration, prediction
	model + verb	produce, provide, generate, exhibit, become, differ
	verb + model	develop, apply, employ, calibrate, construct, generate, design, adopt
	adjective + model	spatial, digital, linear, statistical, general, dynamic
12	**movement**	
	movement + noun	behaviour, vector, interaction, pattern, direction, area, configuration
	noun + movement	pedestrian, surface, sensor, human, ground
	movement + verb	increase, become, enhance, allow, facilitate, count, affect, include
	verb + movement	gain, analyse, concentrate, characterise, coordinate, correlate, track, capture
	adjective + movement	directional, circular, relative, rapid, integrated, horizontal, distributional, dynamic
13	**network**	
	network + noun	coverage, distance, optimisation, design, structure, link, time, model
	noun + network	road, transportation, sensor, route, space, image
	network + verb	remain, provide, contain, cover, use
	verb + network	build, sustain, support, enter, protect, represent, map, develop
	adjective + network	social, neural, multimodal, organisational, artificial
14	**point**	
	point + noun	cloud, dataset, location, feature, distribution, pattern, map, trajectory
	noun + point	control, anchor, feature, access, data, location, target, image, entry
	point + verb	increase, affect, occur, represent, fall
	verb + point	select, contain, locate, generate, extract, connect, move, remove, detect, transfer, calculate, convert
	adjective + point	central, peak, random, focal

15	**population**	
	population + noun	flow, density, count, size, growth, distribution, ratio, estimate, change, migration, threshold, concentration
	noun + population	immigrant, street, resident, world, city
	population + verb	become, contribute, live, reside, work, characterise, depend
	verb + population	increase, limit, distribute, include, estimate
	adjective + population	general, estimated, existing, local, heterogeneous, urban, residential, entire
16	**road**	
	road + noun	condition, network, segment, density, buffer, map, length, width
	noun + road	target, slip, trunk
	road + verb	connect, affect, become, remain
	verb + road	develop, enter, extend, explore, access
	adjective + road	major, busy, complex, urban, rural
17	**sensor**	
	sensor + noun	web, network, observation, deployment, movement, imagery, range
	noun + sensor	satellite, microwave, video, radar, gravity, displacement
	sensor + verb	change, move, provide, become, improve
	verb + sensor	set, deploy, use, provide, change
	adjective + sensor	optical, trifocal, mobile, existing
18	**space**	
	space + noun	agency, design, effect, network, model
	noun + space	activity, floor, output, feature, public, information, network
	space + verb	become, have, use, increase, decrease
	verb + space	represent, reduce, map, build, measure, construct, explore, create
	adjective + space	geographic, topological, physical, floor, regional, hybrid, heterogeneous
19	**technology**	
	technology + noun	adoption, product, advancement
	noun + technology	information, communication, liberation, wireless, management, visualisation
	technology + verb	facilitate, advance, serve, change, support, provide, become, make
	verb + technology	integrate, adopt, apply, develop, use, employ

	adjective technology	+	geospatial, digital, complex, spatial, key, advanced
20	**visualisation**		
	visualisation noun	+	function, system, change, tool, pipeline, workflow, mode, design, approach
	noun visualisation	+	information, noise, data, computer
	visualisation verb	+	rely, start, illustrate, represent, provide, change
	verb visualisation	+	facilitate, support, create, provide, include
	adjective visualisation	+	scientific, realistic, interactive, dynamic, cartographic