# A Corpus-Based Study of Lexical Bundles of Keywords Found in Online News Articles

**Pattarin Metang and Arthitaya Narathakoon***

***Corresponding author's email: arthitaya.n@litu.tu.ac.th**

**Language Institute, Thammasat University, Thailand**

*Received: May 24, 2024* *Revised: March 24, 2025* *Accepted: March 31, 2025*

## Abstract

This corpus-based study investigates lexical bundles of keywords found in online news articles, aiming to equip high school students in Thailand with the linguistic tools necessary for English admission exams. The primary objectives of this investigation are twofold: 1) to compile a list of 100 essential exam preparation keywords from the self-constructed corpus; and 2) to identify the lexical bundles involving the first 10 keywords. The study categorizes these keywords and lexical bundles by using 350 online news articles from CNN. The corpus was developed by aggregating online news articles and analyzing them using AntConc software. The findings reveal a keyword distribution where nouns are predominant (57%), alongside verbs, adjectives, and adverbs, reflecting the dynamic nature of news language. The structural classification reveals that 60% of the lexical bundles are noun phrase-based, indicating a predominant use of noun phrases in news articles. Meanwhile, 40% are verb phrase-based. Functionally, 75% of the lexical bundles serve referential purposes, including specifying attributes, identifying entities, and providing temporal references. Only 5% function as stance expressions. This highlights the focus on descriptive and contextual information. These findings not only enrich vocabulary learning for students, but they also guide educators in designing more targeted teaching materials. This study concludes with pedagogical implications and suggestions for further research in the field of English language education, emphasizing the critical role of authentic textual analysis in preparing students for successful academic and professional futures.

**Keywords:** lexical bundles, corpus-based study, English admission exams, online news articles, vocabulary learning

## Introduction

In Thailand, mastering English is essential for students aspiring to enter higher education. Proficiency in English is critically assessed through high-stakes tests such as TGAT and A-levels. These exams are integral to the university admission process, governed by the Council of University Presidents of Thailand (CUPT) through the Thai University Central Admission System (TCAS). TCAS has evolved over the years, introducing a multi-round competitive system with increasingly complex criteria.

Despite the high stakes, Thai students often fall short of national standards. In 2024, the average score for English in TGAT1 was only 40.63 percent, highlighting a significant gap between student abilities and exam demands. This gap is exacerbated by the complex vocabulary and lexical bundles (LBs) in test materials, which are often sourced from international news outlets such as CNN. The reading passages were selected from CNN's website to align with the TGAT exam's official sample for the reading section (MyTCAS, 2024), ensuring that students are exposed to recent vocabulary and collocations relevant to the exam. Previous TGAT exams and sample tests have frequently included articles from CNN, underscoring the need for targeted preparation using lexical bundles from these texts. A critical aspect of English admission exams, including TGAT, is their reliance on authentic reading materials that reflect real-world language use. While official blueprints provide limited details on passage sources, they indicate an emphasis on internationally recognized texts (MyTCAS, 2024). Additionally, widely used mock exams shared by TGAT test prep communities and admission websites consistently feature CNN news articles as part of their reading sections (GATEngThailand, 2024; Sangfans, 2024). This trend suggests that CNN articles serve as a key text type in TGAT, reinforcing the importance of constructing a corpus based on such sources. Furthermore, linguistic research on standardized testing supports the role of journalistic writing in academic reading assessments, emphasizing that international news sources contain complex lexical bundles that challenge students' comprehension skills (Gunning, 2002). Given the prominence of CNN articles in both mock exams and test preparation materials, it is essential to analyze their lexical structures to better prepare students for the demands of university admission exams. Additionally, studies on English language testing have noted that authentic news sources contribute to lexical complexity in high-stakes exams, making CNN an ideal corpus choice for exam-focused linguistic analysis. Educational reforms in Thailand, particularly the Basic Education Core Curriculum B.E. 2551 (2008), have emphasized English across various learning dimensions, from communication and culture to global contexts. These reforms highlight the need for high English proficiency to ensure students can participate effectively in the global arena and meet rigorous educational standards. As university entrance criteria become more demanding, proficiency in complex lexical items and lexical bundles has become crucial.

One primary challenge for Thai high school students is mastering a robust vocabulary and decoding complex lexical bundles. These skills are crucial for navigating the reading passages of English proficiency tests. Educational theorists like Wilkins (1972) and Nation (2010) emphasize the importance of vocabulary in achieving language proficiency. The study of lexical bundles, words that frequently occur together, is especially vital. They often predict language fluency and enhance reading comprehension and overall language performance.

Accordingly, this study is guided by two research questions:

1. What are the 100 keywords found in online news articles that are essential for high school students to prepare for English admission exams?
2. What are the most frequent lexical bundles involving these keywords in online news articles?

By addressing these questions, the study aims to bridge the gap between the current capabilities of students and the linguistic demands of high-stakes English proficiency tests, ultimately facilitating a better alignment between teaching strategies, test preparations, and the actual language use in academic and testing contexts.

## Literature Review

### Admission Exams

The evolution of university admission criteria in Thailand has been significantly shaped by reforms in education, particularly the Basic Education Core Curriculum B.E. 2551 (2008), which emphasizes English language education across various strands. The introduction of the Thai University Central Admission System (TCAS), modified by the Council of University Presidents of Thailand (CUPT), has made the admission process highly competitive and complex. High-stakes language tests such as TGAT, TPAT, and A-levels play a pivotal role in these admissions, where the proficiency in English directly influences students' future educational and career opportunities. The rigorous demands of these exams necessitate a strong command of English, highlighting the importance of effective language education strategies to enhance global participation and career readiness (Cherngchawano & Jaturapitakkul, 2014).

### Challenges Faced by High School Students

Mastering a substantial vocabulary is fundamental for high school students to achieve fluency in English and excel in high-stakes exams. Wilkins (1972) noted that vocabulary is crucial for effective communication, as students struggle to comprehend and participate in conversations without it. Schmitt (2010) and Nation (2010) emphasized the importance of intentional vocabulary learning for understanding complex academic texts. Additionally, Jones and Sinclair (1974), supported by Henriksen (2013) and Hill (2000), highlighted the importance of lexical bundles for developing communicative competence and achieving near-native fluency. Understanding and using lexical bundles effectively is crucial for reading comprehension and performance in English tests.

### The Gap in Language Proficiency

The sophisticated English vocabulary required in admission exams poses a challenge, as students often encounter specialized terms and lexical bundles not used in everyday language. This disconnect hinders their exam performance, which is critical for university admissions. Complex texts in exams, often sourced from international news outlets such as CNN, underscore the need for familiarity with contemporary and specialized language (MyTCAS, 2024). This shift towards more complex language forms has created a gap between students' current abilities and the advanced skills required, necessitating targeted educational interventions (Gunning, 2002).

### Corpus-Based Studies and Educational Implications

A corpus software has been used to identify high-frequency lexical bundles across various domains, providing insights into the language requirements of high school students preparing for English tests (Shin, 2007; Ackermann & Chen, 2013; Molavi et al., 2014). With the integration of this software, corpus-based research into lexical bundles provides valuable insights for educational practices. Studies by Ackermann and Chen (2013) and Molavi et al. (2014) illustrate the utility of corpus analysis in identifying prevalent lexical bundles and designing language teaching materials that align with real-world use. These studies highlight the disparity between language taught in educational settings and that used in native contexts, suggesting curriculum adjustments to better prepare students for academic and professional environments.

Biber et al.'s (2004) classification of lexical bundles is widely recognized. Lexical bundles are grouped into three categories: referential bundles, discourse organizers, and stance bundles. These sequences of words frequently co-occur in specific texts and serve various functional roles in communication. Biber's classification covers significant combinations of parts of speech and their functional use in context. For example, referential bundles like *in terms of* or *on the basis of* provide context or specify relationships. Discourse organizers such as *on the other hand* or *as a result of* guide the reader through the discourse, indicating contrast or causation. Stance bundles like *is likely to* or *can be seen as* express attitudes or assessments, adding an evaluative dimension to the discourse.

Lexical bundles are frequently used in academic texts, aiding students in developing comprehension and production skills in English. In the study of Priyatno et al. (2023), it has shown the importance of incorporating lexical bundles into EFL textbooks to enhance vocabulary acquisition and fluency. By integrating Biber's framework, educators can better prepare students for the linguistic demands of academic and professional environments.

Recent studies have revealed that lexical bundles are prevalent across many different registers, both written and spoken. Particularly, academic prose has garnered significant attention for its use of lexical bundles (Biber et al., 1999, 2003, 2004; Cortes, 2002, 2004). These studies highlight that lexical bundles possess distinct structural characteristics and serve various discourse functions in academic texts. Previous research has primarily focused on defining the characteristics and functions of lexical bundles in specific genres and comparing the use of lexical bundles by native speakers (NS) and non-native speakers (NNS) (Cortes, 2002, 2004; Biber & Barbieri, 2007; Hyland, 2008a, 2008b). However, there has been a growing interest in studies comparing lexical bundles in the writings of NS and NNS (e.g., Römer & Arbor, 2009; Chen & Baker, 2010; Salazar, 2014).

In the study of Chen and Baker (2010), lexical bundles were found in the academic writing of NS and NNS students, revealing variations between the two groups. They found that L2 students tend to overuse certain lexical bundles, such as *all over the world*, while underusing others like *in the context of*, which are more common in academic prose. Similarly, Ädel and Erman (2012) compared the writings of NS and NNS, noting a wider range of lexical bundle types in the writings of NS. DeCock (2000) reported that NNS used more lexical bundles than NS in undergraduate writing. Moreover, Römer and Arbor (2009) investigated the use of lexical bundles in expert and apprentice academic writings by NS and NNS, concluding that both groups lacked the use of academic English bundles and highlighting the need for learning the language requirements of academic writing for both native and non-native speakers.

**Classifications of Lexical Bundles**

There are two major classifications of lexical bundles: structural classification and functional classification. According to Biber et al. (1999), they categorized lexical bundles in academic texts into 12 major structural categories:

Noun phrase with of-phrase fragment: e.g., *the end of the*

Noun phrase with another post-modifier fragment: e.g., *a good example*

Prepositional phrase with embedded of-phrase fragment: e.g., *in the middle of*

Other prepositional phrase fragment: e.g., *as in the case*

Anticipatory it + verb phrase/adjective phrase: e.g., *it is possible to*

Passive verb + prepositional phrase fragment: e.g., *is based on the*

Copula be + noun phrase/adjective phrase: e.g., *is one of the*

(verb phrase +) that-clause fragment: e.g., *has been shown that*

(verb/adjective +) to-clause fragment: e.g., *are likely to be*

Adverbial clause fragment: e.g., *as shown in figure*

Pronoun/noun phrase + be (+...): e.g., *there was no significant*

Other expressions: e.g., *as well as the*

Later, Biber et al. (2004) and Cortes (2002, 2004) further developed another classification: functional classification, which is a functional taxonomy of lexical bundles, dividing them into three primary categories: stance expressions, discourse organizers and referential expressions. The first category, stance expressions, expresses an author's or speaker's attitudes, feelings, judgments, or commitment concerning the message. It includes epistemic stance: e.g., *I think it was*, *are more likely to*, and attitudinal/modality stance such as desire: e.g., *if you want to*, obligation/directive: e.g., *it is* *necessary to*, intention/prediction: e.g., *is going to be* and ability: e.g., *to be able to*. Secondly, it focuses on discourse organizers. These bundles help structure the text by presenting, clarifying, and elaborating on topics. It includes topic introduction/focus: e.g., *in this chapter we* and topic elaboration/clarification: e.g., *on the other hand*. The last type of this is referential expressions. These bundles relate to specified attributes, conditions, or refer to numbers, quantities, sizes, as well as time and place. They include identification/focus, such as *one of the most*, and imprecision, such as *and things like that.* Specification of attributes is another subcategory, which includes quantity specification like *a lot of people*, tangible framing attributes like *in the form of,* and intangible framing attributes like *in the case of.* Additionally, referential expressions cover time/place/text references. Examples include place reference: *in the United States,* time reference: *at the same time,* text deixis: *as shown in Figure N,* and multi-functional reference: *at the end of.* This framework was used to analyze the data, providing a comprehensive understanding of the structural and functional patterns of lexical bundles in this corpus.

## Related Literature Review

The examination of lexical bundles has seen significant contributions from various studies that explore the impact of these linguistic constructs on language learning and proficiency. For instance, Sarjono et al. (2022) conducted a comprehensive corpus analysis to identify lexical bundles in an English textbook for Indonesian high school students, underscoring the educational value of integrating lexical bundles analysis into language teaching to enhance proficiency. This aligns with Farooqui (2016), who analyzed lexical bundles used in academic writing, highlighting the differences between native and non-native speakers, particularly in their overuse of certain noun lexical bundles, which could reflect a narrower range of language mastery.

Furthering the field of corpus linguistics, Ackermann and Chen (2013) developed the Academic Collocation List from a large corpus to aid English for Academic Purposes (EAP) learners, emphasizing the practical application of lexical bundles studies in educational settings. Similarly, Molavi et al. (2014) analyzed lexical bundles in popular English textbooks to gauge their effectiveness in reflecting actual language use, discovering a discrepancy between the lexical bundles taught and those used by native speakers. This study suggests a potential misalignment in language teaching materials, which could impact learners' ability to engage with natural English usage effectively. While these studies provide valuable insights, they also have limitations. For example, Ackermann and Chen (2013) and Molavi et al. (2014) primarily focused on textbook content, which may not fully represent the dynamic and context-specific nature of lexical bundles in real-world usage. Similarly, Farooqui's (2016) analysis of academic writing highlighted differences in lexical bundle usage between native and non-native speakers, but it did not address the impact of these differences on test performance.

In Thailand, several articles delve into lexical bundles. Sukman et al. (2022) embarked on a study to explore lexico-grammatical elements within a specialized corpus of online business news. They found that noun-based lexical bundles were used a lot, which is important for getting complicated ideas across in certain situations. This study is particularly relevant as it connects the analysis of lexical bundles to real-world applications, similar to the studies by Trinant and colleagues (2019, 2021), who focused on academic corpora in nursing and tourism, further illustrating the significance of noun phrases in scholarly communication.

These investigations collectively underscore the crucial role of lexical bundles in enhancing language proficiency and the need for educational practices to incorporate these insights to better prepare students for academic and professional challenges. However, despite these advancements, there remains a gap in applying these findings to the preparation of students for English admission exams, particularly in understanding how lexical bundles are featured in online news articles, a key component of contemporary language assessment tests. Official sample exams and blueprints indicate that reading passages are frequently sourced from online news outlets such as CNN, aligning with the increasing emphasis on real-world language exposure in high-stakes assessments (MyTCAS, 2024). This review sets the stage for further research aimed at bridging this gap, focusing on the specific needs of students preparing for high-stakes English tests by enhancing their comprehension of complex textual materials through targeted vocabulary and lexical bundles studies.

## Methodology

### Data Collection and Corpus Creation

The data for this study were collected from 350 online news articles published between 2023 and 2024, sourced from CNN, one of the world's largest news broadcasters, covering various subjects. The news was selected from nine groups: world, business, health, entertainment, tech, style, travel, sports, and weather. In each group, the news was selected equally to compile into 350 news stories. The corpus comprised 214,666 tokens, which is well within the recommended range for corpus-based lexical studies. While small-scale corpora often range from 2,000 to 5,000 words for pilot studies (Evans, 2007), research on lexical bundles and corpus linguistics typically utilizes corpora containing at least 100,000 tokens for meaningful frequency analysis (Biber et al., 1999; McEnery & Hardie, 2011). Previous lexical studies have successfully analyzed corpora of similar sizes, such as the Michigan Corpus of Academic Spoken English (MICASE) and sub-samples of the British National Corpus (BNC), both of which contain sub-corpora of 200,000–300,000 tokens for specific lexical and grammatical research. Given that this study focuses on lexical bundle extraction, a corpus of over 200,000 tokens provides a robust dataset for meaningful linguistic analysis.

The collection period spanned a year from the date of the news publication. This study compiled a specialized corpus, the Corpus of Online News Article References (CONAR), designed to analyze the lexical bundles and vocabulary found in online news articles. The reading passages were selected from CNN's website to align with the TGAT exam's official sample for the reading section (MyTCAS, 2024), aiming to expose students to recent vocabulary and lexical bundles pertinent to TGAT exams. CNN provides a wide range of current and relevant topics, ensuring that the language used in the corpus is contemporary and widely recognized. Additionally, CNN's international reach and reputation for high-quality journalism make it a suitable choice for developing a corpus that reflects the lexical demands of English proficiency tests.

To contextualize the lexical bundles in CONAR, this study compares its findings with a general American English (AmE) corpus database, specifically AmE06, which contains 1,017,879 tokens. The AmE06 corpus is a widely used linguistic dataset that represents standard American English usage across different registers, including spoken and written texts. By incorporating this comparative analysis, the study ensures that the lexical bundles identified in CNN news articles are not only examined within a journalistic framework but also evaluated against broader patterns of American English usage. This comparison serves two key purposes: first, to determine whether the lexical bundles found in CNN news articles are unique to journalistic discourse or commonly used across various domains of American English; and second, to assess how well CNN-derived lexical bundles align with general English usage, ensuring their relevance for English proficiency exams that test academic and professional language skills.

All news articles were retained in their original length and stored electronically in CONAR. Non-textual elements like tables, charts, diagrams, figures, references, and photos were excluded. AntConc 4.2.2 (Anthony, 2023) facilitated data analysis. It's crucial to note that CONAR is a specialized corpus tailored to replicate the language used in news articles, providing insights into the lexical structures students encounter in English proficiency exams.

## Data Analysis

The present study used a corpus-based approach to identify keywords and generate a compilation of keyword lexical bundles present in online news articles. The two methods required to accomplish these objectives are 1) collecting a corpus and determining keywords, and 2) extracting lexical bundles of keywords.

AntConc was used to compile a corpus and generate a keyword list from the news (214,666 tokens). Keywords were identified based on their statistical significance compared to the AmE corpus database (AmE06 1,017,879 tokens), chosen for its size and representation of general English usage. This database was selected to ensure that the analysis reflects American English usage, as the primary goal is to compare these keywords with those found in CNN news articles, which is a prominent American press. Using the AmE corpus allows for a more accurate and relevant comparison, aligning the linguistic characteristics of the corpus with those of the American news media. In identifying keywords from the collected news articles, the study employed the statistical measure of *keyness* to determine the significance of word frequency in the target corpus (CONAR) compared to a reference corpus (AmE06, 1,017,879 tokens). Keyness measures how much more frequently a word appears in the target corpus than would be expected based on its frequency in the reference corpus, typically using a chi-square test or log-likelihood ratio. For this analysis, a keyness value of ≥ 20 and a frequency threshold of ≥ 50 were set, ensuring that identified keywords are both statistically significant and commonly used. This method allows for a precise comparison between the language used in CNN news articles and general American English, resulting in a top 100 keyword list for students to prepare for the exams. After creating a list of the top 100-keywords, those keywords were categorized into grammatical classes and grouped into different areas in order to understand the nature of the vocabulary and the group of keywords.

After identifying keywords, they were used as nodes to find lexical bundles using AntConc's Cluster function, focusing on a four-word span (Jones & Sinclair, 1974). Lexical bundles with at least four occurrences were considered (Nelson, 2000), excluding clusters with function words at the beginning or end. Function words that are the first or last in the cluster were taken out of the list because the focus of this study is on lexical bundles. Following the completion of the second step, the researcher identified the lexical bundles of the top 10 keywords from the word lists. Subsequently, an analysis was conducted to determine the percentage of different types of lexical bundles found within these pairs. Finally, recommendations and instructional implications were provided based on the findings.

## Results

### The List of the Top 100 Keywords

The quantitative results have illustrated a list of the first 100 keywords, which occurred at a significant frequency in the online news. The following Table 1 gives an impression of what the corpus in this study is all about.

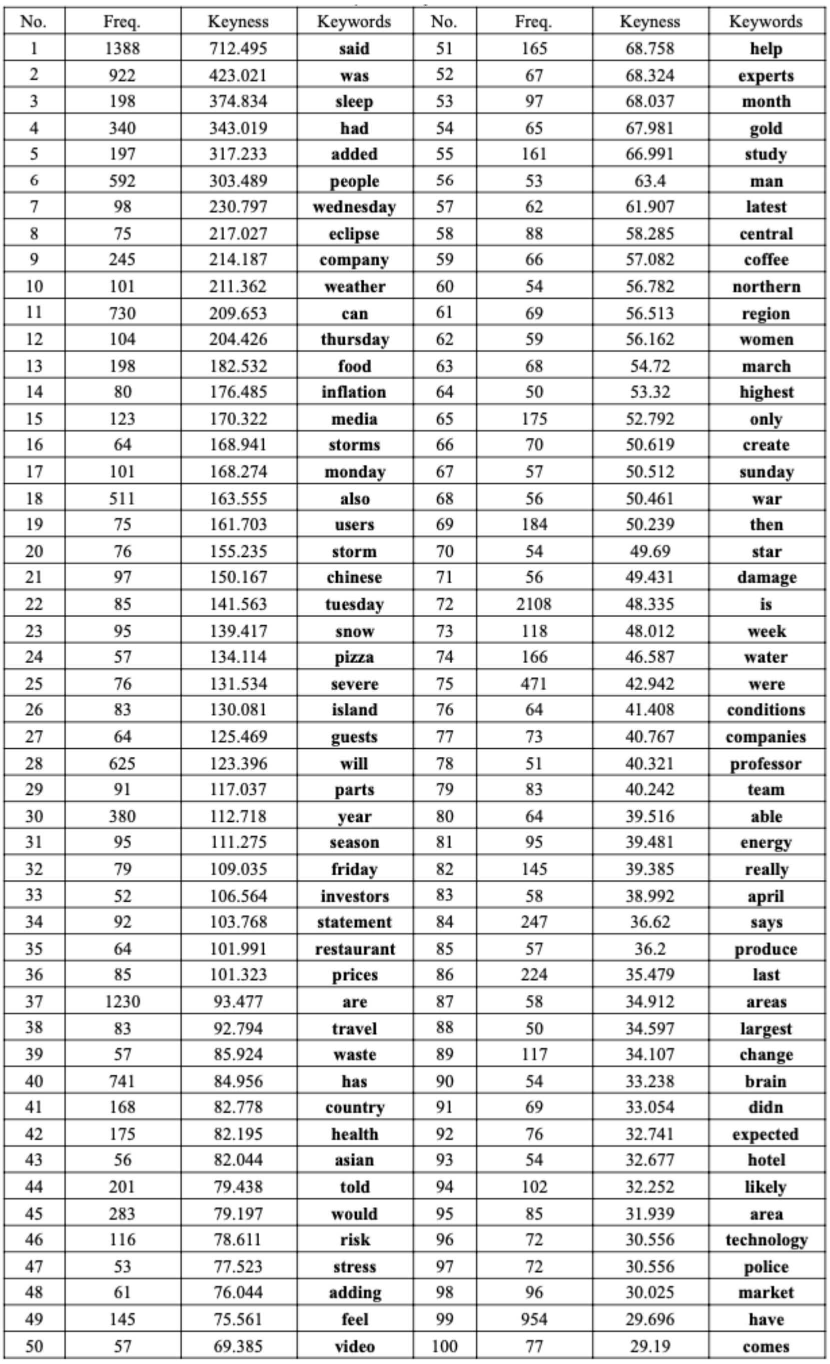
The 100 lexical keywords can be classified into five grammatical types. Out of the total, 57% were nouns, making up the largest share. Verbs accounted for 22%, adjectives for 9%, nouns/verbs (the keywords that were used as both verbs and nouns) for 8%, and adverbs for 4%. These words appear to be associated with diverse subjects, encompassing news, events, and general discourse. Certain terms are associated with the business world like *inflation, market, and inverters*, whilst others are more universal in nature, such as *food, health, and travel.* The text appears to consist of a combination of nouns, verbs, and adjectives that are frequently used in articles, reports, or debates related to current events.

The top 100 keywords can be categorized into 10 broad groups based on their associated meanings, as outlined below:

1. Keywords referring to news and media
   - said, according, media, statement, experts, latest, professor, says
2. Keywords referring to time and days
   - Wednesday, Thursday, Monday, Tuesday, Friday, March, Sunday, week, month, April

Table 1

*Top 100 Keywords in Online News*



| No. | Freq. | Keyness | Keywords | No. | Freq. | Keyness | Keywords |
|---|---|---|---|---|---|---|---|
| 1 | 1388 | 712.495 | **said** | 51 | 165 | 68.758 | **help** |
| 2 | 922 | 423.021 | **was** | 52 | 67 | 68.324 | **experts** |
| 3 | 198 | 374.834 | **sleep** | 53 | 97 | 68.037 | **month** |
| 4 | 340 | 343.019 | **had** | 54 | 65 | 67.981 | **gold** |
| 5 | 197 | 317.233 | **added** | 55 | 161 | 66.991 | **study** |
| 6 | 592 | 303.489 | **people** | 56 | 53 | 63.4 | **man** |
| 7 | 98 | 230.797 | **wednesday** | 57 | 62 | 61.907 | **latest** |
| 8 | 75 | 217.027 | **eclipse** | 58 | 88 | 58.285 | **central** |
| 9 | 245 | 214.187 | **company** | 59 | 66 | 57.082 | **coffee** |
| 10 | 101 | 211.362 | **weather** | 60 | 54 | 56.782 | **northern** |
| 11 | 730 | 209.653 | **can** | 61 | 69 | 56.513 | **region** |
| 12 | 104 | 204.426 | **thursday** | 62 | 59 | 56.162 | **women** |
| 13 | 198 | 182.532 | **food** | 63 | 68 | 54.72 | **march** |
| 14 | 80 | 176.485 | **inflation** | 64 | 50 | 53.32 | **highest** |
| 15 | 123 | 170.322 | **media** | 65 | 175 | 52.792 | **only** |
| 16 | 64 | 168.941 | **storms** | 66 | 70 | 50.619 | **create** |
| 17 | 101 | 168.274 | **monday** | 67 | 57 | 50.512 | **sunday** |
| 18 | 511 | 163.555 | **also** | 68 | 56 | 50.461 | **war** |
| 19 | 75 | 161.703 | **users** | 69 | 184 | 50.239 | **then** |
| 20 | 76 | 155.235 | **storm** | 70 | 54 | 49.69 | **star** |
| 21 | 97 | 150.167 | **chinese** | 71 | 56 | 49.431 | **damage** |
| 22 | 85 | 141.563 | **tuesday** | 72 | 2108 | 48.335 | **is** |
| 23 | 95 | 139.417 | **snow** | 73 | 118 | 48.012 | **week** |
| 24 | 57 | 134.114 | **pizza** | 74 | 166 | 46.587 | **water** |
| 25 | 76 | 131.534 | **severe** | 75 | 471 | 42.942 | **were** |
| 26 | 83 | 130.081 | **island** | 76 | 64 | 41.408 | **conditions** |
| 27 | 64 | 125.469 | **guests** | 77 | 73 | 40.767 | **companies** |
| 28 | 625 | 123.396 | **will** | 78 | 51 | 40.321 | **professor** |
| 29 | 91 | 117.037 | **parts** | 79 | 83 | 40.242 | **team** |
| 30 | 380 | 112.718 | **year** | 80 | 64 | 39.516 | **able** |
| 31 | 95 | 111.275 | **season** | 81 | 95 | 39.481 | **energy** |
| 32 | 79 | 109.035 | **friday** | 82 | 145 | 39.385 | **really** |
| 33 | 52 | 106.564 | **investors** | 83 | 58 | 38.992 | **april** |
| 34 | 92 | 103.768 | **statement** | 84 | 247 | 36.62 | **says** |
| 35 | 64 | 101.991 | **restaurant** | 85 | 57 | 36.2 | **produce** |
| 36 | 85 | 101.323 | **prices** | 86 | 224 | 35.479 | **last** |
| 37 | 1230 | 93.477 | **are** | 87 | 58 | 34.912 | **areas** |
| 38 | 83 | 92.794 | **travel** | 88 | 50 | 34.597 | **largest** |
| 39 | 57 | 85.924 | **waste** | 89 | 117 | 34.107 | **change** |
| 40 | 741 | 84.956 | **has** | 90 | 54 | 33.238 | **brain** |
| 41 | 168 | 82.778 | **country** | 91 | 69 | 33.054 | **didn** |
| 42 | 175 | 82.195 | **health** | 92 | 76 | 32.741 | **expected** |
| 43 | 56 | 82.044 | **asian** | 93 | 54 | 32.677 | **hotel** |
| 44 | 201 | 79.438 | **told** | 94 | 102 | 32.252 | **likely** |
| 45 | 283 | 79.197 | **would** | 95 | 85 | 31.939 | **area** |
| 46 | 116 | 78.611 | **risk** | 96 | 72 | 30.556 | **technology** |
| 47 | 53 | 77.523 | **stress** | 97 | 72 | 30.556 | **police** |
| 48 | 61 | 76.044 | **adding** | 98 | 96 | 30.025 | **market** |
| 49 | 145 | 75.561 | **feel** | 99 | 954 | 29.696 | **have** |
| 50 | 57 | 69.385 | **video** | 100 | 77 | 29.19 | **comes** |

3. Keywords referring to weather and natural events
    - eclipse, weather, storms, storm, snow, severe, island, parts, conditions, water, damage
4. Keywords referring to business and economy
    - company, inflation, restaurant, prices, investors, ceo, waste, companies, energy, produce, change, market
5. Keywords referring to international relations
    - Chinese, country, war, central, northern, region

6. Keywords referring to technology and AI
   - users, video, brain
7. Keywords referring to food and dining
   - food, pizza, guests, restaurant, coffee
8. Keywords referring to health and wellness
   - sleep, health, stress, feel, risk, health, study, women, highest, brain
9. Keywords referring to travel and transportation
   - travel, guests, area, hotel
10. Miscellaneous keywords
    - had, added, people, can, also, will, are, told, would, as, waste, has, are, only, create, were, able, then, star, were, only, likely, expected, police, have, come, largest, help, able, really

**Most frequent lexical bundles of keywords in CONAR.**

Only lexical bundles for the two-word clusters are displayed here. In relation to two-word clusters, only function words that occur in the first or last position of the cluster are excluded. This is because the main emphasis of this study is on lexical items rather than grammatical ones. Table 2 displays the initial 10 keywords, together with the three most frequent 2-word clusters on both sides of the node.

The keyword *said* predominantly appears with formal sources and announcements in business and official reports. For example, *companies said* and *ministry said* are typical in corporate and governmental contexts, respectively. This reflects the authoritative tone often required in reporting statements from entities or officials, where accuracy and direct attribution are critical, such as *said company* and *said administration's*.

The verb *was* features in lexical bundles that describe states or conditions, such as *idea was* and *response was*. It also appears in descriptions of art and commerce, such as *painting was* and *was exported*. These clusters suggest a focus on past events or states, providing background or historical context in news narratives.

In the domain of health and lifestyle, *sleep* collocates frequently with terms like *sleep medicine* and *sleep apnea*, indicating a focus on sleep-related disorders and their treatments. This highlights the importance of sleep health in public discourse, evidenced by clusters such as *obstructive sleep*, *deep sleep*, and *healthy sleep*.

Table 2

*Examples of the Top 10 Keywords with Their Lexical Bundles*

| No. | Keywords | 2-word clusters | **No.** | Keywords | 2-word clusters |
|---|---|---|---|---|---|
| 1 | **said** | companies said<br>study said<br>ministry said<br>said the company<br>said the authors<br>said the administration's | 2 | **was** | idea was<br>response was<br>painting was<br>was pleasantly (surprised )<br>was particulary (think of)<br>was deeply (concerned) |
| 3 | **sleep** | sleep medicine<br>sleep apnea<br>sleep disorder<br>obstructive sleep<br>deep sleep<br>healthy sleep | 4 | **had** | Nasdaq had<br>noise had<br>trial had<br>had (the) controls<br>had severe (injuries)<br>had (trained) extensivey |
| 5 | **added** | The authors added<br>the airline added<br>the government added<br>added advantage<br>added jobs<br>added pets | 6 | **people** | target people<br>find people<br>help people<br>people try<br>people experiencing<br>people use |
| 7 | **wednesday** | Wednesday afternoon<br>Wednesday night<br>late Wednesday<br>published Wednesday<br>released Wednesday<br>Wednesday report | 8 | **eclipse** | solar eclipse<br>partial eclipse<br>total eclipse<br>eclipse will<br>eclipse glasses<br>eclipse watchers |
| 9 | **company** | company said<br>company's headquarters<br>company blog<br>parent company<br>honest company<br>foreign company | 10 | **weather** | national weather<br>severe weather<br>fire weather<br>weather service<br>weather conditions<br>weather outbreak |

The main verb *had* is used to denote possession or experience, seen in *Nasdaq had* and *trial had*. It frequently appears in financial and legal contexts.

The verb *added* collocates often with organizational entities like *the authors added* and *the government added*, showing the addition of information, features, or resources. This is reflected in clusters such as *added advantage*, suggesting enhancements or expansions in various contexts.

The keyword *people* frequently appears with verbs that describe interactions or services directed at groups, such as *target people*, *find people*, and *help people*. This underscores the societal focus in news articles, where the activities, experiences, and usage of resources by people are central themes, as in *people try*, *people experiencing*, and *people use*.

Lexical bundles featuring *Wednesday*, such as *Wednesday afternoon*, *Wednesday night*, and *late Wednesday*, primarily indicate the use of the day to time-stamp various events. The terms often denote specific parts of the day, enhancing the temporal precision in reporting or scheduling. For instance, *published Wednesday* and *released Wednesday* suggest typical usage in contexts related to media and official announcements, where timing is crucial for the audience's understanding.

Lexical bundles with *eclipse*, like *solar eclipse*, *partial eclipse*, and *total eclipse*, demonstrate its use in astronomical contexts. The keyword is commonly associated with different types of eclipses, emphasizing the event's nature and scope. Additionally, terms like *eclipse glasses* and *eclipse watchers* connect with the behaviors and preparations of people observing this celestial phenomenon.

For *company*, lexical bundles such as *company said*, *company's headquarters*, and *parent company* suggest a focus on business entities and their operations or statements. The usage spans from reporting statements made by a company (*company said*) to referencing physical or structural aspects of businesses (*company's headquarters*). This reflects a common use in business reporting and corporate communications.

The lexical bundles around *weather*, such as *national weather, severe weather,* and *weather service,* reflect its usage in discussions related to meteorological conditions. These clusters indicate the operational and informative aspects of weather reporting, with terms like *weather conditions* and *weather outbreak* suggesting how weather descriptions cater to public safety and awareness.

Table 3 categorizes the top 10 keywords and their associated 2-word clusters using Biber et al.'s (1999) structural classification and Biber et al.'s (2004) functional classification of lexical bundles. Structurally, the lexical bundles in the data predominantly fall into two categories: noun phrase-based bundles and verb phrase-based bundles. The noun phrase-based bundles, such as *sleep medicine*, *solar eclipse*, and *national weather*, account for a significant portion of the data and serve to specify attributes and entities related to the keywords. Additionally, there are instances of noun phrases with other post-modifier fragments, such as *Wednesday afternoon* and *company's headquarters*, which provide more specific contextual information.

Verb phrase-based bundles are also prominent in the data, including examples like *companies said*, *was deeply (concerned)*, and *the authors added.* These bundles typically denote actions or states associated with the keywords.

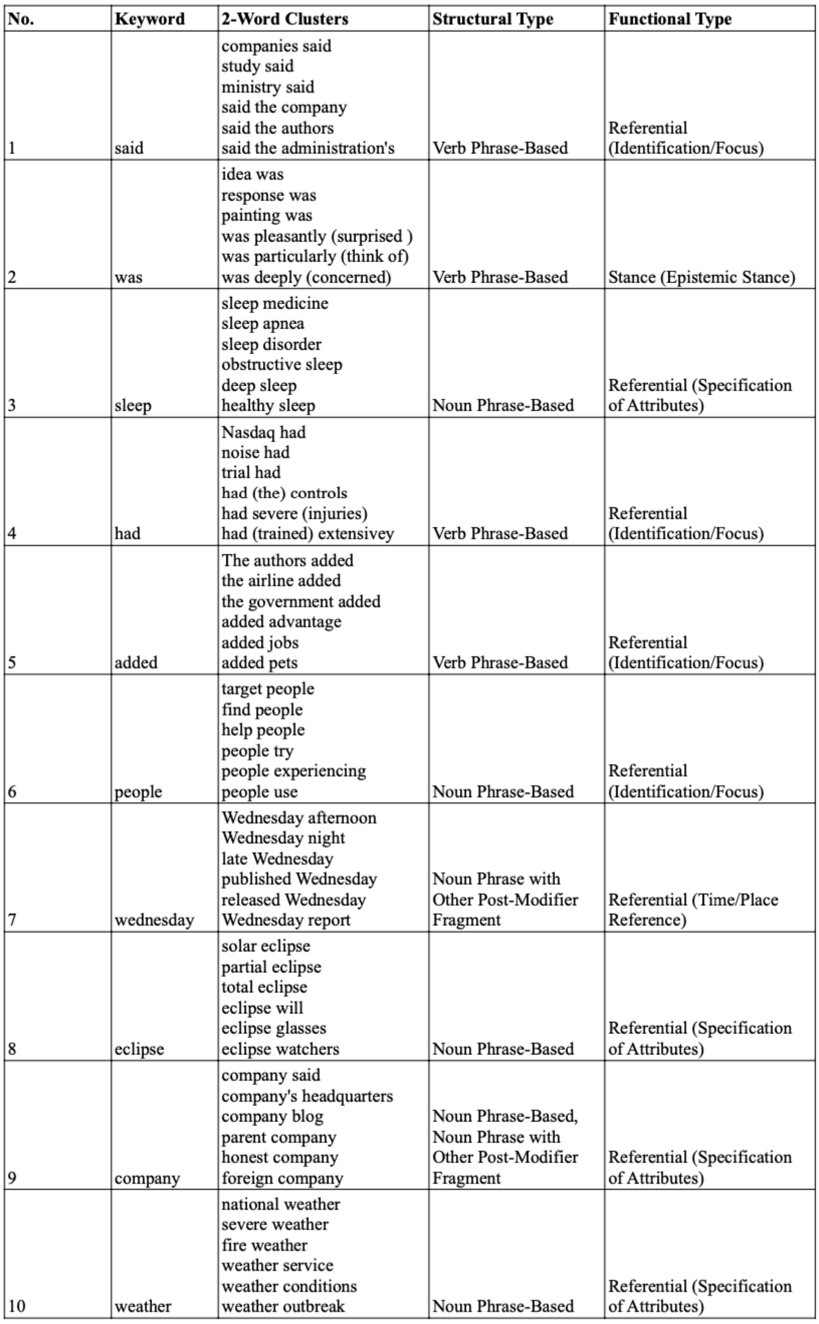
From a functional perspective, the majority of the lexical bundles are referential expressions. These include identification and focus bundles such as *companies said*, *Nasdaq had*, and *target people*, which help to pinpoint specific entities or actions. Specification of attributes is another common function, with bundles such as *sleep disorder*, *solar eclipse*, and *weather conditions* describing specific characteristics related to the keywords. Time and place references, such as *published Wednesday* and *Wednesday report*, are also present, providing temporal and locational context.

Stance expressions, particularly epistemic stance bundles, are evident in phrases including *was pleasantly surprised*, *was particularly think of*, and *was deeply concerned*. These bundles express the author's or speaker's attitudes and evaluations concerning the message.

Examples of lexical bundles produced using the previously collected keywords are displayed in Table 3. To facilitate understanding, Figure 1 provides the percentage of bundles in CONAR.

Table 3

*Structural Distribution of LBs in CONAR*

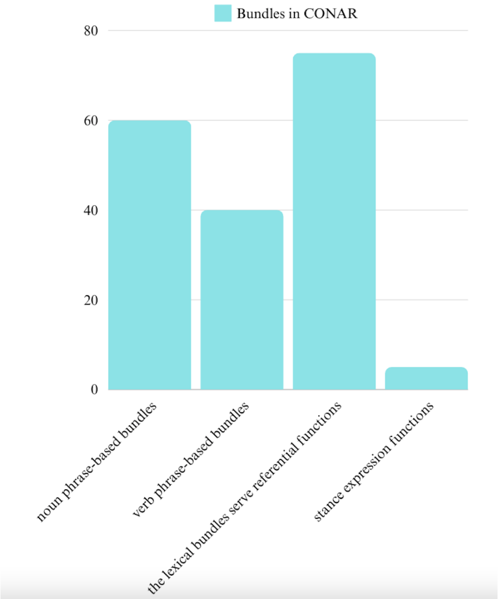


| No. | Keyword | 2-Word Clusters | Structural Type | Functional Type |
|---|---|---|---|---|
| 1 | said | companies said<br>study said<br>ministry said<br>said the company<br>said the authors<br>said the administration's | Verb Phrase-Based | Referential (Identification/Focus) |
| 2 | was | idea was<br>response was<br>painting was<br>was pleasantly (surprised )<br>was particularly (think of)<br>was deeply (concerned) | Verb Phrase-Based | Stance (Epistemic Stance) |
| 3 | sleep | sleep medicine<br>sleep apnea<br>sleep disorder<br>obstructive sleep<br>deep sleep<br>healthy sleep | Noun Phrase-Based | Referential (Specification of Attributes) |
| 4 | had | Nasdaq had<br>noise had<br>trial had<br>had (the) controls<br>had severe (injuries)<br>had (trained) extensivey | Verb Phrase-Based | Referential (Identification/Focus) |
| 5 | added | The authors added<br>the airline added<br>the government added<br>added advantage<br>added jobs<br>added pets | Verb Phrase-Based | Referential (Identification/Focus) |
| 6 | people | target people<br>find people<br>help people<br>people try<br>people experiencing<br>people use | Noun Phrase-Based | Referential (Identification/Focus) |
| 7 | wednesday | Wednesday afternoon<br>Wednesday night<br>late Wednesday<br>published Wednesday<br>released Wednesday<br>Wednesday report | Noun Phrase with Other Post-Modifier Fragment | Referential (Time/Place Reference) |
| 8 | eclipse | solar eclipse<br>partial eclipse<br>total eclipse<br>eclipse will<br>eclipse glasses<br>eclipse watchers | Noun Phrase-Based | Referential (Specification of Attributes) |
| 9 | company | company said<br>company's headquarters<br>company blog<br>parent company<br>honest company<br>foreign company | Noun Phrase-Based, Noun Phrase with Other Post-Modifier Fragment | Referential (Specification of Attributes) |
| 10 | weather | national weather<br>severe weather<br>fire weather<br>weather service<br>weather conditions<br>weather outbreak | Noun Phrase-Based | Referential (Specification of Attributes) |

According to Figure 1, the analysis of the top 10 keywords and their associated 2-word clusters, using Biber et al.'s (2004) framework, provides a detailed insight into the structural and functional characteristics of lexical bundles within the corpus. The structural classification reveals that 60% of the lexical bundles are noun phrase-based, such as *sleep medicine*, *solar eclipse,* and *national weather*, while 40% are verb phrase-based, including *companies said*, *was deeply (concerned)*, and *the authors added.* This indicates a predominant use of noun phrases and verb phrases in the context of news articles.

Figure 1

*The Percentage of Bundles in CONAR*



In terms of functional classification, 75% of the lexical bundles serve referential functions. These include specifying attributes, identifying or focusing on specific entities or actions, and providing temporal references, as seen in bundles like *sleep disorder*, *companies said*, and *Wednesday afternoon*. Only 5% of the lexical bundles function as stance expressions, specifically epistemic stance bundles, which express certainty or evaluation, such as *was deeply (concerned)* and *was pleasantly (surprised).*

This corpus-based study has identified keywords and their lexical bundles behaviors in recent online news. As a result, high school English students can enhance their vocabulary learning efficiency. In addition, practical application can be achieved by developing lessons, learning activities, and exercises that focus on the keywords and lexical bundles lists extracted from the CONAR.

## Discussion

The results of this study on the common use of particular keywords in news articles on the internet and lexical bundles provide important new insights on the language competency resources high school students need to prepare for English admission tests. As other studies have demonstrated, a comprehensive list of terms and their often-occurring lexical bundles partners was found in the research, offering a solid foundation for vocabulary learning that has a firm foundation in everyday usage.

The examination of the top 100 keywords from online news contexts supports the ideas put forth by Nation (2001) and Scott (2012) regarding the importance of academic and high-frequency terms. These keywords play a critical role in language comprehension and competency development, covering a wide range of nouns, verbs, and adjectives. The distribution across grammatical categories, particularly the emphasis on nouns, supports the academic focus on effectively explaining complex ideas, as highlighted by both the CONAR analysis and Sukman et al. (2022). This finding aligns with

Carter's (1998) distinction between lexical and grammatical words, emphasizing that content words carry substantial informational weight compared to functional words.

Key topics emerging in 2023 and 2024, such as natural events, international relations, health, and astronomical phenomena, have significantly influenced global discussions and concerns. For instance:

- **Natural Events**: Keywords like *storm*, *snow*, and *severe* frequently appeared. Early 2024 saw a large mass of Arctic air bringing cold temperatures and snow across the United States.
- **International Relations**: The word *war* was prominent due to ongoing conflicts, including the Israel-Hamas conflict, Sudanese Civil War, and Russia-Ukraine war.
- **Health and Wellness**: The keyword *sleep* appeared frequently due to reports of poor sleep quality in 2023.
- **Astronomical Phenomena**: *Eclipse* was significant due to notable solar eclipses in October 2023 and April 2024.

These findings underscore the need for students to stay informed about current events to understand the trends in vocabulary and issues that may impact upcoming exams. The identification of common lexical bundles, especially two-word clusters, highlights the importance of comprehending word combinations in natural language usage, which is crucial for second language fluency. This supports Bennett (2010), who demonstrated the predictive power of lexical bundles in native language use.

Furthermore, findings from the larger academic community, such as the world of Trinant and Yodkamlue (2019) and Trinant and Kijpoonphol (2021), are consistent with the prominence of nouns in the keyword list and their lexical bundle frameworks. These studies demonstrated a comparable frequency of noun-based lexical bundles, essential for successfully communicating academic and technical information.

The frequently found lexical bundles like *said* in journalistic contexts or *weather* in climate and weather conversations resonate with language use in everyday situations where certain lexical bundles are essential for impact and clarity. Similar attention was paid to this practical application of language use in earlier research by Molavi et al. (2014), who criticized standard English textbooks for failing to adequately represent native speakers' actual language use, particularly when it came to lexical bundles management. This criticism is addressed in the current study by offering a focused examination of lexical bundles in an area that is directly relevant to the students of test prep materials. As a result, potential test-takers will be able to learn and comprehend language patterns that are common in test materials and other contexts with greater complexity.

As this study has identified 100 crucial keywords from various topical areas, reflecting the language students need to master for exams. These keywords cover essential topics in news articles, preparing students for real-world language use. The analysis revealed common lexical bundles associated with these keywords, highlighting the patterns and combinations of words that students need to understand. This emphasizes the importance of lexical bundles in achieving language fluency and comprehension. lexical bundles with keywords like *people*, *said*, and *weather*, which are relevant to the material found in news articles, corroborate the findings of Farooqui (2016) and other researchers who have highlighted the importance of practicing vocabulary in context. These lexical bundles help readers comprehend the text and help them learn how to use these terms correctly, which is a necessary skill for success in school and beyond. The practical implications of this research are further highlighted by the emphasis on educational uses of these findings in teaching materials, as proposed by Ackermann and Chen's (2013) integration of the Academic Collocation List. Through the integration of these keywords and lexical bundles into instructional activities and evaluations, teachers can greatly improve their students' language readiness for exams measuring English competence. Furthermore, the exploration of lexical bundles in the Corpus of Online News Article References (CONAR) in this study provides a foundation for understanding how certain keywords are used in context, which is crucial for developing reading comprehension and production skills in English.

However, there are limitations to the corpus analysis. The reliance on CNN articles means that the language used may reflect specific stylistic choices unique to this source, which might not fully represent the diversity of English used in other contexts. Additionally, the AntConc software, while powerful, has limitations in its ability to capture nuanced meanings and contextual variations, which could affect the analysis of lexical bundles.

To sum up, this study fills the gap left by earlier research by focusing on the vocabulary and lexical bundles that have the greatest influence on students getting ready for entrance tests in English. The results make significant contributions by indicating that a focus on lexical bundles found in authentic sources, such as online news, can improve the efficacy of language instruction and better prepare students for future academic and professional pursuits, as well as exams.

## Conclusion

This study has centered on the pivotal role of specific keywords and their lexical bundles within online news articles, providing high school students with a valuable advantage in preparing for their English admission tests. By identifying and analyzing the top 100 keywords, this research contributes to existing literature underlining the importance of acquiring academically relevant and high-frequency vocabulary for effective English communication and comprehension. The results show that nouns predominately appear in these terms, which highlights their crucial role in communicating significant information in a variety of scholarly and professional contexts. Furthermore, the study's emphasis on lexical bundles patterns, particularly those combining nouns and verbs, highlights how useful these word pairings are in everyday situations for improving language ability. These patterns support a curriculum that stresses contextually rich and frequently used lexical bundles, which is in line with the pedagogical needs found in earlier studies.

Additionally, the study of two-word lexical bundles commonly seen in online news articles provides practical insights that may be immediately applied to the language learning procedures required for passing English proficiency exams. It is expected that this practical implementation of the findings will transform instructional tactics and learning resources by bringing them closer to real language use in academic and testing contexts. For example, incorporating the identified top 100 keywords and their common lexical bundles into the curriculum, developing reading materials and comprehension exercises based on current online news articles that use these keywords and lexical bundles and designing mock exams that mimic the format and content of the TGAT and A-levels, using the identified keywords and lexical bundles.

In conclusion, this study provides tangible evidence supporting the integration of specific lexical bundles into instructional and evaluative approaches. It offers teachers valuable insights into enhancing their lesson plans to effectively prepare students for success in English admission tests. Moreover, it equips students with a more streamlined path to proficiency by enhancing their understanding of the terminology prevalent in online news media discourse, essential for their future academic and professional pursuits.

In addition to filling a significant research gap, this work sets the way for further studies on the dynamic interactions among language proficiency, keyword frequency, and lexical bundles across a range of text genres and learning contexts.

## 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.

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes, 31*(2), 81–92.

Anthony, L. (2023). AntConc (Version 4.2.4) [Computer software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software/antconc/

Baker, W., & Jarunthawatchai, W. (2017). English Language Policy in Thailand. *European Journal of Language Policy, 9*(1), 27–44.

Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Ann Arbor, University of Michigan Press.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263–286.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371–405.

Biber, D., Conrad, S., & Leech, G. (2003). *Student grammar of spoken and written English*.

Longman.

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

Cherngchawano, W., & Jaturapitakkul, N. (2014). Lexical profiles of Thailand University Admission Tests. *PASAA: Journal of Language Teaching and Learning in Thailand, 48*, 1–27.

Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30–49.

Cheng, W. (2012). *Exploring corpus linguistics: Language in action*. Routledge.

Cortes, V. (2002). Lexical bundles in freshman composition. In R. Reppen, S. M. Fitzmaurice, & D. Biber (Eds.), *Using corpora to explore linguistic variation* (pp. 131–145). John Benjamins.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397–423.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238.

DeCock, S. (2000). Recurrent sequences of words in native speaker and advanced learner spoken and written English: A corpus-driven approach. In C. Mair & M. Hundt (Eds.), *Corpus linguistics and linguistic theory* (pp. 51–68). Rodopi.

Evans, D. (2007). Corpus building and investigation for the Humanities. University of Nottingham. http://www.corpus.bham.ac.uk/corpus-building.shtml

Farooqui, A. S. (2016). A corpus-based study of academic-lexical bundles use and patterns in postgraduate Computer Science students' writing (Doctoral dissertation, University of Essex).

GATEngThailand. (2024, February 17). *TGAT English mock exam*. Facebook. https://www.facebook.com/GATEngThailand/

Gunning, T. G. (2002). *Assessing and correcting reading and writing difficulties*. Allyn & Bacon A Pearson Education Company.

Henriksen, B. (2013). Research on L2 learners' lexical collocational competence and development–A progress report. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *Vocabulary Acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 29–56).

Hill, J. (2000). Revisiting priorities: From grammatical failure to lexical bundlesal success. In M. Lewis (Ed.), *Teaching lexical bundles: Further developments in the lexical approach* (pp. 47–69). Commercial Colour Press Plc.

Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4–21.

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41–62.

Jones, S., & Sinclair, J. McH. (1974). English lexical bundles. *Cahiers de Lexicologie, 24*, 15–61.

Kheovichai, B. (2022). lexical bundles and discursive construction of Covid-19 in WHO Director

General's discourse: A corpus-based study. *LEARN Journal: Language Education and Acquisition Research Network, 15*(1), 10–32.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, P. (2010). *Teaching vocabulary strategies and techniques*. Heinle Cengage Learning.

Nelson, M. A. (2000). Corpus-based study of business English and business English teaching materials. Unpublished PhD thesis. University of Manchester.

McEnery, T., & Hardie, A. (2011). Corpus linguistics: Method, theory and practice. Cambridge University Press.

Molavi, A., Koosha, M., & Hosseini, H. (2014). A comparative corpus-based analysis of lexical bundles used in EFL textbooks. *Latin American Journal of Content and Language Integrated Learning, 7*(1), 66–81.

MyTCAS. (2024, July 16). *Exam structure and sample exams*. https://www.mytcas.com/blueprint/ tgat1-91/

Phoocharoensil, S. (2020). A genre and collocational analysis of consequence, result, and outcome. *3L: Southeast Asian Journal of English Language Studies*, *26*(3).

Priyatno, A., Dinda, O. Y., Nugraheni, W., & Utami, W. (2023). Lexical bundles in Indonesian EFL textbooks: A corpus analysis. *Journal of Language and Education*, *9*(2), 25–39.

Römer, U., & Arbor, A. (2009). English in academia: Does nativeness matter? *International Journal of Corpus Linguistics, 14*(2), 257–278.

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. John Benjamins Publishing Company.

Sangfans. (2024). *TGAT English mock exam 2024*. Sangfans. https://www.sangfans.com/test-tgateng1/

Sarjono, R. I. L., Heda, A. K., & Bram, B. (2022). Exploring lexical bundles in Bahasa Inggris textbook: A corpus study. *Jurnal Penelitian Humaniora, 23*(2), 84–94.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.

Schmitt, N. (2010). *Researching vocabulary – A vocabulary research manual*. Palgrave Macmillan.

Scott, M. (2012). WordSmith Tools (Version 6). [Computer software]. Liverpool, UK: Lexical Analysis Software.

Shin, D. (2007). The high frequency lexical bundles of spoken and written English. *English Teaching, 62*(1), 199–218.

Sukman, K., Triwatwaranon, W., Munkongdee, T., & Chumnumnawin, N. (2022). A corpus-based study of lexical bundles of keywords found in online business news articles. *European Journal of English Language Teaching, 7*(3). https://doi.org/10.46827/ejel.v7i3.4275

Thongvitit, S., & Thumawongsa, N. (2017). A corpus-based study of English lexical bundles found in the abstracts of research articles written by Thai EFL writers. *International Journal of Social*

*Science and Humanity, 7*(12), 751–755.

Trinant, K., & Kijpoonphol, B. (2021). lexical bundles in a sample corpus of tourism research articles (SCTRA). *NKRAFA Journal of Humanities and Social Sciences, 9*, 94–108.

Trinant, K., & Yodkamlue, B. (2019). lexical bundles in a sample corpus of nursing research articles (SCNRA). *13*(1), 45–72.

West, M. (1953). *A general service list of English words*. Longman, Green and Company.

Wilkins, A. (1972). *Linguistics and language teaching*. Edward Arnold

## About the Authors

**Pattarin Metang** is a graduate student at the Language Institute, Thammasat University. Her research interests include corpus linguistics, lexical bundle analysis, and English language assessment.

**Arthitaya Narathakoon** is an English lecturer at the Language Institute, Thammasat University. Her research interests include teacher development, teacher training, teacher literacy, and teacher assessment literacy.