Jurnal Penelitian, Pengembangan, Pembelajaran, dan Teknologi (JP3T)

Volume 3 Number 2 (2025)

https://jurnalcendekia.id/index.php/jp3t

Exploring Lexical Demands in the IELTS Reading Test: A Corpus-Based Analysis on the Academic IELTS

Nur Rahma Yuli Rusmita [1]*, Restu Santyarini [2]
[1 2] Universitas Airlangga, East Java, Indonesia
* Corrенspondent Author: nur.rahma.yuli-2023@fib.unair.ac.id

Abstract: Acquiring knowledge of vocabulary is essential for learning a second language, since vocabulary serves as the foundation for the language. In the Academic IELTS, one of the most common English proficiency tests, vocabulary is highly crucial to master, especially in reading. This study uses a corpus-based methodology with *AntWordProfiler* software to analyze 18 IELTS books (Cambridge Practice Tests for IELTS 1996-2023) and identify 246 high-frequency AWL words. Data reduction techniques produce 217 words and filtered into 172 distinct word families. The findings aim to enhance IELTS exam preparation by providing insights into AWL vocabulary, potentially guiding future research to include broader word lists for a more comprehensive understanding.

Keyword: Lexical, Reading, Test, Academic, IELTS

Introduction

The International English Language Testing System (IELTS) is one of the most recognized assessments for measuring English language proficiency. Divided into four elements: listening, reading, writing, and speaking, it plays a pivotal role in evaluating candidates' abilities for academic, professional, and migration purposes (Hashemi & Daneshfar, 2018). Among Indonesian students, IELTS Academic has become increasingly significant due to its requirement in scholarship applications, particularly for one of the most notable ones, LPDP (Lembaga Pengelola Dana Pendidikan), which mandates a minimum score of 6.5 for overseas university admissions (LPDP, 2025). Despite its widespread use, many test-takers face challenges achieving their desired score, which often results in significant costs in preparation courses and materials.

Vocabulary acquisition has emerged as a critical factor in English proficiency tests, particularly in the IELTS Reading section, which is frequently identified as the most challenging. As stated by Kovalenko (2018) and Michell (2022) that limited vocabulary is a key barrier for test takers. Given its importance, teachers, students, and material developers in the English for Academic Purposes (EAP) context must prioritize vocabulary instruction to ensure learners possess an adequate lexical repertoire (Schmitt, 2010). However, determining which words to focus on remains a complex task (Coxhead, 1998). The Academic Word List (AWL), introduced by Coxhead (2000), is widely used in EAP programs but has been criticized for being overly generic and not fully aligning with the lexical demands of IELTS reading texts. This necessitates the creation of more specialized academic word lists tailored to a specific context, such as IELTS.

Vocabulary knowledge is foundational for language learners, particularly for those preparing for academic purposes. Nation (1993) categorizes vocabulary into four types: high-frequency words, academic words, technical words, and low-frequency words. High-frequency words encompass commonly used vocabulary in daily communication and general texts. These words are often compiled in lists like the General Service List (GSL) by West (1953), which includes approximately 2,000 word families and provides significant coverage for various types of texts, including 80% of running words in academic contexts. Such lists enable learners to prioritize essential vocabulary, making their language learning goals more attainable.

Academic words, as distinct from high-frequency vocabulary, are widely used across academic texts but are less common in general English contexts. The Academic Word List (AWL) by Coxhead (2000) includes 570 word families selected for their high frequency in academic texts across various disciplines. These words play a pivotal role in academic success as they bridge general and technical vocabulary, making them critical for students engaging in English for Academic Purposes (EAP). However, while the AWL is beneficial, its generality poses challenges for students preparing for specific tests like IELTS, where the vocabulary demands may not align precisely with the AWL.
Technical words are domain-specific terms relevant to specialized fields such as medicine, engineering, or economics. These words typically constitute about 5% of the total vocabulary in academic texts and vary significantly across disciplines. Lastly, low-frequency words include proper nouns, rare words, and other specialized terms not commonly encountered in everyday language. While their role in academic texts is limited, they can pose challenges for language learners, especially in tests requiring extensive reading comprehension.

The distinction among these vocabulary types underscores the importance of targeted instruction.Coxhead & Nation (2001) emphasize that academic words occur across disciplines, making them universally valuable in EAP settings. However, for learners aiming to excel in specific contexts like the IELTS Academic Reading test, there is a need for refined word lists derived from test-specific corpora. This study addresses this gap by analyzing the lexical demands of the IELTS Academic Reading section to identify high-frequency academic words that are essential for effective preparation.

Numerous studies have addressed the lexical requirements of various English proficiency tests. For instance, Chujo & Oghigian (2009) explored high-frequency vocabulary in TOEFL, TOEIC, and EIKEN, while Jin et al. (2012) examined the General Service List (GSL) and AWL in the TOEFL Reading Test. More discipline-specific lists, such as the Medical Academic Vocabulary List (MAVL) by Lei & Liu (2016) and the Academic Article Word List for Social Sciences (AAWL-SS) by Kwary & Artha (2017), have also been developed. In the context of IELTS, Michell (2022) generated a word list for the Reading section using WordSmith Tools, providing insights into the lexical items frequently encountered in the test. However, gaps remain in understanding the specific lexical demands of the Academic IELTS Reading test, particularly in identifying word families most critical for test preparation.

This study aims to address this gap by utilizing a corpus-based methodology to analyze lexical demands in the IELTS Academic Reading section. Unlike previous studies, this research employs AntWordProfiler 1.5.1 to generate a specialized word list based on 18 Cambridge IELTS practice test books published from 1996 to 2023. The corpus includes texts from three passages in each reading test, providing a comprehensive representation of the vocabulary required. By identifying high-frequency academic words and their distribution across the corpus, this study seeks to offer practical insights for students, educators, and material developers, enabling more targeted vocabulary instruction and preparation strategies for the IELTS Academic Reading test.

Method

This study employed a corpus-based quantitative approach to analyze the lexical demands of the IELTS Academic Reading test, focusing specifically on identifying high-frequency academic word families. The design of this study is situated within the framework of data-driven learning and corpus-informed vocabulary instruction, aiming to generate a specialized wordlist that reflects the authentic linguistic input presented in the IELTS Academic module.

The corpus was constructed using texts from 18 volumes of the *Cambridge Practice Tests for IELTS*, published between 1996 and 2023. Each volume contains four tests, with three reading passages per test, resulting in a total of 12 reading texts per book. In total, 216 reading passages were compiled, representing a longitudinal and representative collection of IELTS Academic Reading materials. This dataset was chosen due to its widespread use in IELTS preparation and its authenticity, as the texts are modeled on actual IELTS exam content and validated by the University of Cambridge. All reading passages were digitized and converted into plain text (.txt) format to ensure compatibility with corpus analysis tools. Non-textual elements such as figures, charts, and task instructions were excluded to maintain textual consistency and analytical clarity.

The primary tool used for lexical profiling was *AntWordProfiler 1.5.1*, a widely recognized software in corpus linguistics developed by Laurence Anthony. AntWordProfiler allows users to classify vocabulary items into predefined lists, such as the General Service List (GSL) and the Academic Word List (AWL). For this study, the first and second 1000 words of GSL (West, 1953) were employed to exclude high-frequency general English vocabulary. The remaining lexical items, not included in the GSL, were considered as candidates for an IELTS-specific academic word list.

The corpus was analyzed using *AntWordProfiler's* range and frequency functions. To identify high-prevalence academic words, two criteria were applied: a minimum range of 5, indicating that a word must appear in at least five different reading passages, and a minimum frequency of 20, signifying the word must occur at least 20 times across the entire corpus. These thresholds were established to ensure that only widely distributed and contextually salient words were included, following methodological guidance from Anwar & Hardjanto (2023).

Result & Discussion

Result

Table 1 provides a summary of the corpus, detailing the number of passages and tokens from each book. The passages were arranged in alignment with the structure of the IELTS Academic module, maintaining the order in which they appear within each test.

Table 1. General Information of the Corpus IELTS

Registers	Passages	Tokens
Academic 1	12	9135
Academic 2	12	9350
Academic 3	12	10261
Academic 4	12	10420
Academic 5	12	10760
Academic 6	12	10270
Academic 7	12	11023
Academic 8	12	10505
Academic 9	12	10549
Academic 10	12	10873
Academic 11	12	10328
Academic 12	12	10041
Academic 13	12	10724
Academic 14	12	10539
Academic 15	12	10585
Academic 16	12	10535
Academic 17	12	10911
Academic 18	12	10751
Total		187560

Source: Reading IELTS 1-18 processed with AntWordProfiler 1.5.1

The analysis using *AntWordProfiler 1.5.1* (Laurence, 2021) began by categorizing the tokens into three groups based on the GSL 1st and 2nd 1000 lists (West, 1953) and the IELTS-specific AWL. After inputting the corpus data, the AWL base word list was excluded to focus on vocabulary specific to IELTS reading texts. Words not classified under the GSL lists were designated as IELTS-specific AWL. Table 2 displays the token distribution of GSL and IELTS-specific AWL.

Table 2. Token ff GSL and AWL of IELTS Reading

Level	File	Token	%
1	1_Gsl_1st_1000.Txt	139832	74.66
2	2_Gsl_2nd_1000.Txt	11225	5.99
0	(IELTS_Awl)	36240	19.35
Total		187297	

Source: Authors

To find the high prevalence base word exclusively, the data were then analyzed and filtered using a minimum range of 5 and a minimum frequency of 20. These criteria indicate that a word must appear in at least 5 reading texts and must appear in all texts at least 20 times. The following table displays a sample list of IELTS reading AWL derived from the IELTS reading corpus.

Table 3. Sample List of IELTS Reading AWL

Words	Range	Frequency
research	85	171
major	65	92
process	62	100
evidence	59	107
environment	56	90
similar	56	68
technology	53	83
area	51	71
researchers	49	96
role	47	109
approach	47	78
areas	46	62
involved	46	59
species	44	181
energy	42	78
individual	42	56
individual	42	56
range	42	58

Source: Authors

An initial dataset of 246 high-frequency words was reduced through meticulous filtering to meet the study's objectives. The reduction process excluded data based on specific criteria (Anwar & Hardjanto, 2023), including: (1) single letters (apart from articles), (2) syllables or word fragments, (3) hyphenated words that were not repeated, (4) personal names, and (5) foreign terms. Following this process, 217 distinct word families were identified.

These 217 word families were then organized following the methodology of the AWL. To further enhance the analysis, word derivatives and inflections were generated using *Familizer*, a program available on the *Lextutor* website (Cobb, 2023). This tool allowed the creation of comprehensive word lists encompassing all inflected and derived forms of each word family. Figure 1 illustrates the *Familizer* interface after data input, while Figure 1 provides a sample of the word member list output.

Figure 1. Familizer Web Page with Data Inserted

Programming the *Familizer* software produced a formatted text file listing all word members in table form. For example, the word "research" and its related forms ("researched," "researcher," "researchers," "researches," and "researching"), were systematically displayed. This step reduced the dataset further, yielding 172 word families from a total of 1,080 tokens. Figure 2 shows a sample of the output of word families made with *Familizer.*

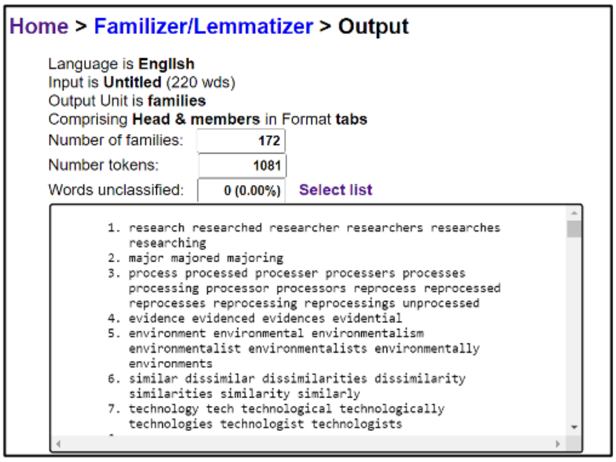

Figure 2. A Sample of the Lists Made with the Familizer

The finalized word list serves as a focused resource for addressing the lexical demands of the IELTS Academic Reading test (see Appendix 1).

Discussion

The findings of this study reveal a substantial lexical demand in the IELTS Academic Reading section, as demonstrated by the identification of 172 word families with high frequency and wide distribution across the corpus. These results align with the current understanding in the field, which consistently highlights limited academic vocabulary as a significant barrier for IELTS test-takers, particularly in the reading section. Michell (2022), for example, emphasized that insufficient mastery of academic lexis is directly correlated with lower performance in reading comprehension tasks. This study also affirms Coxhead's (2000) seminal argument that academic vocabulary plays a pivotal role in academic success, especially for second-language learners preparing for standardized tests.

However, this study makes a noteworthy contribution by showing that not all word families included in the Academic Word List (AWL) are equally relevant or frequently used in IELTS reading texts. Only a subset of these families appeared with sufficient frequency and range to be considered pedagogically essential. This finding supports critiques raised by Lei and Liu (2016) and Kwary and Artha (2017), who argue that general academic word lists may not adequately capture the specific lexical demands of test-specific contexts such as IELTS. Therefore, adapting the AWL based on corpus-informed evidence is essential to ensure more effective and focused vocabulary instruction.

Unlike previous studies that primarily focused on lexical coverage percentages, this research adopted a more refined approach by filtering vocabulary based on minimum frequency and distribution thresholds. This method allowed for the identification of high-yield word families that are realistically teachable and practically relevant. This distinguishes the present study from earlier research by Chujo and Oghigian (2009) and Jin et al. (2012), which examined vocabulary in TOEFL and TOEIC but did not provide test-specific wordlists filtered for both frequency and range. Thus, the approach used in this study adds methodological value and serves as a model for designing more targeted and efficient vocabulary instruction.

Furthermore, the use of tools such as AntWordProfiler and Familizer strengthens the methodological rigor of this study. These tools enabled systematic classification of derived and inflected word forms, ensuring that the final word list represents not only frequent lexical items but also their full morphological range. This reflects contemporary trends in corpus linguistics, where word family analysis is favoured for its ability to capture both syntagmatic and paradigmatic relationships within texts. Such depth of analysis is especially relevant in high-stakes assessments like IELTS, where comprehension depends on more than just word recognition.

The pedagogical implications of these findings are significant. Teachers, IELTS instructors, and materials developers can use the 172 identified word families as a foundation for designing vocabulary lessons, reading materials, and practice tests that are closely aligned with the actual demands of the IELTS Academic Reading test. The list is not only empirically validated but also contextually relevant, offering a much-needed solution for learners, particularly in regions like Indonesia, where IELTS preparation is often tied to competitive scholarship programs such as LPDP.

Conclusion

The preparation for the IELTS exam, particularly the reading section for academic IELTS, is undeniably tough for both teachers and students. The challenge for teachers arises in determining the specific vocabulary that students will encounter during the examination. A corpus-based study of IELTS academic reading texts was conducted to identify the most frequently occurring words from the Academic Word List in these texts. The findings revealed a total of 246 high frequency words. The study excluded words that consisted of single letters (excluding articles), syllables or word fragments, hyphenated words, personal names, and foreign keywords. In addition, a total of 217 words were identified following the process of data reduction, and these words were further classified into 172 distinct word families. To enhance the findings of this research, future investigations could broaden the scope by expanding the corpus parameters to encompass more word lists. This would require not just prioritizing keywords from the Academic Word List (AWL) but also integrating words from the General Service List (GSL). Despite the limitations, the researcher is confident that this study can improve IELTS exam preparation courses, primarily by offering useful insights into the Academic Word List (AWL) vocabulary.

References

Anwar, A. D., & Hardjanto, T. D. (2023). High Frequency Words in The Reading Section of TOEFL PBT Exercises. *Protasis: Jurnal Bahasa, Sastra, Budaya, Dan Pengajarannya*, *2*(1), 204–214. https://doi.org/10.55606/protasis.v2i1.90
Cambridge practice tests for IELTS 1. (1996). Cambridge: Cambridge University Press
Cambridge practice tests for IELTS 2. (2000). Cambridge: Cambridge University Press
Cambridge practice tests for IELTS 3. (2002). Cambridge: Cambridge University Press.
Cambridge practice tests for IELTS 4. (2005). Cambridge: Cambridge University Press.
Cambridge practice tests for IELTS 5. (2006). Cambridge: Cambridge University Press.
Cambridge practice tests for IELTS 6. (2007). Cambridge: Cambridge University Press.
Cambridge practice tests for IELTS 7. (2009). Cambridge: Cambridge University Press.
Cambridge practice tests for IELTS 8. (2011). Cambridge: Cambridge University Press.
Cambridge practice tests for IELTS 9. (2013). Cambridge: Cambridge University Press.
Cambridge practice tests for IELTS 10. (2015). Cambridge: Cambridge University Press.
Cambridge practice tests for IELTS 11. (2016). Cambridge: Cambridge University Press.
Cambridge practice tests for IELTS 12. (2012). Cambridge: Cambridge University Press.
Cambridge practice tests for IELTS 13. (2018). Cambridge: Cambridge University Press.

Cambridge practice tests for IELTS 14. (2020). Cambridge: Cambridge University Press
Cambridge practice tests for IELTS 16. (2021). Cambridge: Cambridge University Press
Cambridge practice tests for IELTS 17. (2022). Cambridge: Cambridge University Press.
Cambridge practice tests for IELTS 18. (2023). Cambridge: Cambridge University Press.
Chujo, K., & Oghigian, K. (2009). How Many Words Do You Need to Know to Understand TOEIC, TOEFL & EIKEN? An Examination of Text Coverage and High Frequency Vocabulary. *The Journal of AsiaTEFL*, *6*(2), 121–148.
Coxhead, A. (1998). Academic Word List An Introduction to the. *Word Journal Of The International Linguistic Association*, 1–9.
Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, *34*(2), 213. https://doi.org/10.2307/3587951
Coxhead, A., & Nation, P. (2001). The specialised vocabulary of English for academic purposes. *Research Perspectives on English for Academic Purposes*, *1971*, 252–267. https://doi.org/10.1017/cbo9781139524766.020
Hashemi, A., & Daneshfar, S. (2018). A Review of the IELTS Test: Focus on Validity, Reliability, and Washback. *IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics)*, *3*(1), 39. https://doi.org/10.21093/ijeltal.v3i1.123
https://lpdp.kemenkeu.go.id/. (2023, November 24). Retrieved from https://lpdp.kemenkeu.go.id/: https://lpdp.kemenkeu.go.id/storage/beasiswa/kebijakan-umum/file/public_policy_file_1686560352.pdf
https://www.idp.com/. (n.d.). Retrieved December 24, 2023, from IDP Indonesia: https://www.idp.com/indonesia/ielts/prepare-for-ielts/about-score/?lang=en
IELTS. (2023, November 17). Retrieved from ielts.org: https://ielts.org/about-ielts
Jin, N. Y., Seng Tong, C., Binti Mohamed Nor, M., Ariff Ahmad Tarmizi, M., & Fairus Nor Mohamad, A. (2012). Corpus Based Analysis of the TOEFL Course Books: What are the Words We Should Teach Our Students? *International Review of Social Sciences and Humanities*, *3*(2), 152–160. www.irssh.com
Kovalenko, M. (2018). The validation process in the IELTS reading component: Reading requirements for preparing international students. *Journal of Language and Education*, *4*(1), 63–78. https://doi.org/10.17323/2411-7390-2018-4-1-63-78
Kwary, D. A., & Artha, A. F. (2017). The Academic Article Word List for Social Sciences. *MEXTESOL Journal*, *41*(4), 1–11.
Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, *22*, 42–53. https://doi.org/10.1016/j.jeap.2016.01.008
Michell, C. (2022). *Using Corpus Linguistics to Better Prepare Students for the IELTS Reading Exam*. 367–384. https://doi.org/10.1007/978-981-16-8888-1_23
Nation, I. S. P. (1993). Measuring readiness for simplified material: a test of the first 1,000 words of English. *Simplification: Theory and Application. RELC Anthology Series No. 31*, *31*, 193–203.
Schmitt, N. (2010). *The Developing Language Learner An Introduction To Exploratory Practice-Dickins Program Evaluation In Language Education Online Communication In Language Learning And Teaching*.
West, M. (1953). *General Service List of English Words.* London: Longman.