The JALT CALL Journal
ISSN 1832-4215, 22, 103271 (2026)
https://doi.org/10.29140/jaltcall.2026.103271

Castledown

Academic vocabulary use in research publications in the AI era: human-authored vs AI-mediated texts

Piyapong Laosrirattanachai
Department of Service Innovation and Intercultural Communication, Faculty of Hospitality Industry, Kasetsart University, Thailand
piyapong.l@ku.th
Piyanuch Laosrirattanachai* (Corresponding author)
Department of Service Innovation and Intercultural Communication, Faculty of Hospitality Industry, Kasetsart University, Thailand
piyanuch.la@ku.th

This study addresses an underexplored dimension of computer-assisted language learning (CALL) by developing the ChatGPT-Mediated Academic Word List for Research Publication (CAWL-RP), an exploratory, corpus-specific lexical profile rather than a general-purpose academic word list, and by examining how AI-mediated rewriting may recalibrate academic vocabulary within one controlled corpus design. Drawing on a 16.9-million-token corpus, the data were systematically processed through de-authorship, generalisation, and re-academisation to simulate AI-supported research writing. Subsequently, a set of systematic procedures comprising frequency, range, uniformity, dispersion, lexical profiling, a meaning-based approach, and concordance line analysis was then employed to construct the CAWL-RP, which comprises 869 word families and 2,341 types. Comparative analyses with the Academic Word List, the New Academic Word List, and the Academic Vocabulary List reveal both convergence and divergence. While the CAWL-RP shares a substantial lexical core with established lists, 470 items occur exclusively in the AI-mediated corpus. Moreover, the CAWL-RP is characterised by a markedly higher proportion of general, high-frequency academic vocabulary, whereas the comparison lists contain relatively more low-frequency and field-specific items. Rather than merely documenting lexical variation, the study reconceptualises academic vocabulary as a dynamic construct that may be partially reconfigured under AI-mediated conditions. By situating the CAWL-RP within CALL and English for Academic Purposes theory, the findings empirically demonstrate how generative AI may operationalise academic lexis differently from human-authored corpora. Because the CAWL-RP is derived from a constructed AI-mediated rewriting corpus based on human-authored 2020 source texts and refined under a single model (GPT-3.5) and a fixed prompt-and-session protocol, it is best interpreted as a provisional lexical profile of this specific corpus design rather than a generalisable inventory comparable in status to the AWL, NAWL, or AVL. Within this bounded scope, the study contributes to ongoing debates on how academic vocabulary should be defined, taught, and operationalised in AI-supported learning environments.

Keywords: academic word list, academic vocabulary, ChatGPT, research publications, generative AI

Introduction

English for Academic Purposes (EAP) focuses on the academic English required for scholarly communication. It plays a central role in teaching, learning, and research across higher education. Within EAP, English for Research Publication Purposes (ERPP) equips scholars with the linguistic tools needed to write and publish research articles (Charles, 2013). This function is reinforced by the status of English as the global academic lingua franca, which facilitates international collaboration and knowledge exchange. As universities compete internationally and use research output as a key performance indicator, international publication has increasingly become a requirement for postgraduate qualifications and academic advancement—further entrenching English at the centre of academic discourse (Flowerdew, 2013; Meebangsai et al., 2023).

Against this backdrop, the rapid integration of generative artificial intelligence (AI) is transforming academic writing practices. Large language models (LLMs) are reshaping how academic texts are produced, influencing both the composition process and the expectations of academic discourse (Praphan & Praphan, 2023; Warschauer et al., 2023). These tools are now used by both L1 and L2 writers to generate, revise, paraphrase, and enhance texts within EAP and ERPP settings (Mo & Crosthwaite, 2025). Among them, ChatGPT has attracted particular attention for supporting multiple stages of the writing process, from initial drafting to final revision (Barrot, 2023; Su et al., 2023). As part of OpenAI's GPT series, it offers near-instant access to extensive academic vocabulary and collocations, thereby shaping how learners encounter, internalise, and apply academic language. This technological shift offers opportunities for efficiency and feedback, yet it also raises questions about standards, authorship, and the evolving conventions of academic style.

Reflecting ChatGPT's growing presence, research has begun to compare its linguistic output with that of human writers. Oh and Lee (2024) explored ChatGPT's role as an epistemic agent in science education, assessing its capacity to participate in discussions typically associated with human epistemic agency. In a complementary direction, Jiang and Hyland (2024) analysed three-word lexical bundles in essays by ChatGPT and British students, finding that ChatGPT relied on a more limited and repetitive range of bundles, with fewer markers of epistemic stance and a weaker authorial presence. Similarly, Zhang and Crosthwaite (2025) observed that, although ChatGPT generated technically appropriate and formal vocabulary, human writers demonstrated greater contextual richness and lexical variation. Extending this line of inquiry, Mo and Crosthwaite (2025) compared stance and engagement features across texts produced by ChatGPT, MetaAI, ERNIEBot, and human writers using Hyland's (2005) framework. Moreover, studies evaluating ChatGPT's performance in producing academic abstracts further underscore its growing influence in academic settings (Casal & Kessler, 2023; Kong & Liu, 2024; Zhang & Zhang, 2025). Taken together, these findings suggest the need for systematic investigation of how AI-mediated academic language aligns with, or departs from, human-established norms.

A central determinant of effective academic writing—whether authored by humans or AI—is academic vocabulary. Mastery of such vocabulary is a robust predictor of successful academic writing, with academic words typically comprising 10–14% of academic texts (Chung & Wan, 2025; Csomay & Prades, 2018; Masrai & Milton, 2018; Pecorari et al., 2019). Within the framework of English for General Academic Purposes (EGAP), academic vocabulary refers to words that occur frequently across disciplines but are not field-specific (Coxhead, 2020). Although these items may not encode the central content of a text, they are indispensable for structuring argumentation, managing stance, and signalling relationships between ideas, thereby supporting effective scholarly communication (Charles & Pecorari, 2016).

Since the publication of the Academic Word List (AWL) (Coxhead, 2000), the identification, teaching, and learning of academic vocabulary—particularly for L2 learners—have received sustained attention. The AWL, a cross-disciplinary inventory of high-frequency academic items, has become foundational in pedagogy and research, underpinning academic literacy instruction. To reflect evolving use and broaden coverage, newer lists have been proposed; notably, the New Academic Word List (NAWL) (Browne et al., 2013) and the Academic Vocabulary List (AVL) (Gardner & Davies, 2014) are widely regarded as a reliable and up-to-date reference reflecting contemporary usage. However, these lists are constructed on corpora composed predominantly of human-authored academic texts, thereby implicitly establishing human production as the normative baseline of academic vocabulary. In the era of AI-mediated writing, this assumption warrants reconsideration. Despite the increasing educational relevance of AI-driven writing tools, it remains unclear to what extent AI-assisted vocabulary in systems such as ChatGPT aligns with—or diverges from—the academic vocabulary codified in the AWL, NAWL, and AVL. Although a growing body of research has examined linguistic and lexical features of AI-mediated texts—including lexical bundles, stance markers, and discourse patterns (e.g., Jiang & Hyland, 2024; Mo & Crosthwaite, 2025)—these studies have primarily focused on localised or feature-specific analyses rather than the systematic modelling of academic vocabulary at scale. In particular, while previous research has identified recurring lexical patterns in AI-mediated writing, the construction of an academic word list derived from AI-mediated academic discourse remains underexplored. This gap is theoretically significant, as it raises questions about whether academic vocabulary, traditionally defined based on human-authored corpora, remains stable or is being recalibrated in AI-mediated writing environments.

It is important to clarify the nature of the data examined in the present study. While previous research has often focused on fully AI-generated texts, the corpus constructed here represents AI-mediated academic discourse. Specifically, original human-authored research articles were transformed through a controlled multi-stage rewriting process using ChatGPT. As such, the resulting texts should not be interpreted as spontaneously generated AI output, but rather as reconstructed academic discourse shaped by both human source content and AI-driven rearticulation. This distinction is crucial, as it positions the study within the broader context of human–AI collaborative writing rather than purely AI-authored production.

The present study addresses this gap by investigating whether the academic vocabulary revised using ChatGPT corresponds with—or differs from—that recorded in the AWL, NAWL and AVL. By focusing on vocabulary selection, it evaluates the extent to which AI-mediated language conforms to established academic norms. This issue is particularly salient for L2 academic writing, where increasing reliance on AI may influence learners' exposure to—and acquisition of—academic vocabulary, thereby affecting their ability to engage effectively with academic discourse.

Identifying academic vocabulary in AI-mediated texts, however, requires a principled approach. As with the AWL, NAWL, and AVL, it is neither feasible nor desirable to include every lexical item that appears in AI-mediated research articles or corpora. Accordingly, the present study introduces the ChatGPT-Mediated Academic Word List for Research Publication (CAWL-RP). It should be stated at the outset that the CAWL-RP is not proposed as a new academic word list in the same conceptual class as the AWL, NAWL, or AVL. Those lists are derived from large, naturally occurring corpora of human-authored academic writing accumulated over years of use. The CAWL-RP, by contrast, is derived from a single constructed corpus in which human-authored 2020 research articles were deliberately rewritten through a fixed ChatGPT (GPT-3.5) prompting protocol. It is therefore more accurately described as an exploratory, corpus-specific lexical profile of one AI-mediated corpus design — one that is informative precisely because it is tightly controlled, but not because it is broadly representative of AI-assisted academic writing in general. The CAWL-RP aims to capture frequently occurring academic vocabulary within a corpus of research articles refined using ChatGPT. By constructing this list, the study seeks to identify a distinct lexical set that may differ from existing word lists and, in so doing, illuminate how academic language is being affected by emerging AI technologies. From a CALL perspective, this study extends current research by shifting attention from AI as a pedagogical tool to AI as a source of linguistic input that may shape learners' lexical development. As learners increasingly rely on generative AI for drafting and revising academic texts, the vocabulary they are exposed to is no longer derived solely from human-authored corpora but also from AI-mediated discourse. Understanding the lexical profile of such input is therefore essential for informing CALL pedagogy, particularly in relation to vocabulary instruction, materials design, and learner awareness of AI-influenced academic language. The next section will present a detailed review of the literature on AI and language learning, as well as the methodology and criteria for developing word lists.

Literature review

Artificial Intelligence in CALL: from pedagogical integration to AI-specific lexical research

Within the field of Computer-Assisted Language Learning (CALL), generative AI has rapidly emerged as a transformative development. Early discussions frame ChatGPT as an extension of intelligent CALL systems, capable of providing adaptive feedback, dialogic interaction, and scaffolded language support (Thakkar, 2024). More recent theoretical perspectives position generative AI as a co-constructive partner in digitally mediated learning environments, reshaping learner engagement while maintaining the centrality of human pedagogical oversight (Oh & Lee, 2025). Empirical research in CALL contexts demonstrates that AI-assisted environments enhance multiple language skills, including reading, writing, speaking, listening, and vocabulary acquisition (Butarbutar et al., 2026). Learners also report positive perceptions of AI-integrated corpus-based instruction for addressing vocabulary and comprehension challenges (Oktavianti et al., 2026). These findings suggest that AI not only mediates instruction but also shapes linguistic output within CALL environments.

A second strand of CALL research focuses on AI-supported vocabulary development. Experimental studies consistently report improvements in lexical retention, engagement, and self-regulated learning when AI tools are integrated into instruction (Behforouz & Al Ghaithi, 2025; Namaziandost & Çakmak, 2025; Yu, 2026). Game-based and mobile AI systems likewise demonstrate positive vocabulary gains in technology-enhanced learning contexts (Ahmed et al., 2026; Wen et al., 2025). In addition, AI has been explored as a tool for lexicographic enhancement within digital learning ecosystems. Generative AI can assist in producing learner-oriented grammatical explanations and lexical definitions, although human validation remains necessary (Fuertes-Olivera, 2024; Li & Tarp, 2024). Across these studies, AI is conceptualised primarily as a pedagogical scaffold within CALL rather than as an independent source of linguistic patterns.

More linguistically oriented CALL research begins to examine the lexical characteristics of AI-mediated texts themselves. Studies comparing AI and human academic writing reveal systematic differences in lexical bundles, stance markers, and lexical sophistication (Jiang & Hyland, 2025; Mo & Crosthwaite, 2025; Zhang & Crosthwaite, 2025). Vocabulary profiling analyses further show discrepancies between prompted proficiency levels and the lexical distribution of AI outputs (Uchida, 2025). While these investigations contribute to understanding AI-mediated discourse within CALL, they primarily address surface lexical diversity, bundle frequency, or proficiency alignment.

They do not systematically model academic vocabulary across large, cross-disciplinary AI-mediated corpora.

While CALL research has extensively examined generative AI as a pedagogical tool that enhances vocabulary learning and influences lexical and discourse features in learner writing, there remains limited corpus-based investigation into how AI-mediated academic discourse constructs academic vocabulary as a lexical category. Specifically, no prior CALL-oriented study has developed and evaluated an academic word list derived exclusively from AI-mediated research writing across disciplines. This gap underscores the need to move beyond evaluating AI as an instructional aid and to examine how generative models may reproduce, reshape, or recalibrate the lexical foundations of academic discourse within AI-mediated CALL environments.

The development of academic word lists

Numerous studies have investigated the vocabulary required for academic study, leading to the development of academic word lists that prioritise high-utility, frequently occurring items in academic English. These lists aim to support learners and educators by identifying vocabulary central to academic communication.

Research on academic vocabulary development has evolved from relatively small-scale, manually compiled lists to large-scale corpus-driven approaches. Early studies (e.g., Campion & Elley, 1971; Ghadessy, 1979; Lynn, 1973; Praninskas, 1972) primarily relied on frequency counts and pedagogical intuition, which, while useful, lacked systematic coverage and replicability. The subsequent development of the Academic Word List (AWL) (Coxhead, 2000) marked a methodological shift towards a more principled corpus-based approach, incorporating frequency and range across disciplines to identify a common core of academic vocabulary. More recent lists, such as the New Academic Word List (NAWL) (Browne et al., 2013) and the Academic Vocabulary List (AVL) (Gardner & Davies, 2014), have further refined these methodologies by expanding corpus size, updating lexical coverage, and incorporating more sophisticated statistical criteria, including dispersion and keyness. Despite these advancements, however, existing lists share a fundamental assumption: they are derived exclusively from human-authored corpora and therefore reflect human production as the normative basis of academic vocabulary.

This assumption has not been critically examined in the context of AI-mediated writing. While methodological refinements have improved the representativeness and coverage of academic word lists, they have not addressed how emerging forms of text production—particularly those involving generative AI—may influence the composition of academic vocabulary itself. As such, the extent to which existing word list methodologies remain applicable to AI-mediated discourse remains an open question.

Methods for developing word lists

The development of academic word lists is typically guided by a set of core methodological considerations, including corpus design, unit of analysis, and selection criteria such as frequency, range, and dispersion (Nation, 2016; Paquot, 2010). While these criteria are widely accepted, their implementation varies considerably across studies, reflecting differences in research aims, corpus composition, and target users. A key issue in the literature concerns the balance between inclusivity and specificity. Frequency-based approaches risk over-representing general high-frequency words, whereas stricter criteria—such as dispersion and specialised occurrence—may exclude lexically important items that function differently across contexts (Gardner & Davies, 2014). More recent approaches, including keyword analysis (Laosrirattanachai & Laosrirattanachai, 2025) and meaning-based classification (Gong et al., 2025), attempt to address these limitations by incorporating statistical and semantic dimensions of vocabulary use.

An effective word list combines quantitative metrics with qualitative methods rather than relying on a single approach. This research aims to identify potential differences between human-authored and AI-assisted academic texts. By identifying these discrepancies, the findings can help refine AI algorithms to align with academic standards, enhancing the quality of AI-mediated content. As ChatGPT's use in academic writing becomes more common, its vocabulary will increasingly appear in research articles. Therefore, L2 learners should prepare to recognise this emerging vocabulary to improve their comprehension of academic texts. To achieve this, the following research questions will be addressed:

1. How was the ChatGPT-Mediated Academic Word List for Research Publication (CAWL-RP) developed and refined through a multi-stage filtering process?
2. To what extent do the lexical items in the CAWL-RP overlap with those in the AWL, NAWL, and AVL?
3. How do the lexical characteristics of the CAWL-RP differ from those of the AWL, NAWL, and AVL in terms of lexical level and frequency band?

Research methodology

Data preparation and corpus compilation

This study aimed to develop a word list of academic vocabulary typically employed in the writing of research articles edited by ChatGPT. To ensure the linguistic data covered a wide range of academic disciplines, a total of 2,700 research articles were collected from journals indexed in Scopus, representing all 27 broad subject areas classified by Scopus (see scopus.com). Specifically, 100 articles were randomly selected from each subject area. To ensure that the collected articles had not been refined by ChatGPT, only articles published in the year 2020 were included, as this predated the public release of ChatGPT.

To generate a corpus that reflects academic language output authentically produced by ChatGPT, a three-stage transformation process was employed. First, the authorship section and the references were removed. Second, each of the 2,700 original research articles compiled in the previous step was rewritten into general English using ChatGPT. The version of ChatGPT being used is GPT-3.5, as it is the free model available. This ensures that everyone can access it, making it more likely to be used than other versions. This step was crucial to avoid simple paraphrasing or lexical substitutions that might occur if ChatGPT were prompted to convert original academic texts directly into ChatGPT-mediated academic English, thereby ensuring that the lexical choices in the final output were characteristic of ChatGPT's generative capacity rather than influenced by the original lexical content.

In the third stage, the general English versions of the articles were input into ChatGPT, prompting it to regenerate them in academic English. To preserve data integrity, each transformation (from academic English to general English and from general English to ChatGPT-mediated academic English) was performed in a new chat session using different ChatGPT accounts. This measure was taken to prevent the system from drawing on prior conversational memory, thus avoiding potential contamination of the refined output. The three-stage transformation was deliberately designed to minimise lexical carryover from the original texts and to isolate the lexical reconfiguration introduced by generative AI. Direct paraphrasing of academic articles into AI-mediated academic English would risk preserving lexical traces from the source corpus, thereby obscuring the model's generative tendencies. The procedures of de-authorship, generalisation, and re-academemisation were therefore implemented to attenuate lexical priming effects (Hoey, 2005) and reduce residual disciplinary phrasing while preserving semantic continuity. By maintaining comparable propositional content while altering the discursive producer, the design enables a controlled contrastive analysis of how academic vocabulary and discourse patterns are reconstructed when mediated by a large language model. It is therefore important to clarify that the CAC-RP does not represent spontaneously generated AI discourse; rather, it constitutes a mediated reconstruction of human-authored research writing under controlled conditions. In this sense, the CAC-RP should be understood as AI-mediated academic discourse rather than purely AI-originated text, and this methodological framing should be taken into account when interpreting the findings. Furthermore, the refined academic English texts were required to closely match the original source texts in terms of word count. This constraint was implemented to minimise variation in text length and ensure comparability across original and refined versions. Upon completion of the data preparation process, the resulting corpus, called the ChatGPT-Mediated Academic Corpus for Research Publication (CAC-RP), consisted of 16,992,881 tokens and 245,847 types. The degree of lexical restructuring introduced by this transformation strategy was subsequently examined through an independent keyword analysis comparing the original and transformed corpora. As reported in the Results section, the findings indicate substantial lexical divergence between the two corpora, providing evidence that the transformation procedure produced measurable lexical restructuring relative to the source texts while preserving overall propositional content. This evidence should be read as an indication of lexical reconfiguration accompanying the transformation process, rather than as confirmation that the resulting corpus is fully de-linked from the source texts.

The three-stage transformation design employed in this study may appear unconventional, as it does not aim to reproduce either naturally occurring human-authored academic texts or fully autonomous AI-generated texts. Instead, it is deliberately constructed to model AI-mediated academic discourse under controlled conditions. By transforming human-authored texts into general English and subsequently re-academising them using ChatGPT, the design seeks to minimise direct lexical transfer while preserving propositional content. This approach enables a more controlled examination of how academic vocabulary is reconfigured when mediated by a large language model. While the resulting corpus does not represent spontaneous AI-generated writing, it captures a form of hybrid discourse that increasingly characterises real-world academic writing practices, where human input and AI assistance interact. The three-stage transformation design inevitably yields a hybrid construct. Whilst the intermediate generalisation stage was rigorously implemented to disrupt surface-level lexical carryover, the underlying conceptual architecture, rhetorical move structures, and propositional logic remain inherently anchored to the original human-authored texts from 2020. Consequently, the CAC-RP must be theorised not as an autonomous representation of pure AI discourse, but as a constrained collaborative output where the human provides the cognitive blueprint and the AI dictates the lexical surface realization. This hybridity is a core epistemological characteristic of the data that bound the generalisability of the CAWL-RP.

Prompt design and control procedures

To ensure procedural transparency and replicability, the prompts used in the three-stage transformation process were standardised across all texts. In the first transformation stage (academic to general English), the following prompt was consistently applied: "*Rewrite the following academic text in clear and simple general English while preserving its original meaning. Avoid specialised terminology and maintain coherence. Ensure that the rewritten text is of approximately similar length to the original, without omitting or adding substantive information.*"

In the second stage (general English to academic English), the prompt used was: "*Rewrite the following text in formal academic English suitable for research publication. Enhance clarity, coherence, and academic tone while preserving the original meaning. Ensure that the rewritten text is of approximately similar length to the original, without omitting or adding substantive information.*"

All texts were processed using ChatGPT (GPT-3.5) via the standard web interface, utilising default system settings as user-level control over parameters was not available. To ensure consistency across the dataset, each text was processed in a separate session without any iterative prompting or subsequent refinement. Regarding this procedure, it must be explicitly acknowledged that absolute algorithmic reproducibility remains constrained due to the black-box nature of commercial large language models. Because the data collection relied on a standard web interface, fine-grained parameters such as temperature, Top-P, and frequency penalties could not be statically locked. Given that these models operate on stochastic probability, slight lexical variations may occur upon identical replication attempts. However, while the lack of fine-grained parameter control limits exact reproducibility, the rigorous use of standardised prompts and controlled session procedures enhances methodological transparency, thereby allowing for reasonable replication in comparable research environments.

Keyword analysis of lexical restructuring

To examine the extent to which the three-stage transformation procedure reduced direct lexical transfer from the source texts, an additional keyword analysis was conducted. Two corpora were compared: the Original Research Article Corpus (ORAC), comprising the 2,700 human-authored research articles collected for this study (16,928,946 tokens), and the ChatGPT-Mediated Academic Corpus for Research Publication (CAC-RP).

Keywords were extracted independently from each corpus using the British National Corpus (BNC) as the reference corpus. Following established corpus-linguistic practice, keywords were ranked according to log-likelihood values, and the top 1,500 keywords from each corpus were selected for comparison. The analysis focused on two indicators. First, keyword overlap was examined to determine the extent to which salient lexical items were shared between the two corpora. Second, corpus coverage was calculated by determining the proportion of the original corpus accounted for by the keyword sets derived from the original corpus and from the CAC-RP respectively.

These procedures were employed to evaluate whether the transformation process resulted in substantial lexical restructuring rather than simple lexical preservation. If lexical attenuation occurred, the keyword inventories of the two corpora would be expected to diverge and the keyword set derived from the CAC-RP would be expected to account for a smaller proportion of the original corpus than the keyword set derived from the original corpus itself.

Ethical considerations and data use

The corpus used in this study was compiled from publicly available research articles indexed in Scopus. The texts were used solely for research and analytical purposes and were not redistributed in their original or transformed forms. All transformations were conducted to generate aggregated linguistic data rather than to reproduce or republish individual texts. The study adheres to principles of fair use for academic research, as the data were processed for non-commercial purposes and analysed at the corpus level. No identifiable or proprietary content from individual articles is reported. Furthermore, the transformed texts do not allow reconstruction of the original sources, as they were substantially modified through a multi-stage rewriting process. While formal copyright clearance was not obtained for individual texts, the use of large-scale textual data for linguistic analysis is consistent with established practices in corpus-based research. Nevertheless, this issue remains an important consideration, and future studies may benefit from incorporating licensed or open-access datasets to further strengthen compliance with data use regulations.

Data analysis

Word list development. To develop the ChatGPT-mediated Academic Word List for Research Publication (CAWL-RP), the CAC-RP was systematically analysed based on a set of criteria, namely frequency, range, uniformity, dispersion, specialised occurrence, and removal of irrelevant words. Two main tools were employed for this purpose: the #LancsBox X programme (Brezina & Platt, 2024) for frequency-related analyses and dispersion calculations, and AntWordProfiler (Anthony, 2025a) for lexical profiling.

The frequency threshold adopted in this study was derived through a corpus normalisation approach rather than a simple proportional calculation. This was achieved by standardising the frequency metric to a baseline of words per million (WPM) to ensure direct comparability with Coxhead's (2000) original benchmark of 100 occurrences in a 3.5-million-word corpus (approximately 28.57 WPM). Given that the CAC-RP comprises approximately 17 million tokens, the selection threshold was adjusted to 485 occurrences to maintain a virtually identical normalised density of approximately 28.53 WPM across the datasets. Whilst standardising frequency profiles based on a normalised million-word baseline is a recognised practice in corpus-based word list development to compare datasets of different sizes (e.g. Gardner & Davies, 2014; Laosrirattanachai & Ruangjaroon, 2021; Valipouri & Nassaji, 2013), it assumes a linear scaling relationship that may not fully reflect the non-linear Zipfian distribution inherent in natural language. To safeguard methodological validity and mitigate the limitations of this linear assumption, this normalised threshold was not utilised in isolation, nor was it intended as a fixed universal cut-off. Instead, its primary purpose was to serve as an initial filter ensuring sufficient recurrence, which was then systematically cross-verified against rigorous non-linear criteria—namely multi-disciplinary range, uniformity, and Juilland's *D* dispersion index ($\geq$0.6). The integration of these secondary statistical filters effectively controlled for any distributional skewness or artificial frequency inflation caused by the initial frequency normalisation, thereby enhancing the empirical robustness of the final CAWL-RP.

Building on Coxhead's original range criterion, which required words to occur in at least 15 of the 28 sources, this study broadened the scope by incorporating data from 27 distinct academic disciplines to ensure the general academic nature of the resulting list and to avoid discipline-specific lexical bias. To meet the range requirement, a word had to appear in texts from at least 14 of the 27 disciplines. The range threshold of 14 disciplines (out of 27) was established to ensure that a word is truly 'general academic' rather than discipline-specific. This 50% coverage threshold follows the rigorous standard set by Coxhead (2000), who required words to appear in half of her subcorpora, thus balancing broad utility with lexical inclusivity. In addition, uniformity was assessed to ensure even distribution across disciplines. Whereas Coxhead (2000) stipulated a minimum of 10 occurrences in each of four main disciplinary groupings (Arts, Commerce, Law, and Science), the present study adopted a more inclusive yet balanced threshold: a word was required to occur at least five times within each of the disciplines in which it appeared. The uniformity minimum of 5 occurrences per discipline was adopted to prevent the list from being skewed by a very high frequency in only one or two disciplines. This ensures that the lexical items are consistently 'active' across different domains, providing a more reliable resource for general EAP purposes.

Dispersion was calculated using the statistical method developed by Juilland and Chang-Rodríguez (1964), which has been widely applied in the development of word lists (e.g. Laosrirattanachai & Laosrirattanachai, 2024; Laosrirattanachai et al., 2026; Lei & Liu, 2016; Nation, 2006). The dispersion cut-off of 0.6 was selected based on the precedent set by Dang et al. (2017) in the development of the Academic Spoken Word List. A value of 0.6 is widely considered a robust indicator that a word is spread evenly across a large-scale corpus, effectively filtering out words that appear frequently but are clustered in only a few specific texts (Juilland & Chang-Rodríguez, 1964). However, it is acknowledged that alternative dispersion metrics have been proposed, including Gries's DP and other normalised dispersion indices, which address certain limitations of Juilland's measure, such as sensitivity to corpus size and distribution skewness (Gries, 2008). Despite these alternatives, Juilland's *D* was selected in the present study to maintain methodological comparability with established academic word list research. Moreover, given the large and balanced nature of the CAC-RP across disciplines, the potential distortions associated with this measure are mitigated to some extent. Future research may explore the use of alternative dispersion metrics to further validate and refine AI-derived academic vocabulary lists. Although these parameters (frequency, range, uniformity, and dispersion) involve calibrated judgement, they are aligned with established practices in corpus-based word list development (Nation, 2016; Paquot, 2010).

To ensure that the words included in the CAWL-RP extended beyond the scope of general English, lexical profiling was employed using the GSL as a reference. Any word identified by the previous four criteria but also present in the GSL was excluded from the list. Only those words not found in the GSL were retained for consideration in the next criterion.

Retrieving academic vocabulary hidden in the GSL: Applying the specialised occurrence criterion through lexical profiling with the GSL as the reference list removes many high-frequency general words, thereby avoiding unnecessary repetition for learners. However, it has been criticised for excluding essential academic vocabulary that also appears in the GSL (Gardner & Davies, 2014). Furthermore, academic vocabulary spans high-, mid-, and low-frequency words, and a term's status may shift between academic and general use depending on context (Nation, 2016). To recover excluded items, words meeting the frequency, range, uniformity, and dispersion criteria but removed due to their presence in the GSL were compared with the Meaning-based Academic Vocabulary List (MAVL) (Gong et al., 2025); those found in the MAVL were recalled and combined with words meeting the specialised occurrence criterion for further consideration in the next stage.

Removal of irrelevant words using concordance lines: To ensure that the CAWL-RP consisted of vocabulary items common across academic disciplines but not field-specific within the context of general academic English for publication purposes, irrelevant words were removed. These included symbols used in mathematical and scientific equations (e.g. *α*, *β*, *γ*, *π*), field-specific terminology (e.g. *alloy*, *photovoltaic*, *phosphorylation*), abbreviations (e.g. *QPCR*, *RPM*), and proper nouns (e.g. *Africa*, *Liu*, *Zhang*). This filtering process was carried out by examining the occurrences of words in their contexts through concordance lines in the CAC-RP. Expert-based validation was not incorporated into the present study. This decision reflects the primary aim of the study, which was to establish a reproducible, corpus-based pipeline for deriving academic vocabulary from an AI-mediated corpus, rather than to replicate the subjective expert-review procedures used in the development of earlier word lists. In traditional specialised word list development, subjective review typically relies on professionals with established pedagogical or domain expertise; identifying an equivalent expert capable of judging AI-mediated lexical choices in this relatively new linguistic context was not straightforward, and the study instead prioritised distributional criteria — frequency, range, uniformity, and dispersion — over qualitative lexical judgement at every stage of list development. Strict mathematical filters were applied accordingly, including Juilland's D dispersion index of at least 0.6 and a uniformity threshold requiring a minimum of five occurrences per discipline, to ensure that only stable and widely distributed items were retained. It should be acknowledged, however, that the absence of expert review remains a genuine limitation of the present study, rather than a methodological advantage. Distributional criteria alone cannot fully resolve borderline cases in which a word's academic status is context-dependent, cannot inform judgements about the pedagogical suitability of individual items for language learners, and cannot establish the conceptual validity of the final list — that is, whether the retained items are recognisable as 'academic vocabulary' in a sense meaningful to teachers and learners. This limitation is revisited in the Limitations and Recommendations for Future Studies subsection.

Regarding the unit of word counting for the CAWL-RP, it is important to note that this list is not a technical word list (for which the type unit would be appropriate). However, the target users of the CAWL-RP are highly likely to be individuals aiming to read and write research articles and are therefore expected to have at least an intermediate level of English proficiency (for which the word family unit would be more suitable). Although there has been considerable debate regarding the most appropriate unit, each has its own advantages and disadvantages, offering different benefits to users. Consequently, this study presents the vocabulary in both word family and type units in the final list.

Comparing academic vocabulary of the four word lists. Once the final version of the CAWL-RP had been compiled, it was compared with three widely accepted academic word lists: the AWL, the NAWL, and the AVL. These lists are widely used in both pedagogical and research contexts and are integrated into various corpus analysis tools such as AntWordProfiler and VocabProfile. The unit of word count for all four word lists used in the comparison comprises both the headword of a family and its types, in order to ensure consistency in the unit of word counting. This approach allows the analysis to yield more detailed and insightful results than a comparison based solely on the word family level, as a single family may consist of numerous members, each of which can fall into different lexical levels. Therefore, comparisons should not be made solely on the basis of the headword of the family. It is important to note that this comparison is undertaken for descriptive and exploratory purposes, not to position the CAWL-RP as an equivalent or competing list to the AWL, NAWL, or AVL. The three comparison lists were built from large corpora of naturally occurring human-authored academic writing, whereas the CAWL-RP reflects the lexical output of one AI model operating under a specific rewriting protocol on a constructed corpus. The comparison is therefore best understood as an examination of how vocabulary in one controlled AI-mediated corpus design aligns with, or departs from, vocabulary attested in established human-authored reference lists, rather than a comparison between four lists of equivalent status. The four word lists are compared focusing on two main aspects:

Overlap and divergence in word membership: The vocabulary items contained in the four lists are directly compared in terms of quantity and overlap, highlighting the similarities and differences among the academic vocabulary featured in each list.

Lexical level and frequency band: Using Nation's (2016) framework, lexical level categorises vocabulary into 25 lists based on word frequency in a large corpus. The most frequently occurring words appear in the earliest lists, while less common words are included in the later ones. Schmitt and Schmitt (2014) further classify these into three lexical frequency bands: high (1–3,000), mid (3,001–9,000), and low (over 9,000). Lexical levels and frequency bands are analysed using the VocabProfile (Cobb, 2025). The vocabulary items in the four lists are examined for the proportion of words that fall within the different levels and frequency bands, in order to determine whether, and to what extent, the difficulty levels of the vocabulary in the four lists vary.

Results and discussion

Keyword analysis of lexical restructuring between ORAC and CAC-RP

As the three-stage transformation design was intended to reduce lexical carryover from the source texts, it was necessary to examine the extent to which the transformed corpus retained the same salient lexical profile as the original corpus. To examine this, a keyword analysis was performed comparing the Original Research Article Corpus (ORAC) and the CAC-RP. Keywords were identified independently for each corpus using log-likelihood statistics against the British National Corpus (BNC) as a reference corpus. The top 1,500 keywords from each corpus were then compared in terms of overlap and corpus coverage. If substantial lexical attenuation had occurred, divergence between the keyword inventories and reduced coverage of the original corpus by the CAC-RP keyword set would be expected. Table 1 summarises the findings.

Table 1. Keyword overlap and corpus coverage between ORAC and CAC-RP

Measure	Result
Top keywords compared	1,500
Shared keywords	1,160
Divergent keywords	680
ORAC keyword coverage of ORAC	37.06%
CAC-RP keyword coverage of ORAC	29.20%
Coverage difference	**–7.86 percentage points**

The results revealed that 1,160 keywords were shared between the two corpora, while 680 keywords were unique to one corpus or the other. Although the substantial overlap indicates continuity in academic content and discourse functions, the presence of 680 divergent keywords suggests that considerable lexical restructuring occurred during the transformation process. Further evidence was obtained from the corpus coverage analysis. The 1,500 keywords derived from the ORAC accounted for 37.06% of the original corpus, whereas the 1,500 keywords derived from the CAC-RP accounted for only 29.20% of the original corpus. This represents a reduction of 7.86 percentage points in coverage. If the transformation process had preserved the original lexical profile to a large extent, the two keyword sets would be expected to demonstrate similar coverage patterns. Instead, the lower coverage achieved by the CAC-RP keyword set indicates that a substantial proportion of salient lexical items in the transformed corpus differed from those characterising the original texts.

Taken together, the overlap and coverage findings indicate that the three-stage transformation procedure produced measurable lexical restructuring relative to the source corpus, evident in both the divergent keyword sets and the reduced coverage of the original corpus by the CAC-RP keyword set. This pattern is consistent with the intended aim of reducing direct lexical carry-over; however, it does not, on its own, establish that the resulting corpus is sufficiently de-linked from the source texts to constitute a fully independent basis for the CAWL-RP. Rather, the keyword analysis should be read as supporting evidence of lexical reconfiguration accompanying the transformation process, to be interpreted alongside the corpus-design limitations discussed later in this paper. While lexical continuity remains evident, the results suggest that the transformation process involved meaningful lexical reconfiguration rather than simple paraphrastic preservation of the original vocabulary.

The development of the CAWL-RP as an exploratory lexical profile

The results addressing Research Question 1 demonstrate that the CAWL-RP was developed through a structured, multi-stage filtering process applied to a large-scale corpus of 2,700 research articles rewritten using ChatGPT to enhance academic tone. Comprising 16,992,881 tokens, this corpus provided a substantial linguistic dataset for systematic word selection, the stage-by-stage results of which are detailed in Figure 1. Crucially, as emphasised earlier, the following account of this process describes the derivation of a provisional, corpus-specific lexical profile rather than the construction of a generalisable academic word list.

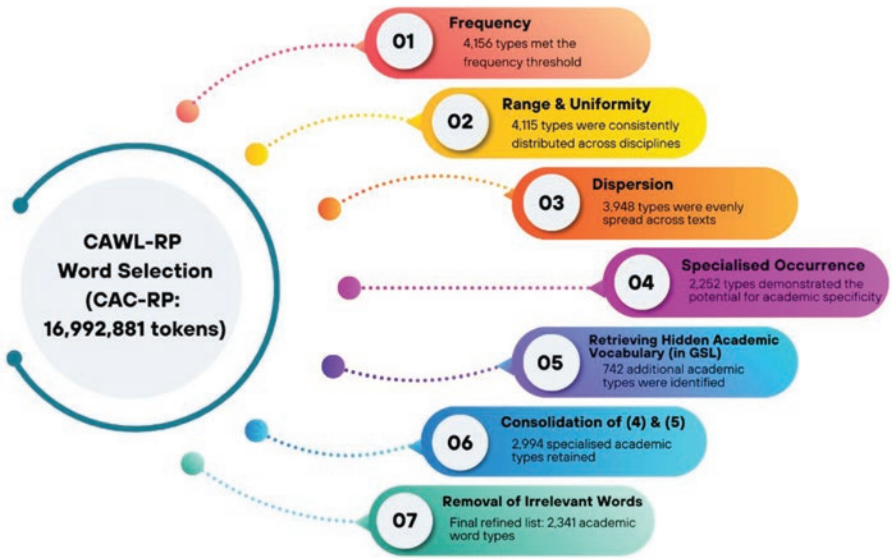

Figure 1. Vocabulary in the CAWL-RP passing each selection criterion

Figure 1 illustrates a multi-stage filtering process, starting with a large corpus and progressively narrowing down the list of lexical items based on a series of rigorous criteria. The colours used in Figure 1 serve solely to visually distinguish the stages of the filtering process and do not imply any hierarchical weighting or statistical significance. One advantage of developing the CAWL-RP lies in the use of multiple criteria, which enhances the credibility of the selected vocabulary and mitigates potential weaknesses that may arise from relying on some criteria.

As can be seen in Figure 1, the frequency stage identified 4,156 words that met minimum occurrence thresholds across the corpus. This set contained essential function words (e.g., *the*, *of*, *and*, *is*) as well as high-frequency academic content words (e.g., *model*, *analysis*, *study*). Words such as *quotations* and *spectrophotometry* were excluded for being too rare or discipline-specific to be of general academic value. While frequency is a foundational metric for word list creation (Vilkaitė-Lozdienė & Schmitt, 2020), this initial count is misleading, as it includes a significant number of common words like *the* and *of* that are not specifically academic.

Next, range and uniformity refined the list to 4,115 words by ensuring items appeared across a wide variety of research articles and fields. Retained words, such as *performance*, *observed*, and *potential*, demonstrated cross-disciplinary applicability, while domain-specific or unevenly distributed terms like *blockchain* and *electrolysis* were excluded. The dispersion stage further reduced the set to 3,948 words by eliminating items disproportionately concentrated in a small subset of documents. Words such as *conducted*, *substantial*, and *enhance* remained due to their balanced presence across the corpus, while terms like *delta* and *synchronization* were disqualified. The subsequent application of range, uniformity, and dispersion criteria is crucial. They ensured that the selected words were not only frequent but also consistently appeared across the corpus, a method strongly advocated by researchers such as Coxhead (2000) and Nation (2016) to ensure a word's broad utility.

In the specialised occurrence stage, the focus shifted towards identifying academic terms with broad research relevance while removing overly general words. The most notable reduction occurred at this stage, which reduced the list by nearly half, from 3,948 to 2,252 word types. This criterion appears to be a key filter for identifying words that are more characteristic of academic discourse, such as *autonomous*, *augmented*, and *equations*, and excluded items such as *importance*, *suitable*, and *exceptional*, which were considered too general. This step effectively distinguished general high-frequency words from those with a more academic flavour, thereby avoiding the pitfall of including irrelevant general words (Scott & Tribble, 2006).

The fifth step focused on retrieving academic vocabulary hidden in the GSL, identifying 742 types that had been overlooked due to their inclusion in the GSL but that nonetheless function with strong academic relevance in the corpus. Words such as *claim*, *findings*, *population*, and *yielded* were reinstated, whereas items like *bottom* and *truth* were excluded. This suggests that some words considered part of a general vocabulary can take on a specific academic sense or be used with particular frequency in academic contexts. This finding supports the 'meaning-based approach' recently proposed by Gong et al. (2025), which aims to distinguish academic from general senses of words. This step demonstrates an awareness that the academic-general distinction is not always clear-cut, a criticism previously levelled at methods like lexical profiling (Gardner & Davies, 2014). Then, a consolidation of steps four and five produced 2,994 candidate types, providing a refined set of academically salient items.

The sixth step involved the removal of irrelevant words, excluding terms with primarily technical, geographical, or scientific specificity—such as *enzyme*, *Africa*, and *metabolism*—while retaining more generalisable academic items such as *accumulation*, *delineate*, and *rigorous*. This process resulted in 2,341 remaining types. Finally, the transition from word types to word families was conducted, producing 869 word families that constitute the final version of the CAWL-RP. This decision is highly pedagogical and aligns with the needs of advanced language learners, who are the target audience for this list, as word families are effective for developing receptive vocabulary skills (Nation, 2016). These 869 word families, together with their 2,341 types, represent the academic vocabulary identified in the present corpus of ChatGPT-mediated research articles re-academised under the specific procedures employed in this study. The CAWL-RP findings on word families and types are revealing. Results show a maximum of 11 types per family, a minimum of 1, and an average of only 3. This indicates that although families may contain many derivational forms, only a few are regularly used in research publications refined using ChatGPT. The pattern highlights the gap between linguistic possibilities and actual usage in academic writing. Pedagogically, this suggests that academic writers, especially L2 learners, benefit more from focusing on a few common types within a family—such as *analysis*, *analyses*, and *analytical*—rather than mastering all variants. This supports previous literature emphasising vocabulary as a predictor of writing success (Charles, 2013; Chung & Wan, 2025; Csomay & Prades, 2018).

The development of the CAWL-RP is a modern contribution to academic vocabulary research, filling a gap in the literature by creating a list from AI-mediated texts. It also validates established methodologies from earlier studies. Its foundation on a specialised corpus of nearly 17 million tokens mirrors the trend toward larger, representative corpora seen in the AWL, NAWL, and AVL, strengthening the credibility of the findings (Browne et al., 2013). The "Specialised Occurrence" parallels the AVL and Academic Spoken Word List (Dang et al., 2017) by distinguishing between academic and non-academic usage, reflecting an adapted corpus-comparison method (Nation, 2016). The "Retrieving academic vocabulary hidden in the GSL" directly engages with debates on academic vocabulary (Gong et al., 2025) by demonstrating that general words can serve specialised academic functions, thereby challenging the AWL's exclusion of GSL words (Coxhead, 2000).

From a CALL perspective, the present study also illustrates what Crosthwaite and Baisa (2023) describe as a useful methodological synergy between generative AI and corpus linguistics. Rather than treating AI as a replacement for corpus-based approaches, the findings demonstrate how large language models can be incorporated into established corpus-linguistic procedures to generate new forms of linguistic evidence. At the same time, the comparison with existing academic word lists highlights the continuing importance of human-authored corpora as reference points for evaluating linguistic authenticity. Consistent with Sinclair's principle of "trusting the text" (Cheung & Crosthwaite, 2025; Sinclair, 2004), AI-mediated language should be examined against naturally occurring language data rather than assumed to represent authentic academic usage automatically. The CAWL-RP therefore represents not only a product of AI-assisted language processing but also an example of how corpus-based scrutiny remains essential within contemporary CALL research.

By culminating in a final list of word families, this research demonstrates a user-centred approach, essential in word list development (Dang, 2020). Using families is appropriate for advanced learners and receptive vocabulary growth, as it enables transfer of knowledge across forms (Nation, 2016). This choice also prepares L2 learners for the increasing role of AI-assisted academic writing, identified as a major emerging challenge in the literature.

Comparing items listed in the AWL, NAWL, AVL, and CAWL-RP

A comparative analysis of the CAWL-RP, AWL, NAWL, and AVL reveals important insights into the evolving nature of academic vocabulary in the age of AI. The results presented indicate that a substantial number of words in the CAWL-RP are also found in existing, human-authored word lists, but a unique core of words remains exclusive to the ChatGPT-mediated corpus.

	Word appearing in				Number of shared words
	AWL	NAWL	AVL	CAWL-RP	
4	✓	✓	✓	✓	186
3	✓	✗	✓	✓	476
3	✓	✓	✓	✗	122
3	✗	✓	✓	✓	133
3	✓	✓	✗	✓	56
2	✓	✗	✗	✓	296
2	✓	✗	✓	✗	731
2	✓	✓	✗	✗	143
2	✗	✗	✓	✓	648
2	✗	✓	✗	✓	73
2	✗	✓	✓	✗	251
1	✓	✗	✗	✗	1,093
1	✗	✗	✗	✓	470
1	✗	✗	✓	✗	4,263
1	✗	✓	✗	✗	1,775

Figure 2. Overlap and divergence of word types across the AWL, NAWL, AVL, and CAWL-RP

Figure 2 shows that a considerable number of words are shared, with 186 types found in all four lists. This suggests a common core of foundational academic vocabulary that transcends different corpora and methodologies, including the human-authored and AI-mediated ones. These are the general academic words that are universally applicable across disciplines (Pecorari et al., 2019). The analysis also highlights a key difference between the NAWL and the other lists. The NAWL contains a significant number of field-specific lexical items, including words such as *yeast*, *vowel*, *accent*, *acid*, *airplane*, *biodiversity*, *ecological*, *lung*, *metaphors*, and *wheat*, which were not identified in the AWL, AVL, or CAWL-RP. These items are typically associated with particular disciplinary domains—for example, *acid, biodiversity, ecological*, and *lung* in the natural sciences; *vowel* and *accent* in linguistics; and *wheat* and *yeast* in agricultural and biological sciences—and therefore function as semi-specialised terminology rather than English for General Academic Purposes (EGAP). Their inclusion suggests that the NAWL operationalises "academic vocabulary" in a broader sense that tolerates greater disciplinary specificity. (See full NAWL list at https://www.linguaeruditio.com/Glossary/NAWL/NAWL_gloss.html.) In contrast, the other three lists—AWL, AVL, and CAWL-RP—are more focused on EGAP words, which are essential for supporting argumentation and scholarly communication regardless of the specific field (Charles & Pecorari, 2016). The presence of such field-specific words in the NAWL might be attributed to the larger and more diverse corpus it was derived from, which likely included more discipline-specific texts.

A notable finding of this study is the identification of 470 word types that occur exclusively in the CAWL-RP, including items such as *absorbance*, *barriers*, *delineates*, *fidelity*, *multimodal*, and *transformative*. This lexical set represents the central focus of the analysis and may reflect distinctive tendencies in AI-mediated academic writing under the specific conditions of this study. Nevertheless, this exclusivity should be interpreted cautiously, as it may be shaped by corpus composition, disciplinary distribution, and the multi-stage preprocessing procedures employed. Accordingly, these items are better understood not as a fixed or "AI-specific" academic vocabulary, but as patterns that emerge in texts mediated by generative AI. While such patterns suggest that AI-assisted writing may influence the distribution and selection of academic vocabulary in particular contexts (Zhang & Crosthwaite, 2025), it would be premature to argue that generative AI is fundamentally reshaping academic vocabulary as a whole, given that academic language is also shaped by broader social, disciplinary, and institutional forces. The findings therefore point to emerging tendencies rather than stable or generalisable lexical change, and further research using alternative corpora and methodologies is needed to examine their extent across different AI systems and settings.

In terms of pedagogical implications, these findings are relevant to EAP and ERPP, particularly as AI tools such as ChatGPT are increasingly integrated into academic writing practices (Barrot, 2023; Su et al., 2023). The vocabulary identified in the CAWL-RP may be understood as reflecting current AI-mediated texts and potentially indicating emerging tendencies in AI-assisted academic writing, although longitudinal evidence is required before any claims of enduring lexical change can be made. Rather than being considered essential, this lexical set may be viewed as beneficial for scholars and L2 learners in supporting academic literacy in an AI-influenced writing environment. Familiarity with such vocabulary can assist learners in recognising and interpreting AI-mediated academic texts, thereby supporting both comprehension and production. Within this context, the CAWL-RP functions as a supplementary resource that helps learners notice and engage with vocabulary frequently associated with AI-assisted academic discourse. For L2 academic writers in particular, it may provide useful guidance in navigating hybrid writing environments in which texts are collaboratively shaped by human and AI contributions, thereby supporting more informed and effective engagement with academic discourse.

Characteristics of academic vocabulary in the AWL, NAWL, AVL, and CAWL-RP

The lexical characteristics of the CAWL-RP, when analysed in terms of lexical level and frequency band, reveal distinct patterns that differentiate it from the AWL, NAWL, and AVL. See Table 2.

Table 2. Percentage of vocabulary items in the four lists by lexical levels and frequency bands

Lexical level	Frequency band	AWL (%)		NAWL (%)		AVL (%)		CAWL-RP (%)	
K-1		3.33		6.38		17.86		13.91	
K-2	High	24.33	84.03	12.39	42.80	20.25	65.31	25.04	76.32
K-3		56.37		24.03		27.20		37.37	
K-4		9.41		22.94		10.21		10.58	
K-5		3.29		8.63		5.61		4.67	
K-6	Mid	1.13	14.92	4.87	40.50	5.10	27.93	3.17	21.89
K-7		0.64		2.85		3.76		1.84	
K-8		0.45		1.21		3.25		1.63	
K-9		0.19		0.99		1.89		0.56	
K-10		0.16		0.63		1.57		0.21	
K-11		0.19		0.32		1.28		0.30	
K-12		0.06				0.69		0.04	
K-13		0.06		0.04		0.40		0.09	
K-14				0.15		0.24			
K-15		0.16		0.04		0.09		0.04	
K-16				0.15		0.12		0.04	
K-17	Low		1.05	0.07	16.70	0.03	6.76		1.79
K-18									
K-19						0.01			
K-20				0.04					
K-21						0.01			
K-22									
K-23									
K-24									
K-25									
Off-list		0.23		14.27		0.43		0.51	

The findings indicate that the CAWL-RP is predominantly composed of high-frequency, general academic words, positioning it as a lexically more accessible resource compared to the NAWL and AVL. Table 2 demonstrates significant variations in the difficulty levels across the four lists. While the AWL is highly concentrated at the K-3 lexical level, the NAWL and AVL are more widely distributed, showing a notable presence in lower-frequency bands and a high proportion of 'off-list' words (14.27% for the NAWL). In contrast, the CAWL-RP relies heavily on high-frequency vocabulary, with a high concentration of words in the K-1 to K-3 levels (37.37% at K-3). This is further reinforced by the lexical frequency band analysis, which reveals that the CAWL-RP has the second-highest proportion of high-frequency words (76.32%), substantially surpassing the NAWL (42.80%).

The distribution of words provides clear insights into the intended purpose of each list. The AWL, with 56.37% of its words in the K-3 level, was designed for foundational academic vocabulary learning, assuming learners have already mastered the first 2,000 most common words (Basturkmen, 2006; Coxhead, 2000). Conversely, the NAWL and AVL, created from larger modern corpora, capture a more nuanced range of academic language, including more specialised terms such as *acid*. The CAWL-RP, however, presents a unique pattern; its highest concentrations are actually in the K-1 (13.91%) and K-2 (25.04%) levels. This suggests that a significant portion of the vocabulary refined using ChatGPT for research publication is drawn from common word lists not typically considered academic. This tendency should not be attributed to GPT-3.5 alone. It more plausibly reflects a compound effect of several interacting factors: the lexical tendencies of GPT-3.5 itself when rewriting text; the three-stage transformation pipeline, in which the intermediate rewriting into general English at the generalisation stage may itself favour broadly accessible vocabulary ahead of the subsequent re-academisation stage; the disciplinary composition of the underlying corpus; the frequency-, range-, uniformity-, and dispersion-based criteria used to select words for the CAWL-RP; and the human-authored 2020 source texts, which remain structurally embedded in the corpus even after transformation and continue to shape the lexical material available to the model. Disentangling the relative contribution of each of these factors is not possible within the present corpus-based design and would require a dedicated study that manipulates each factor independently.

This lexical preference has practical implications for academic writers, though these should be read in light of the compound explanation above rather than as a property of ChatGPT in isolation. The findings align with observations by Zhang and Crosthwaite (2025) that while AI-mediated texts use "technically appropriate and formal vocabulary," they may lack the "contextual richness" of human writers. The CAWL-RP's profile suggests that vocabulary produced under this specific corpus design and transformation pipeline tends towards words that are both academically plausible and broadly understood, though whether this reflects GPT-3.5's inherent tendencies, the transformation pipeline, or the interaction between the two cannot be determined from the present data. For L2 academic writers, this points to a potentially useful, though as yet untested, resource: the vocabulary represented in the CAWL-RP may be broadly applicable across disciplines, which could support comprehension and production in an accessible manner, pending empirical investigation of this possibility.

Interestingly, while the CAWL-RP is composed mainly of high-frequency words, it contains a unique subset of 470 types not found in the other three lists. This finding should be interpreted cautiously: it indicates that, within this specific AI-mediated corpus design, GPT-3.5's lexical output under the study's rewriting protocol did not align perfectly with vocabulary attested in the AWL, NAWL, and AVL. It does not indicate that generative AI's conceptualisation of academic vocabulary in general diverges from human-authored norms, since a different model, prompt design, or source-text set could plausibly yield a different set of divergent items. While the concentration in K-1 to K-3 levels implies a conservative approach, the presence of unique words—such as *delve*, *encompassed*, *leverage*, *pertains*, and *surpassing*—demonstrates that certain terms are considered highly functional within AI-mediated research contexts.

Whilst it is premature to definitively claim that generative AI is driving long-term linguistic evolution, the presence of 470 exclusive word types in the CAWL-RP indicates distinct lexical preferences within AI-mediated genres under these specific experimental parameters. Rather than reshaping academic vocabulary at a macro-evolutionary level, these algorithmic tendencies are better theorised as an emerging register shift within human–AI collaborative writing contexts. Therefore, instead of assuming immediate normalisation in broader scholarly communication, these findings suggest a pedagogical need to cultivate learners' AI lexical awareness, enabling them to critically evaluate whether AI-preferred items fit specific disciplinary conventions. Whether these lexical items will become more widespread in future academic communication remains an open question that requires longitudinal investigation. Consequently, these developments warrant further attention in future research examining how learners engage with vocabulary associated with AI-mediated academic texts.

The present study's contribution to CALL is best understood as descriptive corpus evidence about the lexical profile of one form of AI-mediated academic discourse, rather than as a direct pedagogical intervention. This evidence may inform future work in CALL pedagogy, materials design, and AI literacy instruction; however, the pedagogical value of the CAWL-RP itself — for instance, whether learner exposure to these items produces any measurable benefit — has not been tested in this study and remains an empirical question for future classroom-based research. Within this bounded scope, the predominance of high-frequency academic vocabulary in the CAWL-RP suggests one plausible direction for vocabulary instruction: CALL pedagogy could incorporate explicit instruction that encourages learners to treat AI-mediated vocabulary as a starting point for critical evaluation, rather than as a validated target list.

The findings may also be interpreted through the lens of Corpus-Based Language Pedagogy (CBLP). According to Liu and Ma (2025), corpus-derived resources can function as both linguistic and cognitive scaffolds that support learners in navigating complex language systems. In this respect, the CAWL-RP may serve as a scaffold for learners engaging with AI-mediated academic writing by highlighting vocabulary that frequently occurs in texts refined through generative AI. Areas of convergence between the CAWL-RP and established academic word lists may contribute to a reduction in degrees of freedom (Wood et al., 1976), allowing learners to focus on widely accepted academic vocabulary. Conversely, words that diverge from traditional academic word lists may help mark critical features of emerging AI-mediated discourse, encouraging learners to notice differences between conventional academic language and AI-influenced lexical patterns. These applications are proposed as plausible directions consistent with CBLP principles; they have not themselves been tested with learners in the present study.

The CAWL-RP may also support self-directed language learning (SDLL) by providing learners with a structured resource for exploring vocabulary associated with AI-mediated academic writing. From the perspective of the Knowledge-Skills-Attitude (KSA) framework (Nhan et al., 2026), the list could plausibly support the development of knowledge through exposure to academic vocabulary, skills through its application in reading and writing activities, and attitudes through critical engagement with AI-mediated language (Anthony, 2025b), though these are proposed applications rather than outcomes demonstrated by the present study. In particular, the lexical differences identified between the CAWL-RP and traditional academic word lists may encourage learners to evaluate AI-mediated lexical choices critically rather than accept them uncritically. Such evaluation is increasingly important as AI becomes more deeply embedded within academic communication and language learning environments.

Across each of these framings, it is important to emphasise that the pedagogical value of the CAWL-RP is proposed rather than demonstrated: establishing whether engagement with the list produces measurable learning benefits would require controlled, classroom-based investigation, which falls outside the scope of the present corpus-based study.

Limitations and recommendations for future studies

Before detailing specific limitations, it is worth restating the overall epistemological status of the CAWL-RP. The list is derived from a constructed, AI-mediated rewriting corpus rather than naturally occurring AI-assisted academic writing; it is shaped throughout by human-authored 2020 source texts rather than by bottom-up AI composition; and it depends entirely on one model (GPT-3.5) operating under one prompt-and-session protocol. Taken together, these characteristics mean the CAWL-RP should be read as a provisional lexical profile of a single AI-mediated corpus design, not as a broadly generalisable inventory of academic vocabulary in the AI era. The limitations below should be understood within this framing. First, the corpus size, while substantial, is still a small fraction of the data on which large language models like ChatGPT are trained. The CAWL-RP, therefore, may not fully capture the entire range of academic vocabulary that ChatGPT can generate.

Second, the study did not account for the disciplinary-specific usage of words. While it identifies general academic vocabulary, a deeper analysis is needed to determine how ChatGPT's lexical choices vary across fields like medicine, engineering, or social sciences. Third, the absence of expert-based validation constitutes a further limitation. Because the CAWL-RP was compiled entirely through frequency and distributional criteria, no independent pedagogical or domain-expert judgement was sought regarding the conceptual coherence or teaching relevance of individual items. As a result, borderline items whose academic status is context-dependent may have been retained or excluded on statistical grounds alone, and the extent to which the final list corresponds to teachers' and learners' intuitive sense of 'academic vocabulary' has not been independently verified. Fourth, the study focused solely on ChatGPT-mediated texts; it did not compare its vocabulary to that of other AI models like Gemini or Claude, which may have different lexical profiles. Fifth, GPT-3.5 represents an earlier generation of large language models. Subsequent iterations (e.g., GPT-4-class systems) may demonstrate different lexical distributions due to expanded training data and architectural refinements. The CAWL-RP therefore reflects the lexical tendencies of GPT-3.5 under controlled prompting conditions rather than a universal representation of generative AI systems as a whole. A sixth limitation of the present study lies in the structural tethering of the AI-mediated corpus. Because the generative model was fed highly structured academic inputs, the resulting CAWL-RP reflects vocabulary selected under optimal stylistic scaffolding rather than the lexical tendencies of ChatGPT when generating academic prose from scratch.

Based on these limitations, several avenues for future research are recommended. Future studies would benefit from incorporating expert review, particularly for borderline items and pedagogical interpretation, to strengthen the conceptual validity of AI-derived word lists. It should also empirically test the pedagogical value of the CAWL-RP in actual learning settings — for example, through classroom-based or experimental studies examining learners' vocabulary uptake, retention, or critical engagement when exposed to the list — since the present study offers descriptive corpus evidence only and does not itself evaluate pedagogical outcomes. Additionally, exploring the creation of discipline-specific academic word lists by AI would help determine if they exhibit a similar divergence from human-authored lists as the CAWL-RP did from its general academic counterparts. In a related vein, a comparative study of other AI models, such as Gemini or Claude, as well as other GPT versions (e.g., 4, 4o, and 5), is also essential to determine whether their lexical profiles for research publication are similar to or distinct from those of ChatGPT-3.5. Such research would provide valuable insights into similarities and differences in lexical profiles across contemporary AI systems. Future studies should also contrast the CAWL-RP against a corpus generated entirely from bottom-up AI prompts to isolate how the absence of a human structural blueprint alters the model's lexical deployment. A final key area for future study is to move beyond the single-word level (Coxhead, 2024) and examine whether AI models generate collocations and lexical bundles in academic contexts in a similar or different manner to human authors, which could reveal deeper insights into the formulaic nature of AI-mediated academic discourse.

Conclusion

This study has explored the characteristics of academic vocabulary in AI-mediated research writing through the construction of the CAWL-RP. The findings provide initial insights into how lexical patterns may vary when academic texts are processed through generative AI under controlled conditions. In particular, the study identified a set of lexical items that appear to be associated with AI-mediated rewriting; however, these patterns should be interpreted with caution, as they may also reflect corpus design, disciplinary variation, and methodological procedures. The keyword analysis reported in this study demonstrates that the transformation procedure produced measurable lexical restructuring relative to the source corpus, but this finding should not be read as confirming that the CAC-RP is fully independent of its 2020 source texts. Rather than proposing a definitive academic vocabulary list, the CAWL-RP should be viewed as an exploratory resource that captures one possible configuration of academic vocabulary in AI-mediated contexts. Its value lies not in replacing existing word lists derived from human-authored corpora, but in offering a complementary perspective that may inform ongoing discussions about language use in AI-supported academic environments. The study also highlights several methodological and theoretical considerations, including the challenges of distinguishing AI-influenced lexical patterns from corpus artefacts, as well as the need for greater transparency and standardisation in AI-assisted corpus construction. Future research may build on this work by employing alternative corpus designs, incorporating expert validation, and comparing outputs across different AI systems to further examine the stability and generalisability of the observed patterns. In this sense, the present study does not claim to redefine academic vocabulary, but rather contributes to an emerging line of inquiry into how academic language is represented in AI-mediated texts and how such representations compare with those observed in corpora of human-authored academic writing. As such, it provides a foundation for further investigation rather than a conclusive account.

Acknowledgements

This research was supported by the Bachelor of Arts Programme in English for Service Industry (Special Programme), Faculty of Hospitality Industry, Kasetsart University (Project Code: ESI-HI-KUKPS2024-01).

Use of generative AI

This manuscript made use of ChatGPT to assist in two main tasks. First, it was employed to improve the clarity of the text. Second, it was used to edit the language of the collected data in order to enhance its academic quality from the perspective of ChatGPT, which was then utilised to construct the corpus for data analysis. No AI tools were used for generating research hypotheses or writing conclusions. The final manuscript was authored and reviewed by the listed authors without additional AI input.

References

Ahmed, F. E. Y., Ali, E. H. F., Usmani, S., Kottaparamban, M., Ibrahim, M. A. E., Hamed, F. M. A., & Abdallah, N. M. M. (2026). Applying AI-enhanced digital games in teaching English vocabulary. *Research Journal in Advanced Humanities, 7*(1), 1–12. https://doi.org/10.58256/5fqfg723

Anthony, L. (2025a). *AntWordProfiler* (Version 2.2.1) [Computer Software]. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software/AntWordProfiler

Anthony, L. (2025). Integrating AI technology into corpus-based language learning through ChatAI. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2025.2589747

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57*, 100745. https://doi.org/10.1016/j.asw.2023.100745

Basturkmen, H. (2006). *Ideas and options in English for specific purposes*. Lawrence Erlbaum.

Behforouz, B., & Al Ghaithi, A. (2025). AI as a language learning facilitator: Examining vocabulary and self-regulation in EFL learners. *Studies in Self-Access Learning Journal, 16*(4), 616–634. https://doi.org/10.37237/202507

Brezina, V. & Platt, W. (2024) *#LancsBox X* [Computer Software], Lancaster University, http://lancsbox.lancs.ac.uk

Browne, C., Culligan, B., & Phillips, J. (2013). *New academic word list*. Retrieved from http://www.newgeneralservicelist.org/

Butarbutar, R., Sujarwo, Pure, E. A. G., Sulviana, & Rizal, A. (2026). Unlocking the impact of AI-assisted online interactive learning environment in English language learning. *Journal of Asia TEFL, 22*(4), 646–662. https://doi.org/10.18823/asiatefl.2025.22.4.2.646

Campion, M., & Elley, W. (1971). *An academic vocabulary list.* New Zealand Council for Educational Research.

Casal, J. E., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing? A study of research ethics and academic publishing. *Research Methods in Applied Linguistics, 2*(3), 100068. https://doi.org/10.1016/j.rmal.2023.100068

Charles, M. (2013). English for academic purposes. In B. Paltridge & S. Starfield (Eds.), *The Handbook of English for Specific Purposes* (136–153). John Wiley & Sons.

Charles, M., & Pecorari, D. (2016). *Introducing English for academic purposes.* Routledge.

Cheung, L., & Crosthwaite, P. (2025). CorpusChat: integrating corpus linguistics and generative AI for academic writing development. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2025.2506480

Chung, E., & Wan, A. (2025). Examining the use of academic vocabulary in first-year ESL undergraduates' writing: A corpus-driven study in Hong Kong. *Assessing Writing, 63*, 100913. https://doi.org/10.1016/j.asw.2024.100913
Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language, 15*, 103–116.
Chung, T. M., & Nation, I. S. P. (2004). Identifying technical vocabulary. *System, 32*(2), 251–263. https://doi.org/10.1016/j.system.2003.11.008
Cobb, T. (2022). *Vocabprofile*. [Computer Program]. http://www.lextutor.ca/vp/
Cobb, T., & Horst, M. (2001). Reading academic English: Carrying learners across the lexical threshold. In J. Flowerdew & M. Peacock (Eds.), *Research Perspectives on English for Academic Purposes* (pp. 315–329). Cambridge University Press.
Coxhead, A. (2000) A new academic word list. *TESOL Quarterly, 34*(2), 213–238. https://doi.org/10.2307/3587951
Coxhead, A. (2020). Academic vocabulary. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 97–110). Routledge.
Coxhead, A. (2024). Current issues and future research on teaching and learning of academic vocabulary in EMI contexts. *International Journal of TESOL Studies, 6*(2), 105–108. https://doi.org/10.58304/ijts.20240207
Coxhead, A., & Hirsh, D. (2007). A pilot science-specific word list. *Revue Française de Linguistique Appliqueé, 12*(2), 65–78.
Crosthwaite, P., & Baisa, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast! *Applied Corpus Linguistics, 3*(3), 100066. https://doi.org/10.1016/j. acorp.2023.100066
Csomay, E., & Prades, A. (2018). Academic vocabulary in ESL student papers: A corpus-based study. *Journal of English for Academic Purposes, 33*, 100–118. https://doi.org/10.1016/j.jeap.2018.02.003
Dang, T. N. Y. (2020). Corpus-based word lists in second language vocabulary research, learning, and teaching. In S. Webb (ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 288–304). Routledge.
Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning, 67*(4), 959–997. https://doi.org/10.1111/lang.12253
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
Flowerdew, J. (2013). English for research publication purposes. In B. Paltridge & S. Starfield (Eds.), *The Handbook of English for Specific Purposes* (pp. 301–321). John Wiley & Sons.
Fuertes-Olivera, P. A. (2024). Making lexicography sustainable: Using ChatGPT and reusing data for lexicographic purposes. *Lexikos, 34*, 123–140. https://doi.org/10.5788/34-1-1883
Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics, 28*(2), 241–265. https://doi.org/10.1093/applin/amm010
Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics, 35*(3), 305–327. https://doi.org/10.1093/applin/amt015

Ghadessy, M. (1979). Frequency counts, word lists, and material preparation: A new approach. *English Teaching Forum, 77*(1), 24–27.

Gong, T., Liu, L., Shi, J., & Guo, Y. (2025). A meaning-based academic vocabulary list. *Journal of English for Academic Purposes, 77*, 101557. https://doi.org/10.1016/j.jeap.2025.101557

Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics, 13*(4), 403–437. https://doi.org/10.1075/ijcl.13. 4.02gri

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

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies, 7*(2), 173–192. https://doi.org/10.1177/1461445605050365

Hyland, K. (2012). *Disciplinary identities: Individuality and community in academic discourse*. Cambridge University Press.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly, 41*(2), 235–253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Jiang, F., & Hyland, K. (2024). Does ChatGPT argue like students? Bundles in argumentative essays. *Applied Linguistics, 46*(3), 375–391. https://doi.org/10.1093/applin/amae052

Juilland, A. G., & Chang-Rodríguez, E. (1964). *Frequency dictionary of Spanish words*. Mouton.

Khani, R., & Tazik, K. (2013). Towards the development of an academic word list for applied linguistics research articles. *RELC Journal, 44*(2), 209–232. https://doi.org/10.1177/0033688213488432

Kong, X., & Liu, C. (2024). A comparative genre analysis of AI-generated and scholar-written abstracts for English review articles in international journals. *Journal of English for Academic Purposes, 71*, 101432. https://doi.org/10.1016/j.jeap.2024.101432

Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: Challenging conventions. *TESOL Quarterly, 50*(4), 976–987. https://doi.org/10.1002/tesq.329

Laosrirattanachai, P. & Laosrirattanachai, P. (2024). Hope you guys enjoyed this vlog: Lexical bundles and moves within English travel vlog discourse. *LEARN Journal: Language Education and Acquisition Research Network, 17*(2), 249–278. https://doi.org/10.70730/VIZO4161

Laosrirattanachai, P., & Laosrirattanachai, P. (2025). Tracing tourism business research trends in Scopus-indexed journals using corpus-based and judgement-based approaches. *Humanities, Arts and Social Sciences Studies, 25*(1), 32–53. https://doi.org/10.69598/hasss.25.1.268122

Laosrirattanachai, P., Rittikote, N., & Laosrirattanachai, P. (2026). Developing a nature tourism word list for L2 learners in the tourism sector. *Australian Review of Applied Linguistics, Advance online publication*. https://doi.org/10.1075/aral.25109.lao

Laosrirattanachai, P., & Ruangjaroon, S. (2021). Corpus-based creation of tourism, hotel, and airline business word lists. *LEARN Journal: Language Education and Acquisition Research Network, 14*(1), 50–86.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307–332. https://doi.org/10.1093/applin/16.3.307

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

Li, Q., & Tarp, S. (2024). Using generative AI to provide high-quality lexicographic assistance to Chinese learners of English. *Lexikos, 34*, 397–418. https://doi.org/10.5788/34-1-1944

Liu, D., & Lei, L. (2020). Technical vocabulary. In S. Webb (ed.), *The Routledge Handbook of Vocabulary Studies* (pp.111–124). Routledge.

Liu, J., & Ma, Q. (2025). Supporting low-proficiency L2 learners' vocabulary learning with custom GPT-scaffolded corpus-based language pedagogy: a case study. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2025.2539152

Lynn, R. W. (1973). Preparing word lists: A suggested method. *RELC Journal, 4*(1), 25–32.

Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes, 31*, 44–57. https://doi.org/10.1016/j.jeap.2017.12.006

Meebangsai, D., Pongtin, P., Kitipoontanakorn, P., & Laosrirattanachai, P. (2023). Investigating proficiency of academic English in student writing: A comparative case study on vocabulary utilization in student research article writing vis–à–vis national and international research. *PASAA, 67*, 66–100.

Mo, Z., & Crosthwaite, P. (2025). Exploring the affordances of generative AI large language models for stance and engagement in academic writing. *Journal of English for Academic Purposes, 75*, 101499. https://doi.org/10.1016/j.jeap.2025.101499

Namaziandost, E., & Çakmak, F. (2025). Impact of AI-generated storytelling vs. gamified learning on vocabulary retention and engagement in CALL environments. *Computers and Education: Artificial Intelligence, 9*, 100505. https://doi.org/10.1016/j.caeai.2025.100505

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins.

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.

Nhan, T. T., Anh, T. T. N., Chi, N. Q., Giang, T. H., Trang, C. T. (2026). YouTube's impact on students' self-directed language learning (SDLL): A comprehensive evaluation based on the knowledge-skills-attitude trio. *The JALT CALL Journal, 22*(1), 102669. https://doi.org/10.29140/jaltcall.2026.102669

Oh, P. S., & Lee, G. G. (2024). Confronting imminent challenges in humane epistemic agency in science education: An interview with ChatGPT. *Science & Education, 34,* 779–805. https://doi.org/10.1007/s11191-024-00515-1

Oktavianti, I. N., Budiwati, T. R., Prayudha, Prayogi, I., & Hidayah, I. N. (2026). Students' attitudes, challenges, and needs in English reading comprehension: Foundations for AI-integrated corpus-based language pedagogy. *Discover Education, 5*(1), 34. https://doi.org/10.1007/s44217-026-01115-7

Paquot, M. (2010). *Academic vocabulary in learner writing.* Continuum.

Pecorari, D., Shaw, P., & Malmstrom, H. (2019). Developing a new academic vocabulary test. *Journal of English for Academic Purposes, 39*, 59–71. https://doi.org/10.1016/j.jeap.2019.02.004

Praninskas, J. (1972). *American university word list.* Longman.

Praphan, P. W., & Praphan, K. (2023). AI technologies in the ESL/EFL writing classroom: The villain or the champion? *Journal of Second Language Writing,* 62, 101072. https://doi.org/10.1016/j.jslw.2023.101072

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

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484–503. https://doi.org/10.1017/S0261444812000018

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly, 36*(2), 145–171. https://doi.org/10.2307/3588328

Scott, M., & Tribble, C. (2006). *Textual patterns: Keywords and corpus analysis in language education*. John Benjamins.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Routledge.

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing, 57*, 100752. https://doi.org/10.1016/j.asw.2023.100752

Thakkar, A. (2024). Chat GPT: A new academic reality for research and education. *Progress in Education, 82*, 73–88.

Uchida, S. (2025). Generative AI and CEFR levels: Evaluating the accuracy of text generation with ChatGPT-4o through textual features. *Vocabulary Learning and Instruction, 14*(1), 2078. https://doi.org/10.29140/vli.v14n1.2078

Valipouria, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes, 12*(4), 248–263. https://doi.org/10.1016/j.jeap.2013.07.001

Vilkaitė-Lozdienė, L., & Schmitt, N. (2020). Frequency as a guide for vocabulary usefulness: High-, mid-, and low-frequency words. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 81–96). Routledge.

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing, 62*, 101071. https://doi.org/10.1016/j.jslw.2023.101071

Webb, S. (2008). Receptive and productive vocabulary size. *Studies in Second Language Acquisition, 30*(1), 79–95. https://doi.org/10.1017/S0272263108080042

Wen, Y., Chiu, M., Guo, X., & Wang, Z. (2025). AI-powered vocabulary learning for lower primary school students. *British Journal of Educational Technology, 56*(2), 734–754. https://doi.org/10.1111/bjet.13537

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

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976. tb00381.x

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication, 3*, 215–229.

Yu, M. (2026). Optimizing EFL vocabulary acquisition: a randomized controlled mixed-methods investigation of artificial intelligence-driven incidental, contextual, and multimodal strategies. *Education and Information Technologies, 31*(1), 53–97. https://doi.org/10.1007/s10639-025-13803-2

Zhang, M., & Crosthwaite, P. (2025). More human than human? Differences in lexis and collocation within academic essays produced by ChatGPT-3.5 and human L2 writers. *International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2024-0196

Zhang, M., & Zhang, J. (2025). Disciplinary variation of metadiscourse: A comparison of human-written and ChatGPT-generated English research article abstracts. *Journal of English for Academic Purposes, 76,* 101540. https://doi.org/10.1016/j.jeap.2025.101540

Appendix

The ChatGPT-Mediated Academic Word List for Research Publication (CAWL-RP) comprises 869 word families with their of 2,341 types.

Note:

1. The headword of each family may either be the actual headword or the most frequently used representative type within that family, or alternatively, the form that is closest to the headword prior to derivation.
2. Words shown in italics are those that appear exclusively in the CAWL-RP and are absent from other established academic word lists.

Headword	Type member
ability	
absence	absent
absolute	
absorb	*absorbance*, absorption
abundance	abundant
accelerate	accelerated, accelerating, acceleration
acceptable	acceptance
access	accessibility, accessible
accommodate	
accompany	accompanied
accordance	*according*, accordingly
account	*accounted*, accounting, *accounts*
accumulate	
accurate	accuracy, accurately
achieve	achieved, achieves, achieving
acknowledge	acknowledged
acquire	acquired, acquisition
act	*action*, actions, active, actively, activity, *activities*
activate	activated, activation
actual	
adapt	adaptability, adaptation, adapted, adaptive
addition	additional
address	addressed, *addresses*, *addressing*
adjust	adjusted, adjusting, adjustment, adjustments
administer	*administered*, administration
adopt	adopted, adopting, adoption

Continue

Continued

Headword	Type member
adult	adults
advance	*advances*, advancing, advanced, advancement, *advancements*
advantage	advantageous
adverse	adversely
affect	affected, affecting, affects
agency	
aggregate	aggregated, aggregates, aggregation
aim	*aimed, aiming, aims*
align	aligned, *aligning*, alignment, *aligns*
alleviate	
allocate	allocated, allocation
allow	allowed, *allowing*, *allows*
alter	alterations, altered, alternative, alternatively, alternatives
amplify	amplification, amplified
analogous	
analyze	analyse, analyses, analyzes, analysis, analytical, *analytics*, analyzed, analyzing, analysed, analysing
annotate	annotated, annotation
annual	
anomalous	*anomalies*
anticipate	anticipated
apparent	
appear	*appears*
apply	applicability, application, *applications*, applied, applicable, *applying*
approach	approaches
appropriate	
approximate	approximately, approximation
area	areas
arise	*arises*, *arising*
array	arrays
article	*articles*
articulate	articulated
ascertain	
aspect	aspects
assay	*assays*
assemble	

Continue

Headword	Type member
assess	assessed, assessing, assessment, assessments
assign	assigned, assignment
assist	assisted
associate	associated, association, *associations*
assume	assumed, assuming, assumption, assumptions
attain	attained
attenuate	attenuated, attenuation
attribute	attributable, attributed, attributes
augment	augmentation, augmented, *augmenting*
author	authorities, authors
automate	automated, autonomous
available	availability
average	*averaged*, *averaging*
aware	awareness
axis	*axial*
background	
balance	balanced
barrier	*barriers*
base	*based*, basic, basis, *basal*, *baseline*
batch	
benchmark	
benefit	beneficial, benefits
bias	biases
bidirectional	
binary	
body	*bodies*
bond	bonding, bonds
boundary	*boundaries*
bridge	
brief	
broadly	
buffer	
bulk	
burden	
calculation	calculated, calculations
calibrate	calibrated, calibration
candidate	candidates

Continue

Continued

Headword	Type member
canonical	
capable	capabilities, capability
capacity	capacities
capture	*captured*, *captures*, *capturing*
carrier	carriers
categorize	categories, categorise, categorised, categorized, category
causal	
center	*centre*, centered, *centred*, *centers*, *centres*, central
challenge	challenges, challenging
change	*changes*
channel	channels
character	characteristic, *characteristics*, characterize, *characterise*, characterization, characterisation, *characterized*, *characterised*
charge	*charged*
chronic	
circuit	circulation
cite	citation
class	*classes*, classification, classify, classified
clear	clarity, *clearance*
close	*closely*
cluster	clustering, *clusters*
coefficient	coefficients
cognitive	
coherent	coherence
cohort	*cohorts*
collaborate	collaboration, collaborative
collect	collected, collection, collective, collectively
combine	combination, *combinations*, combined, combining
commit	commitment
common	commonly
communicate	communication
community	communities
compact	
compare	comparable, comparative, *compared*, *comparing*, comparison, *comparisons*
compatible	compatibility
compel	compelling

Continue

Headword	Type member
compete	*competencies*, competitive, competition
complementary	
complete	completed
complex	*complexes*, complexities, complexity
comply	compliance
component	components
compose	composed, *composites*, *compositions*
compound	compounds
comprehend	comprehension, comprehensive, comprehensively
compress	*compression*, *compressive*
comprise	comprised, comprises, comprising
compromise	*compromised*, *compromising*
compute	computation, computational, computed, computer, computing
concentrate	concentrated, concentration, *concentrations*
concept	concepts, conceptual, conceptualized, conceptualised
concern	concerning, *concerns*
conclude	conclusion, conclusions
concomitant	
concordance	
concurrent	concurrently
condition	*conditions*
conducive	conduct, conducted, conducting, conduction, conductive, conductivity
configure	configuration, configurations
confine	confined
confirm	confirmed, confirming
conflict	
conjugate	*conjugated*, conjunction
connect	connectivity, connection, *connections*
conserve	conservation, conserved
consider	considered, considering, consideration, *considerations*, considerable, considerably
consist	consistency, consistent, consistently, consisting, consists
constant	constituent, constituents, constitute, constitutes, constituting
constrain	constrained, constraint, constraints
construct	constructed, constructing, construction, constructs
consume	consumers, consumption
contact	contamination

Continue

Continued

Headword	Type member
contain	contained, containing, *contains*
content	*contents*
contemporary	
contend	
context	contexts, contextual
contingent	
continue	continued, continuous, continuously
contrast	
contribute	contributed, contributes, contributing, contribution, contributions
control	controlling, *controls*, controlled
conventional	
converge	convergence
converse	conversely
convert	conversion, converted
convolution	*convolutional*
coordinate	coordinates, coordination
core	
correlate	correlated, correlates, correlation, correlations
correspond	corresponding, correspondingly, corresponds
corroborate	*corroborated*, corroborating
counterpart	*counterparts*
create	creation
crisis	
criterion	criteria
critical	
crucial	
cue	cues
culminate	*culminating*
cumulative	
current	currents
cycle	cycles, cyclic, cycling
data	*database*, databases, *dataset*, *datasets*
decision	
decline	declined, declines
decompose	*decomposition*

Continue

Headword	Type member
dedicate	*dedicated*
deem	*deemed*
defect	defects
defend	defense, defence
define	defined, defining, definition, definitive
deform	*deformation*
degrade	degradation
degree	*degrees*
delineate	delineated, *delineates*, *delineating*
delve	*delves*
demand	*demands*
demographic	
demonstrate	demonstrated, demonstrates, demonstrating
denote	denoted, denotes
dense	*densities*, *density*
depend	dependence, *dependencies*, dependent
depict	depicted, depicting, depicts
deplete	depletion
deploy	*deployment*
deposit	*deposited*, *deposition*
derive	derived
describe	described, description
design	designated, designed, designs
desire	desired
despite	
detail	detailed
detect	detected, detecting, detection, detector
determine	determined, determining
detrimental	
develop	developmental, developed, developing, development
deviate	deviation, deviations
device	devices
diagnose	diagnosis, diagnostic
diagram	
differ	differential, differentially, differentiation, difference, *differences*, *different*, differing
digital	

Continue

Continued

Headword	Type member
dimension	dimensional, *dimensionality*, dimensions
diminish	diminished, diminishes, diminishing
direct	directed, direction, *directions*, directly
discernible	
discharge	
discipline	*disciplines*
disclose	*disclosure*
discover	discovery
discrepancy	*discrepancies*
discrete	
discuss	discussed, discussion, *discussions*
dislocate	dislocation
disorder	*disorders*
disparate	*disparities*, disparity
disperse	dispersed, dispersion
displace	displacement
display	displayed, displays
disrupt	disruption, *disruptions*
disseminate	
distance	*distances*
distinct	distinction, distinctions, distinctive
distinguish	distinguished
distribute	distributed, distribution, distributions
diverge	divergence, divergent
diverse	diversity
divide	*divided*
document	documented, documents
domain	domains
domestic	
dominant	
downregulate	*downregulated*
draw	*drawing*, *drawn*
dual	
due	
durable	*durability*
duration	durations

Continue

Headword	Type member
dynamic	dynamically, dynamics, *dysfunction*
early	
effect	*efficacious*, *efficacy*, *effects*, effective, effectively
efficient	*efficiencies*, efficiently
effort	*efforts*
elaborate	elaborated
element	*elemental*, elements
elevate	elevated, elevation
elicit	*elicited*
elucidate	*elucidated*, elucidates, *elucidating*
embed	embedded, embedding, embeddings
emerge	emerged, emergence, emergent, emerges, emerging
emphasize	emphasis, emphasized, emphasizes, emphasizing, emphasise, emphasised, *emphasises*, emphasising
empirical	empirically
employ	employed, *employing*, *employs*
enable	enabled, enables, enabling
encapsulate	encapsulated
encode	encoded, encoder, encoding
encompass	*encompassed*, *encompasses*, *encompassing*
encounter	encountered
endeavor	*endeavors*, endeavour, *endeavours*
engage	engaged, engagement, engaging
engineer	*engineered*
enhance	enhanced, enhancement, *enhancements*, enhances, enhancing
enrich	enriched, enrichment
ensure	ensures, ensuring
entail	*entails*
entity	entities
environment	environmental, environmentally, environments
equal	
equation	equations
equilibrate	equilibrium
equip	equipment, equipped
equity	
equivalent	
error	errors

Continue

Continued

Headword	Type member
escalate	*escalating*
essential	
establish	established, establishing, establishment
estimate	estimated, estimates, estimating, estimation
ethical	ethics
evaluate	evaluated, evaluating, evaluation, evaluations
evident	evidence, evidenced
evolve	evolution, evolving
examine	examined, *examines*, *examining*
example	*examples*
exceed	exceeded, exceeding, exceeds
except	exception
excess	excessive
exchange	
exclude	excluded, excluding, exclusion, exclusively
execute	executed, execution
exemplify	*exemplified*
exert	*exerted*, *exerts*
exhibit	exhibited, exhibiting, exhibits
exist	existence, existing
expand	expanded, expanding, expansion
expect	expectation, *expectations*, expected
experience	*experiences*, experienced
experiment	experimentally
expert	expertise, experts
explain	explanation, explanatory
explicit	explicitly
explore	exploration, *explored*, *exploring*
expose	exposed, exposure
express	expressed, *expressing*, expression, *expressions*
extant	
extend	*extending*, *extends*, extended, extension, extensive, extensively, extent
external	
extract	extracted, extraction, extracts
extrude	extrusion
fabricate	fabricated, fabrication

Continue

Headword	Type member
facilitate	facilitated, facilitates, facilitating, facilities, facility
factor	factors
failure	
favorable	favourable
feasible	feasibility
feature	features, featuring
feed	feedback
fidelity	
field	*fields*
figure	*figures*
filter	*filtered*, *filtering*
find	finding, *findings*
fix	fixed
flexible	flexibility
flow	*flows*
fluctuate	fluctuations
focal	
focus	focused, focuses, focusing
follow	*followed*, *follows*, following
force	*forces*
forecast	*forecasting*
form	formed, forming, *forms*, formal, formation, format
formula	formulated, formulation, formulations
foster	fostering, *fosters*
found	foundation, *foundational*
fraction	fractions
fracture	
fragment	*fragments*
frame	*frames*
framework	frameworks
frequent	*frequencies*, frequently, frequency
function	functional, *functionalities*, functionality, functioning, functions
fundamentally	fundamental
fuse	fusion
fuzzy	
gain	*gained*, *gains*
gap	*gaps*

Continue

Continued

Headword	Type member
garner	*garnered*
gender	
generalize	*generalizability*, generalization, generalized, generalise, *generalisability*, generalisation, generalised
generate	generated, generates, generating, generation, generative
geographic	geographical
global	globally
goal	goals
govern	governing
gradual	gradually
grain	*grained*
graph	graphical, graphs
ground	grounded
group	*groups*
grow	growing, growth
guide	guidance, guidelines, guided
hence	
herein	
heterogeneous	heterogeneity
hierarchical	
highlight	highlighted, highlighting, highlights
holistic	
horizontal	
hybrid	
hypothesis	hypotheses, hypothesized
ideal	
identical	
identify	identification, identified, identifies, identifying, identity
illuminate	illumination
illustrate	illustrated, illustrates, illustrating, illustration
immediate	*immediately*
impact	impacts
impair	*impaired*
impede	*impedance*
imperative	
implement	implementation, implemented, implementing

Continue

Headword	Type member
implicate	implicated, implications
imply	implies, implicit, implying
important	importance
impose	imposed
improve	*improves*, improved, improvement, *improvements*, improving
inadequate	
incentive	incentives
incidence	incident
include	included, *includes*, including, inclusion
incorporate	incorporated, incorporates, incorporating, incorporation
increase	increased, *increases*, increasing, increasingly
incremental	
indeed	
independent	independently
index	*indices*
indicate	indicated, indicates, indicating, indicative, indicator, indicators
indirect	
indispensable	
individual	individually, individuals
induce	induced, induces, inducing, induction
infer	inference, inferior, inferred
infiltrate	*infiltration*
influence	*influenced*, *influences*, influencing, influential
inform	information, informed
inherent	inherently
inhibit	inhibiting, inhibition, *inhibitor*, *inhibitors*, *inhibitory*
initial	initially, initiated, initiation, initiatives
innovate	innovations, innovative
input	inputs
inquiry	
insight	insights
inspect	inspection
inspire	*inspired*
instability	
instance	instances
institution	institutional, institutions
instrument	*instruments*, instrumental

Continue

Continued

Headword	Type member
insufficient	
intact	
integrate	integral, integrated, integrates, integrating, integration, integrative
integrity	
intelligence	intelligent
intend	intended, *intentions*
intensify	intensified, intensities, intensity, intensive
interact	interaction, interactions, interactive
interconnect	interconnected
interest	
interference	
intermediate	
internal	
interval	intervals
intervene	intervention, interventions
interview	*interviews*
intricate	
intrinsic	intrinsically
introduce	introduced, *introduces*, *introducing*, introduction
inverse	
investigate	investigated, investigates, investigating, investigation, investigations
involve	involved, involvement, involves, involving
isolate	isolated, isolation
issue	issues
item	items
iterate	iteration, *iterations*, iterative
kernel	
key	
knowledge	
label	labelled, labelling, labeled, labeling, labels
laboratory	
lack	lacking
landscape	
latent	
latter	

Continue

Headword	Type member
layer	layered, layers
level	*levels*
leverage	*leveraging*
lie	*lies*
likely	
likewise	
limit	*limits*, limitation, *limitations*, limited, limiting
link	linked, linking, links
literature	
local	
locate	localization, localized, located, location, locations, *localisation*, localised
longitudinal	
loop	
machine	*machining*
magnitude	magnitudes
maintain	maintained, maintaining, maintenance
major	majority
manage	managing
manifest	
manipulate	manipulation
manner	
manual	
map	*maps*
marginal	marginally
marked	markedly
material	*materials*
max	maximize, maximal, maximise, maximum
mean	*means*, meaningful
measure	measured, *measures*, measuring
mechanism	mechanisms, mechanistic
mediate	media, *median*, mediated, mediating, medium
meet	*met*
member	*members*
mental	
merely	
method	methodological, methodologies, methodology, methods

Continue

Continued

Headword	Type member
metric	metrics
milieu	
min	minimize, minimal, minimise, minimising, minimizing, minimum
minor	
mitigate	*mitigated*, *mitigates*, mitigating, mitigation
mix	mixed
mobile	mobility
mode	modes
model	modeled, modeling, *models*, *modelled*, modelling
moderate	
modify	modification, modifications, modified
modulate	modulated, *modulates*, *modulating*, modulation
module	*modules*, *modulus*
monitor	monitored, monitoring
moreover	
motivate	
movement	*movements*
multifaceted	
multimodal	
multiple	
mutual	
namely	
narrow	
nature	
navigate	navigation
necessary	necessity, necessitate, *necessitates*, *necessitating*
need	*needs*
negative	negatively, negligible
network	networks
neutral	
nevertheless	
node	*nodes*
nominal	
nonetheless	
normal	normalization, normalized, normative, norms, normalisation, normalised
note	notable, notably, noted

Continue

Headword	Type member
noteworthy	notion
null	
number	*numbers*, numerous
objective	*objectives*
observe	observation, *observations*, observed
obstacles	
obtain	obtained
occur	occurred, occurrence, occurring, occurs
offer	*offering*, offers
ongoing	
operation	operations
optimal	*optimisation*, *optimization*, *optimise*, *optimize*, *optimised*, *optimized*, *optimising*, *optimizing*
option	options
organize	organization, *organizations*, *organisation*, *organisations*, *organise*, organized, organised
orient	orientation, orientations, oriented
originate	originating
outcome	outcomes
outline	outlined
output	outputs
overall	
overarching	
overcome	
overlap	overlapping
overview	
paper	
paradigm	paradigms
parallel	
paramount	
part	partial, partially
participate	participants, participation
particular	particularly
partner	partners
pattern	*patterns*
peak	*peaks*
penetrate	penetration
perceive	perceived, percentage, perception, perceptions

Continue

Continued

Headword	Type member
performance	
period	periodic, periods
persist	persistence, persistent
personalize	personalized, *personnel*, personalise, *personalised*
perspective	perspectives
pertain	pertaining, *pertains*, pertinent
phase	phases
phenomenon	phenomena
pivotal	
place	*placement*
planning	
platform	*platforms*
plausible	
plot	*plots*
policy	policies
policymakers	
population	*populations*
portion	
pose	posed, poses, posited, *position*, positioned, positioning, positions, *posits*, *posterior*
positive	positively
possess	*possesses*, possessing
potent	potential, potentially, *potentials*
practice	practical, practise, *practices*, *practises*
precede	preceding
precise	precisely, precision
predefine	predefined
predict	predicted, predicting, prediction, predictions, predictive, predictor
predominant	predominantly
preference	*preferences*
preliminary	
prepare	preparation
presence	*present*, presented, *presenting*, presents, presentation
preserve	preservation, preserved, preserving
prevail	prevailing, prevalence, prevalent
previous	previously

Continue

Headword	Type member
primary	primarily
principal	principle, principles
prior	prioritise, prioritize
private	
probe	
procedure	procedures
process	processed, processes, processing
produce	produced, producing, product, *products*, production, productivity
proficiency	
profile	*profiles*
profound	
program	programme, programming, *programs*, *programmes*
progress	progression
project	projected, projection, projects
prolong	*prolonged*
prominent	prominence
promote	promotes, promoting
propagate	propagation
propensity	
proportion	proportional, proportions
propose	proposed, *proposes*
prospective	
protect	protective
protocol	protocols
provide	provided, *provides*, providing
proximal	proximity
pseudo	
public	publication, publications
publish	published
purpose	*purposes*
putative	
qualitative	
quantify	quantification, quantified, *quantifies*, *quantifying*, quantitative, quantitatively
question	*questions*
questionnaire	
random	randomised, randomized, randomly

Continue

Continued

Headword	Type member
range	ranged, ranges, ranging
rapid	rapidly
rate	*ratings*
ratio	ratios
react	reaction, reactions, reactive
reason	reasoning
recall	
receive	received, receiving
recognize	recognition, recognized, *recognizing*, *recognise*, recognised, *recognising*
recombine	recombination
recommend	recommendation, *recommendations*
reconstruct	reconstructed, reconstruction
record	recorded, *records*
recover	recovery
recruit	recruitment
recurrent	
recycle	*recycling*
refer	referred, *refers*, reference, *references*
refine	refined, refinement
reflect	reflected, *reflecting*, *reflects*, reflection
regard	*regarded*, regarding
regime	regimes
region	regional, regions
regress	regression
regulate	regulated, regulating, regulation, regulations, regulatory
reinforce	reinforced, reinforcement, reinforcing
relate	related, relation, relationship, *relationships*, relative, relatively
release	released
relevant	relevance
reliable	reliability
rely	reliance, relying
remain	*remained*, *remaining*, *remains*
remodel	*remodeling*, *remodelling*
remote	
remove	removal, removed

Continue

Headword	Type member
render	rendering
repeat	repeated
replicate	replicates, replication
represent	*represented*, representing, *represents*, representation, *representations*, representative
require	required, requirement, requirements, requires, requiring, requisite
research	researchers
resilient	resilience
resistant	resistance
resolve	resolution, resolved
resource	resources
respect	respective, respectively
respond	respondents, response, responses, responsive, responsiveness
restrict	restricted
result	*resulted*, *results*, resulting, resultant
retain	retained, retention
retrieve	retrieval, retrieved
reveal	revealed, revealing, reveals
reverse	
review	reviewed, *reviews*
rigid	
rigorous	rigorously
rise	*rising*
robust	robustness
role	roles
rotate	rotation
salient	
sample	*sampled*, *samples*, *sampling*
satisfy	*satisfaction*
scaffold	*scaffolds*
scale	scaling, *scales*
scan	*scanning*
scenario	scenarios
schedule	scheduling
scheme	schemes, schematic
scholars	scholarly
scientific	

Continue

Continued

Headword	Type member
scope	
score	*scores*
secondary	
section	sectional, sections
sector	sectors
segment	*segmentation*, *segments*
select	selected, selecting, selection, selective, selectively, selectivity
sensitive	
separate	separated, separation
sequence	sequences, sequencing, sequential, sequentially
series	
serve	*served, serves*
session	*sessions*
set	*settings*, *setup*, *sets*, setting
shape	shaped, shaping
share	shared
shift	shifts
signal	signaling, *signals*, *signalling*
significant	significance, significantly, signifies, signifying
similar	similarity, similarly
simulate	simulated, simulation, simulations, simultaneous, simultaneously
situate	situated
size	*sizes*
societal	
sole	solely
sophisticated	
source	sourced, sources
space	spatial, spatially, *spaces*, *spacing*
span	spanning
sparse	
specialized	specialised
specific	specifically, specifications, specificity, specified, specimen, specimens
stable	stability, stabilisation, stabilization
stage	
standard	*standards*, standardize, standardise, standardised, standardized

Continue

Headword	Type member
state	*states*
static	
statistical	statistically, statistics
status	
stem	
stimuli	stimulate, stimulation
stochastic	
store	*storage*
straightforward	
strategy	strategic, strategically, strategies
stratify	stratified
strength	strengths, strengthen, strengthening
strong	*stronger*, strongly
stress	stresses
stringent	
structure	structural, structured, structures
study	*studies*
subject	*subjected*, *subjects*, subjective
suboptimal	
subsequent	subsequently
subset	subsets
substantial	
substantiate	substantiated, substitution
success	successful, successfully
sufficient	sufficiently
suggest	suggested, *suggesting*, *suggests*, suitable
sum	summarised, summarized, summary
supplement	supplementary, supplemented
support	supported, *supports*, supporting
suppress	suppressed, suppression
surpass	*surpasses*, *surpassing*
survey	surveys
sustain	sustainability, sustainable, sustained, sustaining
switch	switching
synthesize	synthesis, synthesise, synthesised, synthesized, synthetic
system	*systemic*, systematic, systematically
table	*tables*

Continue

Continued

Headword	Type member
target	targeted, targeting, targets
task	tasks
temporal	
tend	*tends*
tense	tension
term	terms
test	tested, testing, *tests*
theme	thematic, themes
theory	theoretical, theories
thereby	
thorough	thoroughly
threshold	thresholds
tool	*tools*
train	trained
trait	traits
transfer	transferred
transform	transformation, transformations, *transformative*, transformed, transformer
transient	
transit	transition, transitions
treatment	*treatments*
trend	trends
trigger	triggered
typically	
ultimately	
uncertainty	*uncertainties*
undergo	undergoes, underwent
underlie	underlying
underpinning	
underscore	*underscored*, *underscores*, *underscoring*
unify	unified, uniform, uniformly
unique	
unit	
update	updated
upregulate	*upregulated*, *upregulation*
use	used

Continue

Headword	Type member
utilize	utility, utilization, utilized, utilizes, utilizing, utilisation, utilise, utilised, utilises, utilising
valid	validate, validated, validation, validity
value	*values*
vary	variability, variable, variables, variance, variant, variants, variation, variety, various
version	
versus	
vertical	
via	
view	
virtual	
vision	visible, visual, visualization, visualisation, visualised, visualized
vital	
volume	volumes
warrant	warranted
whereas	
whereby	
wherein	
widespread	
workflow	
yield	*yielded*, yielding, *yields*