




Academic Vocabulary in Applied Linguistics Research Articles: A Corpus-based Study

Ismail Xodabande¹ , Shima Torabzadeh², Mohammad Ghafouri³ , Azadeh Emadi⁴ 

1 Kharazmi University

2 Islamic Azad University

3 Guilan University

4 University of Tehran

Background: Generally operationalized as the words used more frequently in academic discourse for describing abstract ideas and processes, academic vocabulary poses a major learning burden for native and non-native speakers of English. Recent developments in corpus-based technologies and tools have made it possible to analyze large bodies of texts for profiling vocabulary items, and a growing number of studies investigated such vocabulary in research articles published in different disciplines.

Purpose: Despite significant progress in academic word list development, research focusing on the contribution of the newly developed word lists in academic texts remained largely limited. Accordingly, the majority of studies used outdated lists for general and academic vocabulary as the starting points in their studies.

Methods: The current study investigated a large corpus of applied linguistics research articles (2000 RAs, 15.5 million words, 20 journals) to identify frequently used academic words based on New Academic Word List (NAWL). In analyzing the data, predefined criteria were used and the study used flemma for counting and defining words.

Findings: The findings indicated that 310 out of 960 academic words in NAWL were used frequently in the corpus and provided 4.19% coverage. This coverage differs considerably with the previous studies that investigated similar corpora using the Academic Word List (AWL) and reported around and more than 10% coverage for academic vocabulary. Since the base lists used for profiling the corpus in this study were different from those employed by the previous studies, such differences mainly arise as a result of improvements in operationalizing general service and academic vocabulary.

Implications: In light of these findings and recent calls for more replication research in vocabulary studies, the study draws some implications for researching and teaching academic vocabulary. Additionally, in order to facilitate academic vocabulary learning in applied linguistics, the study presents a list of frequently used NAWL items divided into six bands based on their frequency in the corpus.

Keywords: vocabulary, academic vocabulary, research articles, applied linguistics, corpus-based word lists, NAWL

Citation: Xodabande, I., Torabzadeh, S., Ghafouri, M., & Emadi, A. (2022). Academic Vocabulary in Applied Linguistics Research Articles: A Corpus-based Study. *Journal of Language and Education*, 8(2), 156-166. <https://doi.org/10.17323/jle.2022.13420>

Correspondence: Ismail Xodabande, ismail.kh.tefl@gmail.com

Received: Nov 24, 2021

Accepted: Jun 14, 2022

Published: Jun 30, 2022

Copyright: © 2022 The Authors



INTRODUCTION

Nowadays, with the establishment of the English as the academic lingua franca

(Hyland, 2013), a considerable number of university students and researchers around the world are required to read and publish in English (Flowerdew, 2015;

Li & Flowerdew, 2020). Nevertheless, it has been argued that non-native speakers of English constantly face serious linguistic barriers in research publication (Corcoran, 2017; Li & Flowerdew, 2020). Insufficient vocabulary knowledge is among the crucial factors that adds to non-native English users' inability to successfully participate in discursive practices of their scientific communities (Bazerman et al., 2012; Laufer, 1996), which is deemed essential for their professional identity development (Hyland, 2013). In this regard, focusing on the vocabulary learning needs of university students and researchers in specialized areas remained an important research agenda. In line with emphasizing disciplinary literacy (Airey et al., 2017), the study of vocabulary in established genres such as research articles can inform material development for instructional purposes, and also helps students and teachers in identifying the most important vocabulary related to their disciplines. Moreover, developing subject specific vocabulary lists helps university students in self-directed and autonomous learning of those items, and English for academic purposes (EAP) teachers can make their instruction more aligned with the learning needs of their students with prioritizing such vocabulary (Webb & Nation, 2017).

Over the past years, the study of vocabulary in academic discourse attracted considerable attention, and the field of applied linguistics has long been interested in identifying specialized lexis across different domains of language use. In this regard, in a paradigmatic classification of vocabulary in English, Coxhead and Nation (2001) made distinctions among (1) general service (high frequency), (2) academic, (3) technical, and (4) low frequency words (for different and more recent views on pedagogical description of vocabulary see Beck et al., (2013), Nation (2013), and Schmitt and Schmitt (2014)). General service vocabulary refers to the most commonly used function and content words that encompass the majority of running words used in all types of writings. These pragmatically neutral words (Stubbs, 1986) cover almost 80% of spoken and written texts in general. Given their importance, it has been argued that this vocabulary should be the first step in developing the lexical knowledge of English language learners, and a number of corpus-based word lists have been developed to guide such endeavors (Brezina & Gablasova, 2015; Browne, 2014, 2021; West, 1953). Unlike general service words, the low frequency vocabulary refers to those rarely used terms that occur very infrequently in academic texts, and they are not crucial for comprehension of the discourse (Laufer, 2005). Technical words constitute subject specific terms that are common in a specialized field (e.g. chemistry), and their meaning and usage are considerably different from one subject area to the next. Those working in a particular profession or studying within a specific field are usually well familiar with their domain specific technical terms, and in academic contexts these words are defined in glossaries and field specific dictionaries (Coxhead, 2018). Nonetheless, occurring between general and technical words are the academic vocabulary

that are neither specific to a specialized area of study, nor they are general in being used across various text types. In the literature, these medium frequency words are referred with different labels such as 'sub-technical' or 'semi-technical', and 'academic' vocabulary (Coxhead, 2018; Paquot, 2010), and corpus-based studies revealed a range of 10% to 14% coverage for them in most academic texts (Coxhead, 2000; Gardner & Davies, 2014). Given that academic vocabulary is used for describing abstract ideas and processes in academic discourse (Paquot, 2010), they pose a major challenge for both English as second/foreign language and native English speaking students in academic writing (Coxhead, 2019; Evans & Green, 2007; Evans & Morrison, 2010, 2011; Spencer et al., 2017).

Recognizing the importance of academic vocabulary, a number of general academic word lists have been developed to be incorporated into wide angle EAP programs, and also for setting principled vocabulary learning goals (Coxhead, 2000; Gardner & Davies, 2014; Xue & Nation, 1984). Since its development more than two decades ago, the Academic Word List (AWL) (Coxhead, 2000) that contains 570 word families remained as a predominant source for EAP instruction, materials development, and vocabulary assessment (Coxhead, 2011; Huntley, 2006; McLean & Kramer, 2015; Wells, 2007). However, a number of recent corpus-based studies investigating various academic corpora started to challenge the status of the AWL as the best list of academic vocabulary (Gardner & Davies, 2014; Hyland & Tse, 2007; Masrai & Milton, 2018). In this regard, the AWL has been criticized for using the old and outdated GSL (West, 1953) for representing general service vocabulary in English (Gardner & Davies, 2014), presence of some general rather than academic words in the list (Masrai & Milton, 2018), and the variation in the coverage provided by the list in different disciplines (Chen & Ge, 2007; Liu & Han, 2015; Martínez et al., 2009; Xodabande & Xodabande, 2020). More seriously, the AWL has been also criticized for using level six word families defined as the base word plus its inflected forms and transparent derivations (Bauer & Nation, 1993; Nation, 2016) as the unit of vocabulary analysis, the choice which limits the pedagogical applications of the list (Gardner & Davies, 2014). In light of the new developments in corpus linguistics and associated technologies for analyzing vocabulary in much larger corpora, two general academic word lists namely the New Academic Word List (NAWL), and the Academic Vocabulary List (AVL) (Gardner & Davies, 2014) have been developed. Although these new lists show significant improvements over the AWL, both in terms of text coverage and pedagogical applicability by using lemma (a headword plus its inflected forms) and flemma (a headword plus inflected forms of different parts of speech) for counting words (Brown et al., 2020), research investigating their contributions to academic discourse remained very limited (Coxhead, 2018, 2019; Durrant, 2016). In this regard, the dominant status of the AWL has resulted in giving far less attention to the newly developed core academic word lists. The current study aimed

to fill part of this gap, and set out to investigate the use of the NAWL items in applied linguistics research articles. It should be noted that although the AVL (Gardner & Davies, 2014) is more empirically grounded based on being published in a peer reviewed study, the NAWL also meets the essential requirements of a systematically developed core academic word list, and the availability of base lists, resources, and corpus information in the project's website makes it an easily accessible resource for language teachers and university students. The findings add to our understanding with respect to the use of academic vocabulary in this field, and the results can guide applied linguistics students and researchers in setting sound vocabulary learning goals.

Academic Vocabulary in Research Articles

The study of academic vocabulary in research articles as the preeminent genre in academy (Hyland, 2009) is an expanding and fast growing area of inquiry (Chen & Ge, 2007; Khani & Tazik, 2013; Martínez et al., 2009; Valipouri & Nassaji, 2013; Vongpumivitch et al., 2009; Wang et al., 2008; Yang, 2015). Within this line research, a good number of studies investigated the contribution of the AWL (Coxhead, 2000) in research articles and developed corpus based academic word lists for a number of subject areas (Dang, 2019). Overall, the studies provided evidence for the significant coverage of the AWL in research articles, and the list consistently provided around 10% coverage in most investigated corpora (Coxhead, 2000, 2011; Coxhead & Byrd, 2007). Nonetheless, these studies also highlighted some of the shortcomings and limitations associated with the AWL as a general academic word list, which is intended to serve a wide variety of disciplines. This section lays out a general overview of such studies, and situates the current study within the existing body of knowledge.

Analyzing a multi-genre and multi-disciplinary corpus with around 3.3 million words, Hyland and Tse (2007) examined the use of the AWL in university textbooks, research articles, lectures, laboratory manuals, thesis and dissertations. The findings of the study revealed that the AWL covers around 10.6% of the corpus which was balanced among different disciplines. Nevertheless, further analysis showed that individual academic vocabulary items on the list occurred and behaved differently in terms of range, frequency, collocation, and meaning across the investigated disciplines. This study was among the first studies that systematically investigated academic vocabulary across a number disciplines, and the findings provided strong evidence for the specificity of vocabulary in academic discourse. Moreover, the study was a pioneer in an ongoing attempt to develop discipline-based and more narrow academic vocabulary list to be used in specific disciplines.

Chen and Ge (2007) studied the use of AWL in 50 medical research articles with around 190000 running words. The study found that 292 words in the AWL were frequently used

in medical research articles. Furthermore, the findings revealed that 111 AWL items were used very infrequently in the corpus. The cumulative coverage of the AWL items in the corpus was around 10% percent, and the use of the high-frequency academic vocabulary in the medical research articles were different compared to the original sub lists developed by Coxhead (2000). In another study, using both qualitative and quantitative analysis, Martínez et al. (2009) investigated the use of the AWL in a corpus of agriculture research articles that contained 826416 running words. The findings indicated that the GSL and AWL items accounted for 76.59% of the tokens in the corpus, while the AWL provided around 9.06% coverage. Data analysis also revealed that about 37.50% items in the AWL were not used in agriculture research articles. Although these early studies supported the findings reported by Hyland and Tse (2007) regarding the specificity of academic vocabulary and disciplinary variation, the small sizes of the investigated corpora pose some limitations on the generalizability of the reported findings (Nation, 2016; Sorell, 2013). Moreover, in a study with a focus on research articles in chemistry, Valipouri and Nassaji (2013) examined a corpus with around four million running words for frequency and distribution of the AWL items. Data analysis indicated that 327 AWL word families that accounted for 9.60% of tokens have been used frequently in the corpus. The study also found that 25% of the words in chemistry research articles were beyond general service and academic vocabulary.

Two studies in the literature examined the use of the AWL in applied linguistics research articles. In this regard, Vongpumivitch et al. (2009) investigated a corpus of 200 research articles collected from five journals with 1.5 million running words. The findings of the study showed that the AWL accounted for about 11.17% of the corpus. Furthermore, 475 AWL items (out of 570) have been identified as being frequent in the applied linguistics research articles. Given the cumulative coverage of the GSL/AWL, the study concluded that the academic vocabulary "play a more important role in academic writing than the non-AWL content word forms in the field of applied linguistics" (p. 37). In a similar study, Khani and Tazik (2013) randomly collected 240 research articles (with 1,553,450 running words) from 12 journals published between 2000 and 2009 and developed an academic word list for the applied linguistics field. The findings attained by this study also showed that the AWL accounted for 11.96% of the words in the corpus. General service and the AWL words together provided a total coverage of 88%. In order to create a pedagogical word list, the authors identified 773 words types (defined as orthographic forms) (573 AWL, 200 non-GSL/AWL) that provided 12.48% coverage in the corpus. These two studies concluded that the academic vocabulary plays an important role in the research articles written in the field of applied linguistics. Additionally, Gholaminejad and Anani Sarab (2020) investigated a large corpus of widely used textbooks in applied linguistics with 10.7 million running words. However, unlike previous studies

that used the GSL and AWL for creating their field specific word lists, Gholaminejad and Anani Sarab (2020) employed the New-GSL (Brezina & Gablasova, 2015) as the base list for high frequency vocabulary in English, and established a lemma-based academic word list for applied linguistics. More specifically, the study identified 336 lemmas each occurring with a minimum frequency of 45.7 per million words, and accounting for 7.1% of the words in the corpus. Together with the New-GSL words, these academic lemmas provided 61% coverage in the entire corpus. Furthermore, the study revealed that only 67.85% of the academic lemmas used in the corpus overlapped with AVL items (Gardner & Davies, 2014), with a considerable number of AVL words (i.e. 2679 lemmas) not being frequently used in applied linguistics textbooks.

In sum, the findings reported by earlier studies investigating the AWL in research articles generally indicate that (1) academic vocabulary as defined based on the AWL accounts for a significant proportion of words in research articles across disciplines, (2) despite around 10% coverage of the AWL in different subject areas, the use of such vocabulary also shows significant disciplinary variation (Hyland & Tse, 2007; Martínez et al., 2009), and (3) the AWL provides higher coverage in humanities than hard sciences (Khani & Tazik, 2013). These observations are in agreement with a firmly grounded view on academic literacy that emphasizes the close link between the content knowledge of a given discipline, and the associated uses of specific vocabulary in the discursive practices (Hyland, 2002, 2006, 2013; Woodward-Kron, 2008). In this regard, there is a need to first investigate specialized texts in academic discourse to identify terminological choices in different subject areas, and then make such resources available for those who need them in their professional practices. Such undertakings can result in better outcomes if newly developed words lists with enhanced pedagogical potential incorporated into corpus-based studies of academic texts such as research articles. Following this line of inquiry, the current study aimed to investigate the coverage of the NAWL in the applied linguistics research articles, and to identify highly relevant and pedagogically useful academic words for university students and researchers within the field. In doing so, the following research questions were addressed:

1. What is the coverage of the NAWL in applied linguistics research articles?
2. What are the frequently used academic words in applied linguistics research articles?

METHOD

Corpus

The corpus analyzed in the current study was compiled by systematic selection of 2000 research articles published in

20 well-known journals in the field of applied linguistics. In order to ensure the balance and representativeness of the corpus, principled procedures were followed in selecting journals and research articles. In this regard, first, after searching the SCImago¹ journal ranking data-base, the 50 top ranking journals in the field of applied linguistics were identified. The list then was given to 10 university professors with extensive experience in the field, and they were asked to select 20 journals that best represent the field. After finalizing the list of the journals based on expert recommendations, all published articles between 2011 and 2020 in these journals were collected and classified based on journal name and publication year. In order to create a manageable corpus, stratified random sampling was used in which 10 articles per year were randomly selected for each of the journals. The PDF documents were then converted into text files, and the text files were cleaned in order to be analyzed by computer using corpus analysis software. In the process of cleaning, to reduce the noise in the corpus, all extra data including journal names, running heads, author names and affiliations, page numbers, DOIs, tables, and references were deleted from the text files. Given the large number of files, additional cleaning of the corpus for proper names used in the text was not undertaken. The resulted corpus contained 15569031 running words. The principled collection of the data and systematic selection of the research articles aimed at enhancing the representativeness of the corpus. The list of the selected journals is provided in Table 1.

Software and Base Lists for Analysis

The present study used AntWordProfiler (Anthony, 2014) for analyzing the applied linguistics research articles for the use of academic vocabulary. AntWordProfiler is designed for profiling the vocabulary level and the complexity of the texts. More specifically, the software compares the loaded corpora against available vocabulary lists. The General Service List (GSL) (West, 1953) (1000/2000) and the Academic Word List (570) (Coxhead, 2000) are the default word lists pre-loaded into the program, nonetheless it is possible to remove them and add other vocabulary lists such as BNC/COCA base lists. After analyzing the corpus, the software generates complete statistic and detailed frequency information that could be used in further analysis of vocabulary items. The base lists used in this study for analyzing the corpus included the NGSL (Browne, 2021) and NAWL lists that are created for vocabulary profiling using AntWordProfiler. To be used in AntWordProfiler program, the NGSL is divided into three sub-lists which is based on frequency and the coverage of vocabulary items. The first two sub-lists in the NGSL each contain 1000 words (i.e. lemmas) and the third sub-list contains 801 words. A supplementary list containing words for days, months, and numbers is also available. The NAWL also contains 963 words and has been created based on the

¹ SCImago. (n.d.). SCImago Journal & Country Rank. Retrieved April 19, 2021, from <https://www.scimagojr.com/journalrank.php>

Table 1

Selected journals for compiling the corpus

	Journal	No. of words
1	Modern Language Journal	955281
2	Studies in Second Language Acquisition	794531
3	Applied Linguistics	483062
4	System	520677
5	Language Testing	824622
6	TESOL Quarterly	632505
7	Language Learning	1298987
8	Language Teaching	948607
9	Language Teaching Research	882740
10	English for Specific Purposes	790603
11	English for Academic Purposes	695244
12	RELC	651656
13	ReCALL	842031
14	Computer Assisted Language Learning	572302
15	International Journal of Applied Linguistics	821755
16	Second Language Research	1015795
17	Journal of Second Language Writing	679206
18	Innovation in Language Learning and Teaching	733304
19	ELT Journal	505922
20	Annual Review of Applied Linguistics	920201
	Total	15569031

same principles as the AWL (Coxhead, 2000) which means that the list contains items beyond the NGSL.

Vocabulary Selection Criteria

Previous corpus-based studies investigating vocabulary use in different subject areas employed a variety of units for counting words that include types (orthographic forms), lemmas (defined as the base word plus its inflected forms in the same part of speech, for example the verb walk is considered a different lemma than the walk as a noun), flemmas (headword and inflected forms of different parts of speech, for example the flemma for the headword walk includes walk, walks (third person and plural noun), walking (in all parts of speech), and walked (past and past participle tenses)) and word families (base word plus its inflected forms and transparent derivations) (Bauer & Nation, 1993; Nation, 2016). The majority of the studies within this line of research employed level six word family (Nation, 2016) which is grounded in the assumption that the knowledge of the base word facilitates the understanding of its derived

and inflected forms (Coxhead, 2000; Xue & Nation, 1984). Nevertheless, a growing number of studies started to question this approach, and the use of lemma and flemma are gaining more attention in word list development as these units contain information on parts of speech, and hence regarded to be more appropriate for creating pedagogically useful lists (Brezina & Gablasova, 2015; Brown et al., 2020; Gardner & Davies, 2014; Lei & Liu, 2016). Responding to this debate, Nation (2016) argued that all employed units are indeed different levels of word families scale as delineated by Bauer and Nation (1993) where word types represent level 1, and widely used word families are in level 6. It is now well established that determining the unit of counting from different levels should be in line with the goals for list development. In this regard, lower levels including word types and lemmas are appropriate for productive uses of language (Dang, 2019; Durrant, 2014), and flemmas and word families are more suitable for receptive uses (Dang et al., 2017; Nation, 2016). In light of these considerations, the current study employed flemmas as the unit for counting academic vocabulary.

In order to further analyze the data and identify frequently used academic vocabulary in applied linguistics research articles, output from AntWordProfiler software was copied into a Microsoft Excel spreadsheet. Data analysis followed by using three criteria including specialized occurrence, range, and frequency employed by Coxhead (2000) in developing the academic word list (AWL). Moreover, given the variation in the number of running words in each of the 20 journals, a forth criterion namely dispersion was also used (Brezina, 2018). Based on the first criterion, academic vocabulary is operationalized as being beyond the general service or core vocabulary in English based on the New General Service List (Browne, 2021). As for range, words that occurred in all 20 journals and at least in 500 research articles were selected for further investigation. With respect to frequency, selected flemmas had to occur at least 28.5 times per million words as suggested by Coxhead (2000), which amounted to 440 times in the corpus with around 15.5 million running words. Finally, for the dispersion criteria, the flemmas that met the frequency threshold had to occur with a similar ratio (i.e. 28.5 per million words) in each of the journals.

RESULTS AND DISCUSSION

The results of lexical profile of the corpus are represented in the Table 2. As it is represented below, 10453140 tokens were in the first NGSL list that accounted for 67.1% of the corpus. Next, 1180431 tokens were identified in the second NGSL list. The coverage provided by these items was around 7.58% that indicates a considerable decrease in the coverage of the second base list. The third list provided 3.28% coverage and accounted for 510362 tokens. There were also 2977, 2656, and 1941 types in the corpus occurring in the three lists respectively. Regarding the academic vocabulary, the analysis revealed that 653192 tokens were identified in the New Academic Word List (NAWL). These items provided 4.19% coverage in the corpus, accounting for 2000 word types, and 955 flemmas. Around 0.52% of the corpus that included 81108 tokens was in the supplementary list containing the words for numbers, week days, and months. Finally, 2700798 tokens accounting for around 17.33% of the corpus

were beyond the lists of general service and academic vocabulary and contained proper nouns, in text used numbers, and low frequency vocabulary.

After applying the criteria for selecting the words (i.e. specialized occurrence, range, and frequency), 310 flemmas were selected as the academic vocabulary occurring frequently in the research articles published in the field of applied linguistics (Appendix A). These flemmas that were beyond NGSL items accounted for 587361 tokens, and provided 3.77% coverage in the corpus. The top 10 frequently occurring academic vocabulary included repertoire, classroom, linguistic, vocabulary, discourse, linguistics, feedback, lexical, none, and corpus. These flemmas provided around 1.1% coverage by accounting for 168890 tokens in the corpus. Moreover, further data analysis also revealed that 645 flemmas in the NAWL occurred infrequently in the corpus, and these items accounted for 65831 tokens, and only 0.42% of the entire corpus.

Comparing the findings to the earlier studies that investigated the academic vocabulary in research articles published in different subject areas, the current study found different results with respect to coverage of academic vocabulary. These differences mainly stem from using a different and improved core academic word list for profiling the corpus. In this regard, although the earlier studies reported around and more than 10% coverage for the AWL in medical (Chen & Ge, 2007), agriculture (Martínez et al., 2009), chemistry (Valipouri & Nassaji, 2013), and applied linguistics (Khani & Tazik, 2013; Vongpumivitch et al., 2009) research articles, the current study found just above 4% coverage for the academic words based on the NAWL. Nevertheless, since the base lists used for profiling the corpus in this study were different from those employed by the previous studies, these findings need to be interpreted in light of the differences and improvements in operationalizing general service words and academic vocabulary. As stated before, the NGSL and the NAWL are developed based on much larger and contemporary corpora compared to the old GSL (West, 1953) and the AWL (Coxhead, 2000). Additionally, the old age of the

Table 2

Statistics

FILE	TOKEN	TOKEN%	CUMTOKEN%	TYPE	GROUP
NGSL1	10453140	67.1	67.1	2977	1000
NGSL2	1180431	7.58	74.68	2656	1000
NGSL3	510362	3.28	77.96	1941	801
NAWL	653192	4.19	82.15	2000	955
Supplements	81108	0.52	82.67	98	48
-	2700798	17.33	100	127075	127075
TOTAL:	15569031				

GSL has resulted in classifying some currently in use and high frequent words (Nation, 2012) as academic vocabulary, and some items in the final list have also more general nature and are only marginally academic (Masrai & Milton, 2018). As a result, although the studies that employed the GSL and the AWL in profiling the research articles for academic vocabulary reported higher coverages, it should be noted that their final lists contained a considerable number of high-frequency vocabulary.

In order to further illuminate on the observed differences, and hence to better interpret the results obtained in the current study, a detailed comparison of the findings was conducted with Khani and Tazik (2013) and Gholaminejad and Anani Sarab (2020) that also investigated academic vocabulary in applied linguistics research articles and textbooks. As mentioned earlier, Khani and Tazik (2013) identified 773 words types (573 AWL, 200 non-GSL/AWL) that occurred frequently in the corpus, and provided 12.48% coverage. First, the list of 773 word types was analyzed against the NGSL (with three levels) and NAWL base lists. The results revealed that 22.44% of the words in the list occurred in the first NGSL, 30.8% in the second, and 14.96% in the third base lists. Totally, 68.2% of the academic vocabulary identified by Khani and Tazik (2013) were in fact general service and high frequent words based on New General Service List (NGSL) (Browne, 2021). Only 18.8% of the items in the list were identified as academic words based on the NAWL, and 13.72% of the items were beyond the base lists. Further analysis also indicated that almost 80% of the 537 frequently used AWL items in applied linguistics research articles belonged to the NGSL. Regarding the 200 non-GSL/AWL word types, it was found that 33% of these items were also general service words based on NGSL, and 31.44% were academic based on NAWL. Around 35.57% of these non-GSL/AWL items were beyond the base lists. By excluding the NGSL items from the list of 773 word types, the remaining words provided a coverage of 4% in the corpus, which is very similar to the results obtained in the current study for the new academic vocabulary. The findings are also in line with the previous studies that criticized the AWL for containing general words rather than academic vocabulary (Masrai & Milton, 2018).

Moreover, the comparison of the word list created in this study with Gholaminejad and Anani Sarab (2020) revealed that the two wordlists had only 66 similar words and around 20% overlap. Moreover, it was found that 195 (around 58%) of lemmas identified by Gholaminejad and Anani Sarab (2020) as academic words in applied linguistics belong to NGSL words, although it should be highlighted that these words have special meanings in the field. Finally, 75 lemmas (22%) were beyond the NGSL and NAWL items. One reason for the differences in the findings stems from using different word lists for representing high frequency vocabulary in English. In this regard, since Khani and Tazik (2013) used the GSL (West, 1953), and Gholaminejad and Anani Sarab (2020)

used the New-GSL (Brezina & Gablasova, 2015) for excluding high-frequency vocabulary items, their final lists contain different items compared to the present study that used the NGSL (Browne, 2021). Another factor contributing to the observed variation is related to the size and composition of the investigated corpora in the three studies. These findings underscore the need for more replication research in corpus-based wordlist development with a focus on investigating the contribution of the newly developed academic word lists in research articles (Coxhead, 2018).

The findings of the study have implications for vocabulary learning and teaching in EAP programs, and also for corpus-based studies of academic vocabulary. First, the results of the current study indicated that the use of academic vocabulary is highly affected by the nature of subject areas, and only 310 out of 960 flemmas in the NAWL were employed frequently by the researchers in the field of applied linguistics. This means that a common core view on academic vocabulary is problematic and has serious limitations (Hyland, 2013; Hyland & Tse, 2007). This is the case even with the newly developed and improved versions of the old academic vocabulary lists, as they cannot serve the needs of university students and researchers in different disciplines (Durrant, 2016). In this regard, there is a need to develop more restricted and disciplinary oriented academic vocabulary lists. Given the short span of most EAP courses, such endeavors can bring positive outcomes by aligning the courses with the learning needs of the students and more efficient use of time. In order to facilitate the setting of a vocabulary learning component in an EAP program in applied linguistics, the frequently used academic vocabulary in the field is divided into 6 bands based on their frequency. Unlike the majority of corpus-based studies that produced long lists of academic terminologies, the list of 310 flemmas presented in appendix A is both short and pedagogically useful, and might be covered during an academic semester. This might be best realized via integrating digital technologies into EAP courses for learning vocabulary (Xodabande & Atai, 2020; Zakian et al., 2022). Second, in light of the limitations associated with the old versions of the general service and academic vocabulary, there is a need to revisit earlier findings and test their results against the newly developed lists. The need for more replication studies in corpus-based word list development has been emphasized in the literature (Miller & Biber, 2015; Schmitt et al., 2017), nevertheless, it received far less attention in this line of research and remained a missing component in specialized vocabulary research (Coxhead, 2018). In this regard, by acknowledging the contributions of the earlier studies that enhanced our understanding with respect to the use of academic vocabulary in research articles, more research investigating larger corpora and new lists across various subject areas can provide the field with new insights and references for improving vocabulary learning and teaching.

CONCLUSION

The current study investigated the use of academic vocabulary in a large corpus of applied linguistics research articles. The findings revealed that the academic words provided around 4.19% coverage in the corpus, and that 310 out of 960 flemmas in NAWL were used frequently (Appendix A). These findings provided a different picture with respect to the contribution of academic vocabulary in research articles, as 4.19% percent coverage is significantly lower than the general coverage of 10% reported in the literature for a different list of academic vocabulary. In this regard, the study highlighted the need for more research focusing on the contribution of the recently developed academic word lists in research articles and other academic genres. The study had some limitations that should be acknowledged. One limitation relates to the representativeness of the analyzed corpus. Given the broad scope of the applied linguistics field and the various domains of language related courses and topics within this discipline, creating a well-balanced corpus is a daunting task for the researchers. Obtaining expert opinions for selecting journals, systematic sampling of a 2000 articles for creating a large corpus with around 15.5 million words, and extended time span of 10 years, all aimed at creating a well-compiled data-base for the analysis. Nevertheless, the broad scope of the field necessitates taking such limitations into account in interpreting the findings. Another limitation stems from the operationalizing academic vocabulary as those vocabulary items that are beyond general service and core vocabulary in English. Recently this view has been challenged, and it has been argued that academic vocabulary cuts across high-, mid-, and low-frequency words. In this regard, EAP teachers, researchers, and university students in applied linguistics should bear in mind that the boundaries between general service, academic, and technical vocabulary is not clear cut and as neatly defined by vocabulary researchers. With all these limitations, the findings of the current study contribute to the existing body of knowledge in vocabulary studies for educational purposes, and highlights the importance of replication research in light of the recent developments in corpus-based pedagogy in EAP. Considering the significant role of academic vocabulary, future studies might consider investigating the use of such words in the writings of university students and not only expert users and established researchers. This research direction can shed more light on processes involved in learning and using academic words.

ACKNOWLEDGEMENTS

We would like to convey our special thanks to the editor and the four anonymous reviewers of the Journal of Language and Education for their insightful comments on earlier versions of the manuscript.

DECLARATION OF COMPETING INTEREST

None declared

REFERENCES

- Airey, J., Lauridsen, K. M., Räsänen, A., Salö, L., & Schwach, V. (2017). The expansion of English-medium instruction in the Nordic countries: Can top-down university language policies encourage bottom-up disciplinary literacy goals? *Higher Education*, 73(4), 561–576. <https://doi.org/10.1007/s10734-015-9950-2>
- Bauer, L., & Nation, I. S. P. (1993). Word Families. *International Journal of Lexicography*, 6(4), 253–279. <https://doi.org/10.1093/ijl/6.4.253>
- Bazerman, C., Keranen, N., & Prudencio, F. E. (2012). Facilitated Immersion at a Distance in Second Language Scientific Writing. In M. Castelló & C. Donahue (Eds.), *University Writing: Selves and Texts in Academic Societies* (pp. 235–248). Brill. https://doi.org/10.1163/9781780523873_014
- Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing Words to Life: Robust Vocabulary Instruction* (2nd ed.). Guildford Press.
- Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge University Press.
- Brezina, V., & Gablasova, D. (2015). Is There a Core General Vocabulary? Introducing the New General Service List. *Applied Linguistics*, 36(1), 1–22. <https://doi.org/10.1093/applin/amt018>
- Brown, D., Stoeckel, T., Mclean, S., & Stewart, J. (2020). The Most Appropriate Lexical Unit for L2 Vocabulary Research and Pedagogy: A Brief Review of the Evidence. *Applied Linguistics*. <https://doi.org/10.1093/applin/amaa061>
- Browne, C. (2014). The New General Service List Version 1.01: Getting Better All the Time. *Korea TESOL Journal*, 11(1), 35–50.
- Browne, C. (2021). The NGSL Project: Building Wordlists and Resources to help EFL Learners (and Teachers) to Succeed. *JALTCALL Publications, PCP2020*, 1. <https://doi.org/10.37546/JALTSIG.CALL2020.1>
- Chen, Q., & Ge, G. chun. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes*, 26(4), 502–514. <https://doi.org/10.1016/j.esp.2007.04.003>
- Corcoran, J. (2017). The Potential and Limitations of an Intensive English for Research Publication Purposes Course for Mexican Scholars. In M. J. Curry & T. Lillis (Eds.), *Global Academic Publishing: Policies, Perspectives and Pedagogies* (pp. 217–232). Multilingual Matters. <https://doi.org/10.21832/9781783099245-021>
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213–238. <https://doi.org/10.2307/3587951>

- Coxhead, A. (2011). The Academic Word List 10 Years On: Research and Teaching Implications. *TESOL Quarterly*, 45(2), 355–362. <https://doi.org/10.5054/tq.2011.254528>
- Coxhead, A. (2018). *Vocabulary and English for Specific Purposes Research: Quantitative and Qualitative Perspectives*. Routledge.
- Coxhead, A. (2019). Academic vocabulary. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 97–110). Routledge. <https://doi.org/10.4324/9780429291586-7>
- Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16(3), 129–147. <https://doi.org/10.1016/j.jslw.2007.07.002>
- Coxhead, A., & Nation, I. S. P. (2001). The specialised vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), *Research Perspectives on English for Academic Purposes* (pp. 252–267). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524766.020>
- Dang, T. N. Y. (2019). Corpus-Based Word Lists in Second Language Vocabulary Research, Learning, and Teaching. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 288–303). Routledge. <https://doi.org/10.4324/9780429291586-19>
- 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>
- Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, 43, 49–61. <https://doi.org/10.1016/j.esp.2016.01.004>
- Evans, S., & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes*, 6(1), 3–17. <https://doi.org/10.1016/j.jeap.2006.11.005>
- Evans, S., & Morrison, B. (2010). The first term at university: implications for EAP. *ELT Journal*, 65(4), 387–397. <https://doi.org/10.1093/elt/ccq072>
- Evans, S., & Morrison, B. (2011). Meeting the challenges of English-medium higher education: The first-year experience in Hong Kong. *English for Specific Purposes*, 30(3), 198–208. <https://doi.org/10.1016/j.esp.2011.01.001>
- Flowerdew, J. (2015). Some thoughts on English for Research Publication Purposes (ERPP) and related issues. *Language Teaching*, 48(2), 250–262. <https://doi.org/10.1017/S0261444812000523>
- Gardner, D., & Davies, M. (2014). A New Academic Vocabulary List. *Applied Linguistics*, 35(3), 305–327. <https://doi.org/10.1093/applin/amt015>
- Huntley, H. (2006). *Essential Academic Vocabulary: Mastering the Complete Academic Word List*. Houghton Mifflin Company.
- Hyland, K. (2002). Specificity revisited: how far should we go now? *English for Specific Purposes*, 21(4), 385–395. [https://doi.org/10.1016/S0889-4906\(01\)00028-X](https://doi.org/10.1016/S0889-4906(01)00028-X)
- Hyland, K. (2006). *English for Academic Purposes: An advanced resource book*. Routledge.
- Hyland, K. (2009). *Academic Discourse: English In A Global Context*. Bloomsbury Academic.
- Hyland, K. (2013). Writing in the university: education, knowledge and reputation. *Language Teaching*, 46(1), 53–70. <https://doi.org/10.1017/S0261444811000036>
- 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>
- Khani, R., & Tazik, K. (2013). Towards the development of an academic word list for applied linguistics research articles. *RELJ Journal*, 44(2), 209–232. <https://doi.org/10.1177/0033688213488432>
- Laufer, B. (1996). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition: A Rationale for Pedagogy* (pp. 20–34). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524643.004>
- Laufer, B. (2005). Focus on Form in Second Language Vocabulary Learning. In S. H. Foster-Cohen, M. del P. G. Mayo, & J. Cenoz (Eds.), *EUROSLA Yearbook: Volume 5* (pp. 223–250). John Benjamins Publishing Company. <https://doi.org/10.1075/eurosla.5.11lau>
- 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, Y., & Flowerdew, J. (2020). Teaching English for Research Publication Purposes (ERPP): A review of language teachers' pedagogical initiatives. *English for Specific Purposes*, 59, 29–41. <https://doi.org/10.1016/j.esp.2020.03.002>
- Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes*, 39, 1–11. <https://doi.org/10.1016/j.esp.2015.03.001>
- Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes*, 28(3), 183–198. <https://doi.org/10.1016/j.esp.2009.04.003>
- 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>
- McLean, S., & Kramer, B. (2015). The creation of a New Vocabulary Levels Test. *Shiken*, 19(2), 1–11. http://teval.jalt.org/sites/teval.jalt.org/files/19-02-1_McLean_Kramer.pdf
- Miller, D., & Biber, D. (2015). Evaluating reliability in quantitative vocabulary studies: The influence of corpus design and composition. *International Journal of Corpus Linguistics*, 20(1), 30–53. <https://doi.org/10.1075/ijcl.20.1.02mil>
- Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139858656>

- Nation, I. S. P. (2016). *Making and Using Word Lists for Language Learning and Testing*. John Benjamins Publishing Company.
- Paquot, M. (2010). *Academic Vocabulary in Learner Writing: From Extraction to Analysis*. Continuum International Publishing Group.
- Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching*, 50(2), 212–226. <https://doi.org/10.1017/S0261444815000075>
- 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>
- Sorell, J. (2013). *A study of issues and techniques for creating core vocabulary lists for English as an international language*. Unpublished PhD thesis, Victoria University of Wellington, New Zealand.
- Spencer, S., Clegg, J., Lowe, H., & Stackhouse, J. (2017). Increasing adolescents' depth of understanding of cross-curriculum words: an intervention study. *International Journal of Language & Communication Disorders*, 52(5), 652–668. <https://doi.org/10.1111/1460-6984.12309>
- Stubbs, M. (1986). Language development, lexical competence and nuclear vocabulary. In M. Stubbs (Ed.), *Educational Linguistics* (pp. 98–115). Blackwell.
- Valipouri, 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>
- Vongpumivitch, V., Huang, J. yu, & Chang, Y. C. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28(1), 33–41. <https://doi.org/10.1016/j.esp.2008.08.003>
- Wang, J., Liang, S. lan, & Ge, G. chun. (2008). Establishment of a Medical Academic Word List. *English for Specific Purposes*, 27(4), 442–458. <https://doi.org/10.1016/j.esp.2008.05.003>
- Webb, S., & Nation, I. S. P. (2017). *How Vocabulary is Learned*. Oxford University Press.
- Wells, L. (2007). *Vocabulary Mastery 1: Using and Learning the Academic Word List*. University of Michigan Press.
- West, M. (1953). *A general service list of English words*. Longman, Green & Co.
- Woodward-Kron, R. (2008). More than just jargon – the nature and role of specialist language in learning disciplinary knowledge. *Journal of English for Academic Purposes*, 7(4), 234–249. <https://doi.org/10.1016/j.jeap.2008.10.004>
- Xodabande, I., & Atai, M. R. (2020). Using mobile applications for self-directed learning of academic vocabulary among university students. *Open Learning: The Journal of Open, Distance and e-Learning*, 1–18. <https://doi.org/10.1080/02680513.2020.1847061>
- Xodabande, I., & Xodabande, N. (2020). Academic Vocabulary in Psychology Research Articles: A Corpus-Based Study. *MEXTESOL Journal*, 44(3), 1–21.
- Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, 3, 215–229.
- Yang, M. N. (2015). A nursing academic word list. *English for Specific Purposes*, 37(1), 27–38. <https://doi.org/10.1016/j.esp.2014.05.003>
- Zakian, M., Xodabande, I., Valizadeh, M., & Yousefvand, M. (2022). Out-of-the-classroom learning of English vocabulary by EFL learners: investigating the effectiveness of mobile assisted learning with digital flashcards. *Asian-Pacific Journal of Second and Foreign Language Education*, 7(1), 1–16. <https://doi.org/10.1186/s40862-022-00143-8>

APPENDIX

Academic Vocabulary in AL

Band 1: repertoire, classroom, linguistic, vocabulary, discourse, linguistics, feedback, lexical, none, corpus, cognitive, bilingual, comprehension, aspect, grammatical, oral, pre, explicit, semantic, publish, accuracy, impact, competence, pragmatic, syntactic, curriculum, communicative, translation, questionnaire, psychology, textbook, implicit, domain, usage, verbal, empirical, dynamic, statistical, mediate, appendix, correlation, qualitative, facilitate, intermediate, methodology, dictionary, phonological, accent, utterance, assignment.

Band 2: conference, tutor, autonomy, obtain, orient, criteria, thesis, vowel, distribution, tense, sub, stimulus, retrieve, validity, cue, reliability, semester, effectiveness, conceptual, interact, quantitative, statistics, marker, bundle, undergraduate, orientation, variance, syntax, metaphor, audio, correction, dissertation, morphological, stance, media, developmental, novice, norm, occurrence, similarity, embed, longitudinal, syllable, interface, candidate, diverse, statistically, explicitly, integration, correlate.

Band 3: faculty, regression, meaningful, multi, variability, productive, ideology, overview, dominant, plural, lecturer, morphology, namely, parameter, prediction, informal, overlap, elementary, paradigm, chunk, systematic, initiate, commonly, partial, ex, syllabus, standardize, comparative, manuscript, utilize, gram, sensitivity, practitioner, linear, deviation, span, problematic, behavioral, correctly, complement, elaborate, temporal, indicator, workshop, inclusion, evident, strategic, null, onset, precede.

Band 4: expertise, mid, duration, trajectory, par, scenario, conscious, click, importantly, positively, transcription, encode, transcribe, comparable, protocol, synthesis, discrimination, differential, semi, forum, accurately, mentor, generalize, inference, consonant, classify, generalization, consciousness, singular, logical, intensive, transformation, constrain, activate, emergence, interval, threshold, valid, simultaneously, adolescent, indirect, correspondence, portfolio, independently, thereby, particle, minimal, strand, socially, replication.

Band 5: clarify, identical, entity, critique, actively, authority, spontaneous, reinforce, separately, neutral, node, nominal, descriptor, preliminary, nonetheless, randomly, articulate, conception, junior, simulation, likelihood, matrix, indigenous, hedge, dialect, diagnostic, likewise, spatial, individually, interestingly, differentiate, bound, manual, vocabulary, dominance, rhetoric, lab, partially, consent, micro, proposition, ecological, pi, coefficient, critically, coordinate, disadvantage, graph, trait, facet.

Band 6: artifact, consensus, broadly, depict, admission, prominent, manipulate, pronounce, availability, hierarchy, classification, integral, identification, collective, conditional, optimal, seminar, globalization, adaptation, disability, competent, ethical, sophisticate, replicate, legitimate, contrary, slot, essentially, neural, formulation, subset, induce, superior, selective, ultimate, subjective, scholarship, postgraduate, exploit, congruent, motive, trans, ecology, progression, adaptive, detection, maximize, symbolic, minimize, render, readily, probe, stereotype, assert, marginal, campus, coherent, denote, interviewer, manipulation.