THE JOURNAL OF ASIA TEFL
Vol. 18, No. 3, Fall 2021, 959-966
http://dx.doi.org/10.18823/asiatefl.2021.18.3.15.959

The Journal of Asia TEFL

http://journal.asiatefl.org/

e-ISSN 2466-1511

General Service and Academic Words in Psychology Research Articles: A Corpus Based Study

Mohammadreza Valizadeh
Cappadocia University, Cappadocia, Turkey.

Ismail Xodabande
Kharazmi University, Tehran, Iran.

Introduction

Acquiring vocabulary is associated with various challenges for foreign language learners (Nation, 2013). It has been argued that vocabulary knowledge is a good predictor of reading comprehension (Laufer, 1996; Qian, 2002) and correlates positively with scores obtained on general proficiency tests (Alderson, 2006). Moreover, insufficient vocabulary has been regarded among the most serious problems English for Academic Purposes (EAP) students face (Evans & Morrison, 2010, 2011). Nowadays, Master and PhD students in English as a foreign language (EFL) contexts are required to read and publish research articles in international journals in English (Flowerdew, 2015). In this regard, they constantly face serious linguistic barriers (Corcoran, 2017; Li & Flowerdew, 2020), and limited vocabulary knowledge further adds to their plight and inability to read and write in English, with some consequences for their participation in the scientific communities (Bazerman et al., 2012). Moreover, in most EFL contexts, many university students fail to learn the 2000 most frequent word families in English (Webb & Chang, 2012), and therefore fall far short of the vocabulary demands of reading academic texts in English (Ward, 2009). Hence, the focus on the vocabulary learning needs of university students is an important undertaking for a number of reasons. First, given the current emphasis on promoting disciplinary literacy (Airey et al., 2017), such a focus can help material developers, teachers, and students to identify the most salient and important vocabulary items for different subject areas. Second, as vocabulary instruction receives insufficient attention in language courses (Webb & Nation, 2017), developing vocabulary lists for different disciplines can help students in self-directed and autonomous learning of those items. And third, with such a minimal focus on vocabulary in the classroom, EAP teachers can make their instruction more congruent with the needs of the students, and make the most out of classroom time. The current study reports some findings from a larger corpus-based study of psychology research articles. Given the limited empirical attention to the discipline of psychology in university vocabulary studies (Coxhead, 2018), the findings can help teachers and students in EAP and English for Research Publication Proposes (ERPP) courses in psychology.

Academic Vocabulary

Over the past years, a number of word lists have been developed to assist university students in EAP programs (Coxhead, 2000; Gardner & Davies, 2014; Xue & Nation, 1984). In this line of enquiry, the Academic Word List (AWL) attracted considerable attention and remained as a major source for EAP instruction, materials development, and vocabulary assessment (Coxhead, 2011). Nonetheless, a growing number of studies started to challenge its usefulness as an academic word list (Gardner & Davies, 2014; Hyland & Tse, 2007; Masrai & Milton, 2018). Consequently, the AWL has been criticized on various grounds, including its general rather than specialized nature (Masrai & Milton, 2018), the amount of variation in the use and coverage provided in different disciplines (Chen & Ge, 2007; Liu & Han, 2015; Martínez et al., 2009), and using the outdated GSL (West, 1953) as the source for general service vocabulary (Gardner & Davies, 2014). Recently, in response to these challenges, two additional lists, namely the New Academic Word List (NAWL), and the Academic Vocabulary List (AVL) have been introduced to the field (Browne et al., 2013b; Gardner & Davies, 2014); nevertheless, research investigating their contributions has remained limited (Durrant, 2016).

Moreover, although the research article genre attracted considerable attention in academic vocabulary studies (Chen & Ge, 2007; Khani & Tazik, 2013; Martínez et al., 2009; Valipouri & Nassaji, 2013; Vongpumivitch et al., 2009; Wang et al., 2008; Yang, 2015), the field of psychology has remained under-explored, and only a limited number of studies have investigated the vocabulary profile of different texts in this field (Akbarian et al., 2017; Safari, 2018; Xodabande, 2020; Xodabande & Xodabande, 2020). Xodabande (2020) analyzed a corpus of 30 introductory and general psychology textbooks with around 14.8 million words. Using the BNC/COCA frequency lists (Nation, 2012), the study identified 2048 word families providing around 82.66% coverage of the corpus. For pedagogical purposes, the study divided the list into high- and mid-frequent categories. The 1294 high-frequent words (in the 2000 most frequent words in BNC/COCA lists) provided 71.6%, and 754 mid-frequent (in the third to ninth BNC/COCA lists) provided 11.06% coverage in the entire corpus. In another study investigating psychology research article abstracts, Akbarian et al. (2017) reported that the GSL and the AWL items provided around 88.22% coverage in their corpus containing 307,126 words, while the contribution of the AWL items was 15%.

Two studies in particular investigated the use of the GSL/AWL items in psychology research articles. In this regard, Safari (2018) analyzed a corpus of psychology research articles with 3.4 million words. The findings indicated that 1,077 GSL and 95 AWL word families were used infrequently in the corpus. The study further identified 189 non-GSL/AWL words which occurred frequently, and compiled a list of 1587 most frequent word families for the psychology discipline. Although the study did not provide specific information regarding the coverage of the GSL and the AWL in the corpus, the author claimed that the developed psychology word list improved the coverage provided by the aforementioned lists by 2.2%, nonetheless, it contained many fewer items. Xodabande and Xodabande (2020) investigated a corpus of 8,500 psychology research articles with 74 million words. Their findings indicated that the GSL items accounted for 72.08% of the corpus. The coverage of the AWL items was 13.12%, and the cumulative coverage of the two lists accounted for 85.2% of the words in psychology research articles. The study further identified 693 word types beyond the GSL and the AWL but occurring frequently and covering 6.1% of the words in the corpus. As mentioned earlier, both the GSL and the AWL have been criticized recently (Gardner & Davies, 2014) and there remains a need to investigate the coverage of newly developed general service and academic word lists such as the New General Service List (NGSL) (Browne et al., 2013a), and the New Academic Word List (NAWL) (Browne et al., 2013b) in academic texts. The NGSL is developed using a 273 million-word subsection of the two billion-word Cambridge English Corpus (CEC), and provides considerable improvements over the old GSL. Similarly, the NAWL was created based on a corpus of 288 million words featuring different spoken and written academic texts. In order to fill part of this gap in the existing literature, the current study sets out to investigate the lexical profile of psychology research articles based on the NGSL and NAWL, and aims to identify the most frequent and hence useful words in the corpus with a high pedagogical value for students and teachers in EAP and ERPP courses.

The Study

The Corpus and Software for Corpus Analysis

Using the criteria proposed by Sinclair (2005) in terms of size, balance, and representativeness, a corpus containing 8,500 research articles was analyzed. In this regard, using AntCorGen software (Anthony, 2019), 20,000 psychology research articles were collected from PLOS ONE database. The collected articles were a representative sample of the research article genre, and covered the major sub-areas of the field including cognitive, developmental, and social psychology. Next, with the aim of creating a manageable corpus, 8500 articles were selected randomly and assigned to 20 sub-corpora, each containing 425 research articles and about 3,700,000 running words. AntWordProfiler (Anthony, 2014) was also used for lexical profiling of the articles. This software is a powerful tool that compares the texts against a set of vocabulary lists loaded into the program and generates complete statistics and frequency information.

Data Analysis

The output from AntWordProfiler was copied into a Microsoft Excel spreadsheet and used to identify frequently-used NGSL and NAWL items in the corpus. In this regard, following Coxhead (2000), three criteria of range, frequency, and specialized occurrence were used. As for range, NGSL and NAWL items that occurred in all 20 sub-groups of the corpus were included in the list of the most frequent items in psychology research articles. For frequency, the selected words had to occur at least 2100 times in the entire corpus, and at least 105 times in each of the sub-corpora (28.5 times in a million words). For specialized occurrences, the academic words had to be outside the most frequently occurring words in English based on the NGSL. Regarding the unit of counting for the words, the study used lemmas defined as the base word plus its inflected forms (Nation, 2016). This is in line with a growing number of studies that argued for the use of lemmas as a more appropriate unit for creating pedagogically useful word lists (Brezina & Gablasova, 2015; Brown et al., 2020; Gardner & Davies, 2014; Lei & Liu, 2016).

Results and Discussion

Table 1 shows the lexical profile of psychology research articles based on the NGSL and NAWL. As it is represented, there were 74,016,481 running words (tokens) in the corpus. The first NGSL provided 67.43% coverage, followed by 9.5%, and 4.14% coverage for the second and the third NGSL lists respectively. Taken together, the NGSL provided around 81.07% coverage, and accounted for 7,932 types, and 2,801 lemmas. Regarding the NAWL, the findings indicated that there were 3,609,410 tokens in the corpus providing 4.88% coverage. There were also 2,128 NAWL words types, and 963 lemmas in the corpus. The supplemental list (days of the week, months, numbers) provided 0.57% coverage. And finally, 13.49% of the tokens were beyond the abovementioned lists.

TABLE 1
Lexical Profile of Psychology Research Articles based on NGSL and NAWL

	TOKEN	TOKEN%	CUMTOKEN%	TYPE	GROUP
1st_NGSL	49909814	67.43	67.43	3058	1000
2nd_NGSL	7029764	9.5	76.93	2781	1000
3rd_NGSL	3061573	4.14	81.07	2093	801
NAWL	3609410	4.88	85.95	2128	963
Supplemental	419860	0.57	86.52	108	49
Not in the lists	9986060	13.49		142945	142945
Total	74016481				

The findings indicate that the NGSL provides better coverage in research articles than the GSL. Xodabande and Xodabande (2020) reported 72.08% coverage for the GSL in psychology research articles; nonetheless, the current study shows that the NGSL provides 8.99% more coverage of the same corpus. This is also 7.85% higher than the coverage of GSL items in psychology research abstracts (Akbarian et al., 2017). Regarding the academic words, the previously mentioned studies reported 13.12%, and 15% coverage for the AWL respectively, which are considerably higher than the 4.88% for the NAWL in the corpus analyzed for this study. These findings further support the claims made by other researchers that some of the AWL items should be considered as general service words rather than academic vocabulary (Hyland & Tse, 2007; Masrai & Milton, 2018).

Regarding the frequently-used NGSL and NAWL items in psychology research articles, the findings revealed that 851 lemmas in NGSL1, 531 in NGSL2, and 300 in NGSL3 met the criteria of frequency and range set for the current study. These 1,682 word families accounted for 59,112,388 tokens in the corpus and provided around 79.86% coverage. This means the remaining 1,119 word families in the NGSL provide only 1.21% coverage. As for the NAWL, 361 lemmas also met the criteria and provided 4.32% coverage. The other 602 NAWL lemmas occurred very infrequently with the coverage of 0.58%. Taking the lists together, 2,043 NGSL and NAWL accounted for 84.20% of the corpus, leaving only 1.79% coverage for the 1,721 less frequent NGSL and NAWL words. These findings underscore the specificity of language used in research articles and are in agreement with previous studies investigating this genre in that some general service and academic words are less important for students and researchers in specific fields (Hyland & Tse, 2007; Li & Qian, 2010; Martínez et al., 2009). The findings also highlight that in EAP and ERPP courses, not all words are of equal importance. A complete list of the most frequently-used NGSL/NAWL items is provided in the supplementary file available for download using the following link: https://gofile.io/d/dOlnzS.

Implications

The current study has some implications for teaching vocabulary in EAP and ERPP programs in psychology. First, as the findings revealed, the newly developed general service and academic word lists (Browne et al., 2013a, 2013b) demonstrate considerable improvements over the old GSL (West, 1953) and the AWL (Coxhead, 2000) in terms of lexical coverage in the research articles. Given the fact that the majority of published materials for EAP is based on the AWL (Huntley, 2006; Schmitt & Schmitt, 2005, 2011), language teachers need to bear in mind that some of the vocabulary items introduced in those materials are indeed general service words. In this regard, there is considerable benefit in teaching NAWL items to university students, as this ensures that students can learn valuable academic vocabulary to assist their reading and writing (Xodabande & Atai, 2020). Second, as it was the case for the GSL and AWL, even in using the more updated and well-developed general and academic word lists to set vocabulary learning goals in EAP/ERPP programs, vocabulary items have different values depending on different disciplines. Considering that learning around 3,800 lemmas in the NGSL and NAWL is a difficult undertaking for many university students in EFL contexts, identifying a limited number of words with high value ensures efficient use of time and other resources to meet students' language learning needs. Since many students in such contexts need to read only discipline-related texts, more restricted and disciplinary word lists benefit them greatly. Finally, as corpus studies reveal important information about the use of language in different areas of academic discourse, it seems that there is a need for corpus linguistic courses in English language teacher education programs. This can raise language teachers' corpus literacy which is considered as a valuable addition to their toolbox (Callies, 2019), and is particularly important in EAP/ERPP courses.

Conclusions

The current corpus-based study investigated the lexical profile of psychology research articles based on NGSL and NAWL, and developed a list of frequently-used general and academic words for psychology students. The findings indicated that 1,682 general service and 361 academic words occurred frequently in the corpus, providing a total coverage of 84.20%. In this regard, the study provides a valuable list of vocabulary items for psychology that can be used in instructional programs and materials development in EAP/ERPP courses. Nevertheless, the study had some limitations. First, the analyzed corpus contained only one academic genre, namely research articles. In order to better understand the use and coverage of general and academic words in a given discipline, a well-balanced corpus featuring various academic texts is needed. Second, the study was descriptive and quantitative, providing only a partial picture of vocabulary use in research articles. Since the use of general and academic words is related to communicative practices in different disciplines, there is a need to investigate rhetorical practices associated with such words. Future research can address these areas and investigate the behavior of individual words in academic texts.

The Authors

Mohammadreza Valizadeh holds a PhD in TEFL from Gazi University, Ankara, Turkey and is currently an assistant professor at Cappadocia University, Cappadocia, Turkey. He has published several articles in different scholarly journals. His areas of interests include corrective feedback, flipped learning, learning strategies, learner autonomy, task-based language teaching and individual differences in language learning.

Department of Translation and Interpretation (English)
Faculty of Humanities
Cappadocia University
Cappadocia, Nevşehir, Ürgüp,
Turkey, 50420.

Ismail Xodabande (corresponding author) is a PhD candidate of Applied Linguistics at Kharazmi University, Tehran, Iran. He is interested in doing research on the role of various educational technologies in foreign language education, Mobile Assisted Language Learning (MALL), and the application of corpus linguistics in ELT. He has published in these areas in Cogent Education, Computer Assisted Language Learning, Open Learning, and Journal of Computers in Education.

Department of Foreign Languages
Kharazmi University
No. 43, Shahid Mofatteh Street
Tehran, Iran.
Email: ismail.kh.tefl@gmail.com

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

Akbarian, I., Ghanbarzadeh, Z., & Afzali Shahri, M. (2017). Profiling vocabulary in Psychology journal abstracts: A Comparison between Iranian and Anglo-American Journals. *Iranian Journal of Language Teaching Research, 5*(1), 51-69.

Alderson, J. C. (2006). *Diagnosing foreign language proficiency: The interface between learning and assessment.* Continuum.

Anthony, L. (2014). *AntWordProfiler*. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software/antwordprofiler/

Anthony, L. (2019). *AntCorGen (1.1.2).* Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software/antcorgen/

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

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., Culligan, B., & Phillips, J. (2013a). *New General Service List (NGSL).* www.newgeneralservicelist.org

Browne, C., Culligan, B., & Phillips, J. (2013b). *The New Academic Word List.* http://www.newgeneralservicelist.org/nawl-new-academic-word-list

Callies, M. (2019). Integrating corpus literacy into language teacher education: The case of learner corpora. In S. Götz & J. Mukherjee (Eds.), *Learner corpora and language teaching* (pp. 245-263). John Benjamins Publishing Company. https://doi.org/10.1075/scl.92.12cal

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.

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., & 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., & 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. *RELC 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

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

Li, Y., & Qian, D. D. (2010). Profiling the Academic Word List (AWL) in a financial corpus. *System*, *38*(3), 402-411. https://doi.org/https://doi.org/10.1016/j.system.2010.06.015

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

Nation, I. S. P. (2012). *The BNC/COCA word family lists*. https://www.victoria.ac.nz/__data/assets/pdf_file/0004/1689349/Information-on-the-BNC_COCA-word-family-lists-20180705.pdf

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.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An Assessment Perspective. *Language Learning*, *52*(3), 513-536. https://doi.org/10.1111/1467-9922.00193

Safari, M. (2018). Do university students need to master the GSL and AWL words: A psychology word list. *Journal of Modern Research in English Language Studies*, *5*(2), 101-122. https://doi.org /10.30479/jmrels.2019.10266.1275

Schmitt, D., & Schmitt, N. (2005). *Focus on vocabulary: Mastering the academic word list*. Longman.

Schmitt, D., & Schmitt, N. (2011). *Focus on vocabulary 2: Mastering the academic word list.* Pearson Education.

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. Y., & 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. L., & Ge, G. C. (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

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, *28*(3), 170-182. https://doi.org/10.1016/j.esp.2009.04.001

Webb, S., & Chang, A. C.-S. (2012). Second language vocabulary growth. *RELC Journal*, *43*(1), 113-126. https://doi.org/10.1177/0033688212439367

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.

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

Xodabande, I. (2020). Investigating the vocabulary load of psychology textbooks: A corpus-based study. *Cambridge Open Engage*. https://doi.org/10.33774/coe-2020-jfvvd

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

(Received May 23, 2021; Revised August 20, 2021; Accepted September 18, 2021)

