Journal of Academic Language & Learning
Vol. 18, No. 1, 2024, 21-48. ISSN 1835-5196

The Applied Linguistics Academic Word List (ALAWL): The creation of a domain-specific word list for applied linguistics

Ana K. Barraza, Grant Eckstein, Zachary M. Lambert, Carolee Rogers, and Jesse Vincent

Department of Linguistics, Brigham Young University, 4064 JFSB Provo Utah 84602, United States

Email: akb3713@byu.edu, grant_eckstein@byu.edu, zacklamb@byu.net, carolee_rogers@byu.edu, and vincenjes@gmail.com

(Received 28 August, 2023. Published online 9 February, 2024.)

Because specialized vocabulary knowledge is an essential component of disciplinary expertise, scholars in many domains, such as agriculture, economics, and medicine, have developed robust, register-specific word lists to supplement general service and academic word lists. However, similar lists for applied linguistics are less robust and may provide superficial vocabulary options. We developed an applied linguistics academic word list (ALAWL) of 664 words by sampling more than 48 million tokens taken from academic articles in 12 leading applied linguistics journals from the years 2012 to 2021. Our list covers 5.34% of words in a reference corpus of nearly 3 million words and, when combined with the new general service list and new academic word list, provides greater coverage than earlier attempts at such a list, accounting for 93.49% of words in the corpus. We describe our methods and present the ALAWL as an option for learners and educators in the field of applied linguistics and TESOL.

Key Words: Word list; applied linguistics; corpus; TESOL; academic word list; general service list.

1. Introduction

Vocabulary knowledge is an essential component of language learning, particularly because second language (L2) learners need a large vocabulary in order to communicate effectively in the second language (Nation, 2006). However, second language (L2) learners typically do not have the time to naturally assimilate large amounts of vocabulary incidentally from their environment as is more typical when acquiring a first language (L1). Thus, a major concern for teachers is finding ways to facilitate the vocabulary learning process for their students. This is often accomplished through intentional learning or instruction (Barcroft, 2009) and using vocabulary lists to help students focus their efforts on important, high-frequency words (Gardner, 2013). Researchers and practitioners have laboured to develop such lists for language-learning purposes.

When discussing word lists and their creation, it is useful to first define certain terminology used in corpus linguistics. While "word" is often used to refer to a unit of language that conveys meaning, there are times when more specific terms such as "type", "token", and "lemma" are needed to accurately differentiate between the forms of words being discussed. Gardner has defined "type" as "one or more contiguous letters of the Roman alphabet that form a distinct word" and defined "token" as "the frequency ... of types or other word units within a text" (2013, p. 9). Therefore, in the sentence *A bird stood on the bird feeder*, there are six types and seven tokens.

A "lemma" is a set of one or more types with the same base form and part of speech and for which spelling only changes for grammatical purposes (Gardner, 2013). An example of a set of words that are considered one lemma are "play", "plays", "playing", and "played"; "player" is considered a different lemma as it is a noun rather than a verb. A lemma may also be called a "headword". These terms are often used in corpus research and in the creation of word lists such as the GSL and NGSL.

The earliest English list to gain broad appeal was the GSL or General Service List (West, 1953), which was originally developed with 2,000 headwords, each representing a word family and based on objective and subjective inclusion criteria (Gilner, 2011). Some word families included words that should have been their own headwords, like *efficient* being listed under *effect* (see Bauman & Culligan, 1995) and thus the list was revised in the mid-1990s to 2,284 headwords. With increasing computational power for corpus techniques to identify relevant high-frequency words (Gardner, 2013), the New General Service List (NGSL) was created by Browne et al. (2013) as a more modern replacement for the GSL. The organizational structure was also changed from word families to modified lemmas in which lemma groups included all inflected forms of a headword (Browne, 2014) rather than more distantly related forms that tend to occur in word families (e.g., *unpleasantly* included with *please*). In comparing the two lists, the NGSL includes 92% of words found in most general English texts, whereas the GSL covers 84%. The advantage of the newer lemmatized list is that it offers greater vocabulary coverage with fewer headwords while taking advantage of basic grammatical knowledge since learners need only to master the headword and its inflected forms.

About this same time, Brezina and Gablasova (2015) created their own version of the GSL – also called the new General Service List. Brezina and Gablasova believed there were flaws in the qualitative criteria used to develop the GSL and so created their own using lemmas rather than word families. Their updated list was developed using four different corpora and included 2,494 items, 2,116 of which are shared by all four corpora. It claims to have similar coverage as the GSL but with fewer words.

Other applications of corpus linguistics have led to the development of the Academic Word List (AWL) which was organized by word families like the original GSL. The AWL was created by Coxhead in 2000 to work in conjunction with the GSL, representing the most common word families outside of the GSL that a first-year university student would encounter in their readings. A specialized corpus of 3.5 million words across arts, commerce, law, and science was used to create the list, which accounted for roughly 10% of the total vocabulary across the four disciplines. Despite being widely known and used, the AWL has been criticized for being an appendage of the GSL (assuming GSL competence) and for using word families instead of lemmas (Gardner and Davies, 2014). In response to these criticisms, the NAWL (Browne et al., 2013) and the AVL (Gardner & Davies, 2014) were created.

The NAWL was created by Browne et al. (2013) using a similar method to the AWL, though with a significantly larger and more modern corpus of 288 million words. The NAWL uses lemmas instead of word families and is meant to be used alongside the lemmatized NGSL to cover roughly 92% of words found in academic texts (Browne et al., 2013; Browne 2014). Gardner and Davies (2014) developed the Academic Vocabulary List (AVL) drawing on an academic sub-corpus of the Corpus of Contemporary American English (COCA). The AVL covers about 14% of the academic materials in COCA and the British National Corpus. Like the NAWL, it uses lemmas rather than word families to determine word frequencies and is ratio-based, meaning that words which appeared 1.5 times more in academic than non-academic texts were considered; this rendered a larger and more modern list than the AWL. However, unlike the NAWL, it is not meant to work in conjunction with the GSL or NGSL.

Academic word lists like the NAWL play a crucial role in setting general vocabulary goals for language courses, guiding learners in their independent study, and informing course and material

designers in selecting texts and developing learning activities. Yet one of the most challenging aspects of teaching vocabulary is making principled decisions about which words are worth focusing on (Coxhead, 2000), and academic word lists may not be sufficient to provide the kind of coverage needed for discipline-specific literacy skills. O'Flynn (2019) argues that such general lists "can mislead students by not providing a sufficiently nuanced account of the academic vocabulary of their discourse community" (p. 38). Simply put, teaching from the AWL, NAWL, or other general lists may require students to learn vocabulary that they may not need and not expose them to vocabulary that is crucial for understanding information in their field.

While exploring this notion, Martinez et al. (2009) found the AWL lacking for use in the field of agriculture and created a field-specific vocabulary list for agricultural students. The rationale for this list was the belief that the complexity of disciplinary communication necessitates focusing vocabulary instruction on the uses and conventions of specific disciplinary communities. The authors believed that such lists should include all frequent academic lexical items necessary for successfully reading and writing academic articles in a specific field. They created the AgroCorpus, a corpus of 826,416 words taken from 218 articles on agricultural science. While examining the overlap between the AWL and their corpus, the authors found that only 9.06% of words in the AgroCorpus were found on the AWL and 37.5% of the types in the AWL did not occur at all in the AgroCorpus. They then compared the AgroCorpus to similar word lists from other fields. Computer science had the highest rate of overlap – containing approximately 16% of words from the AWL – while biology had the lowest with its list containing 6.2% AWL words. These observations led Martinez et al. to claim that the AWL is too general to be functional for those who are teaching or studying English for a specific purpose. They instead proposed using field-specific corpora to create specialized high-frequency word lists to maximize learning efficiency.

In a more recent study, O'Flynn (2019) created the Economic Academic Word List (EAWL) using both a target corpus, an economics subcorpus of the Directory of Open Access Journals (DOAJ), and a validation corpus, the EcoCorpus, to pull relevant vocabulary from academic articles and studies on economics. Of the 887 EAWL words, only 354 overlap with the NAWL, meaning roughly 60% of words are unique to the EAWL. With a fairly uniform rate of coverage using the EAWL, representing between 5.4-5.9% of the words found in various major economics journals, O'Flynn explained that the EAWL "provides the means to design and select the most relevant teaching materials, establish vocabulary goals and target academic language instruction more specifically" (p. 38) while also guiding students in their independent studies. From this study, one can see yet more evidence of the need for and usefulness of specialized academic word lists. Furthermore, it is apparent that creating such lists in a methodologically-sound manner is possible.

Many register-specific corpora and word lists have been designed for specialized fields other than applied linguistics. For corpora, this includes the Hong Kong Financial Services Corpus, HKFSC (Research Centre for Professional Communication in English at the Department of English of the Hong Kong Polytechnic University, n.d.); and the aforementioned AgroCorpus (Martinez et al., 2009). Word lists for specialized fields outside of applied linguistics—using both single- and multi-word approaches—include the Academic Formulas List (AFL) (Simpson-Vlach & Ellis, 2010); Medical Academic Word List (MAWL) (Wang et al., 2008); and Economic Academic Word List (EAWL) (O'Flynn, 2009). These corpora and word lists fill the need in their fields for a list of register-specific, specialized vocabulary.

While there are some register-specific corpora and word lists designed for the field of applied linguistics, they are limited in their scope and generalizability. These include the Academic Word List for English Language Teaching (ELT) (Huong, 2018), the Applied Linguistics Research Article Corpus (ALC) (Vongpumivitch et al., 2009), and the Applied Linguistics AWL (Khani & Tazik, 2013). Huong (2018) created an academic word list for English Language Teaching using nine English language teaching textbooks which were converted to PDF and run through the

Range program to create a 702,828-word corpus. The first 2,000 words of the GSL were then removed from the list. Words from the remaining list were selected based on frequency and range (number of textbooks they appeared in). The minimum range to be included in Huong's (2018) English language teaching word list was 50%. The result was a list of 665 academic words specific to English language teaching. The ALC, on the other hand, was created from a corpus of 200 applied linguistics research articles from five journals with 1,237,574 words total for a final list of 603 words (Vongpumivitch et al., 2009). Lastly, the Applied Linguistics AWL was created in 2013 using 11,553,450 running words from 240 articles from twelve research journals published between 2000 and 2009. The resulting list has 773 types.

Despite their utility to language teachers and learners, these three lists are drawn from a small sample of applied linguistics texts. For instance, Huong (2018) drew the ELT from only nine English Language Teaching books with 702,828 running words total. Both the sizes of the ALC, created from 1.5 million words, and Applied Linguistics AWL, created from approximately 11.5 million words, though being drawn from an objectively larger and more robust corpora, are still relatively small given the volume of publications in the Applied Linguistics field. Furthermore, when creating the Applied Linguistics AWL, Khani and Tazik used articles of roughly the same length randomly selected from twelve journals to "reduce the possibility of word reoccurrences" (2013, p. 213). However, this results in a list that may not fully represent the most frequent words that applied linguists will encounter as they read journal articles in the field.

Our study aims to create an Applied Linguistics Academic Word List (ALAWL) corpus of academic vocabulary specifically for the specialized field of applied linguistics from a large, representative selection of professional applied linguistics journals using articles without regard to their size. By doing so, we seek to identify the most frequent academic and technical words that would be beneficial for students to learn in the specific domain of applied linguistics.

2. Methods

This section outlines the quantitative and qualitative methods used in the development of the ALAWL. These methods are divided into three stages: (2.1) designing and compiling the corpus and reference corpus, (2.2) quantitative analysis of the corpus, (2.3) qualitative refinement of the word list.

2.1. Designing and compiling the corpus and reference corpus

The ALAWL corpus is a collection of 48,244,153 tokens from across twelve open-access TESOL journals. We selected the twelve academic journals for our corpus based on several criteria. First, we chose those from a range of impact factors (>1 up to 5). Impact factors indicate how widely viewed and cited journals are within the Applied Linguistics community (see Table 1). Another requirement was to limit entries to those published between 2012 and 2021 to maintain a focus on current vocabulary in the field. We then consulted with an expert in applied linguistics and TESOL to determine the final list. We collected articles from each journal using Python and WebMage, which is a module within Python for facilitated data collection (https://pypi.org/project/webmage/). Each article was then converted into plain text and each text was tagged for part of speech and lemma using a tagging software called SpaCy (https://spacy.io/). Finally, the tagged corpus was indexed with another Python module called TextElixir (https://pypi.org/project/textelixir), which helped us to measure frequency distribution, dispersion, concordance lines, collocations, and n-grams.

Once the corpus was created, a reference corpus was later compiled to determine the coverage of the finalized word list. This reference corpus consists of texts from the same academic journals from 2021 to 2022, but the texts do not overlap with the main corpus. This corpus contains 2,920,002 tokens across 516 documents.

Table 1. Applied linguistics journals used in corpus.

Journal Name	Impact Factor (2021)
Applied Linguistics	5.741
Language Teaching	5.327
The Modern Language Journal	4.759
Language Learning	4.667
Studies in Second Language Acquisition	3.988
Language Teaching Research	3.899
TESOL Quarterly	3.692
Second Language Research	2.178
ELT Journal	2.028
Annual Review of Applied Linguistics	1.440
International Journal of Applied Linguistics	1.267
TESOL Journal	0.891

2.2. Quantitative analysis of the corpus

Our quantitative analysis of the corpus included two elements: (2.2.1) Specialized Occurrence and (2.2.2) Frequency and Dispersion Threshold.

2.2.1. Specialized occurrence.

The frequency for each lemma in our target corpus was calculated, and blacklists were used to remove common function words and non-academic or non-field-specific vocabulary respectively from our final list. The blacklists were meant to help us reach our goal of a final word list that, when combined with the common words of the NGSL, would account for 95% of our applied linguistics (AL) corpus.

For an occurrence of a word to be deemed specialized in the AL field, high-frequency words of English, as represented by Browne et al.'s NGSL and NAWL (2013), had to be excluded. The NGSL was used instead of Brezina and Gablasova's (2015) list as the former was created to work in tandem with the NAWL, while Brezina and Gablasova did not create an associated academic word list. The NGSL project was meant to create "a list of high-frequency words that was as useful as possible for students, teachers and researchers around the world" (Browne, 2014, p. 9). For this reason, the NGSL was chosen to represent the high-frequency words of English to be avoided in the creation of the ALAWL. There are two lists in the NGSL, namely the NGSL1 and the NGSL2. For the purposes of our word list, we combined both lists. The combined NGSL list comprises a total of 2,800 words, while the NAWL has 960 words. These lists were then used as blacklists, or closed sets of items to be excluded from the frequency list.

Another blacklist was needed to exclude all proper nouns, adjectives associated with proper nouns (e.g. nationalities), and ordinal numbers (e.g. first, second). It was assumed that prospective AL students would already know these words or that they could be learned from other sources. Our

goal in making this blacklist was that our final corpus would include high-frequency technical words specific to the AL field.

2.2.2. Frequency and dispersion threshold.

In this study, our aim was to produce a list that covered approximately the same percentage as NAWL. As such, we determined two thresholds for the word list: frequency and dispersion. Dispersion for this study is measured using Deviation of Proportions (DP) as proposed by Gries (2008). After testing the coverage of the wordlist with several thresholds onto the reference corpus, we determined that the coverage is closest to NAWL when using words with a frequency greater than 550 times and a DP of less than or equal to 0.95. This leaves the final count of the word list at 664 words.

2.3. Qualitative Refinement of the List

The qualitative refinement process involved examining and manually removing unnecessary words from the list. In addition to our blacklists which included the NGSL, NAWL, proper nouns, ordinal numbers, and proper adjectives (e.g., *American*), it was decided to remove all gerunds whose headwords appeared in the NGSL (e.g., *learning, teaching*). Our reasoning was that AL students and professionals will need to be able to use gerunds of these headwords in the field, but they are learnable from the NGSL with high transfer despite the difference in word form. We also removed a small number of abbreviations that were not specific to the AL domain, such as *e.g.*, *i.e.*, *vs.*, and *etc.* We kept nominalizations like "placement" and "alignment" because they are not gerunds and thus have a more tenuous link to the NGSL. We also combined several tokens with variants. For example, British spellings were merged with American spellings while hyphenated words were merged with non-hyphenated words.

3. Results

The resulting word list represents 664 of the most frequently used lemmas in the field of applied linguistics. The first 100 of these are shown in Table 2, with the full list available in Appendix A. The raw frequency of these lemmas in the journals pulled from is 1,765,279 with an average word frequency of 2,659. The potential usefulness of the ALAWL as a teaching resource can be ascertained by noting the word coverage achieved. The ALAWL covered 5.37% of words in our original corpus, and when combined with the NGSL and the NAWL, the word coverage was 93.19%. In order to validate this, we also measured the coverage of the combined NGSL, NAWL, and ALAWL on our reference corpus of nearly 3 million words collected from the same 12 journals used for our original corpus but drawn from the years 2021 and 2022 such that there is no overlap with the original corpus. Based on this analysis, the ALAWL accounted for 5.34% of words, and the three corpora together accounted for 93.49% of all words leaving just 6.51% off-list words. Based on this coverage, it is likely that readers familiar with these three word lists will be able to read and comprehend content in the field of applied linguistics well given the typical threshold researchers give of 95% to 98% vocabulary knowledge in order to reasonably comprehend a text (Hu & Nation, 2000; Laufer, 1989; Schmitt et al., 2011).

Of the 664 lemmas on the Applied Linguistic Academic Word List (ALAWL), the largest portion were nouns, which accounted for 39.8% of the total lemmas on the list. The second-largest word group was adjectives which comprised 35.2% of the ALAWL. Verbs accounted for 16.9%, while adverbs accounted for 8.0%. The list also contains a single conjunction, 'ALBEIT' (See Appendices A-B for complete part-of-speech word lists).

Table 2. The 100 most frequent words in applied linguistics. (*Note.* Abbreviations listed in the table are explicated in Appendix A.)

Word	Frequency	Word	Frequency	Word	Frequency
LEARNER	194384	COLLABORATIVE	6226	AFFECTIVE	3912
L2	152640	TOKEN	6041	REVIEWER	3860
L1	62128	SE	5993	NP	3793
PROFICIENCY	46648	INTERCULTURAL	5956	FAMILIARITY	3789
LITERACY	14835	BILINGUALISM	5873	NNS	3777
EFL	14043	IMMERSION	5853	INTONATION	3737
FLUENCY	13709	METHODOLOGICAL	5654	TEXTUAL	3723
PRONUNCIATION	13549	PHONETIC	5560	LINGUA	3715
ESL	12813	NON-NATIVE	5539	LEXICON	3703
MULTILINGUAL	12443	DESCRIPTIVE	5452	ALOUD	3673
PP	12331	CONTEXTUAL	5215	MODALITY	3640
PEDAGOGICAL	11874	AUTHENTIC	5101	REVISION	3635
INSTRUCTOR	10726	PROFICIENT	5065	INTERACTIVE	3635
SD	10663	METALINGUISTIC	5061	RESPONDENT	3546
INSTRUCTIONAL	10116	ADDITIONALLY	4992	PERCEPTUAL	3508
CF	9860	ANOVA	4977	RETENTION	3469
PEDAGOGY	9736	INSTRUCT	4970	INTERLAN-GUAGE	3429
GENRE	9642	INTERLOCUTOR	4849	FRANCA	3407
MONOLINGUAL	9549	CORRECTIVE	4785	PHONOLOGY	3379
FL	9486	ADMINISTER	4725	MORPHEME	3372
APTITUDE	8954	FOSTER	4630	ORTHOGRAPHIC	3365
TESOL	8505	MULTILINGUALISM	4542	TRANSCRIPT	3285
EXCERPT	8401	SOCIOCULTURAL	4481	PLACEMENT	3254
COLLOCATION	8257	RECEPTIVE	4465	ACTIVATION	3251
NS	8101	INCIDENTAL	4372	ANALYTIC	3226
HERITAGE	8032	FORMULAIC	4346	UNGRAMMATI-CAL	3208
COGNITION	7788	MOTIVATIONAL	4234	MULTIMODAL	3205
INTERACTIONAL	7282	EFFICACY	4220	SALIENT	3179
ELT	7137	COMPREHENSIBIL-ITY	4114	LINGUIST	3178
PREDICTOR	7106	COLLABORATION	4103	COGNATE	3152
ELICIT	6868	SA	4056	ONGOING	3140
RATER	6796	RELEVANCE	4024	REFLECTIVE	3094
ADJECTIVE	6770	AUDITORY	3915	COMPLETION	3067
PRONOUN	6592				

4. Discussion and Conclusions

The Applied Linguistic Academic Word List (ALAWL) created by this study gives a list of 664 of the most frequent academic or technical lemmas used in the field of applied linguistics. The ALAWL draws on articles from twelve applied linguistics journals, 48,244,153 tokens. Thus, this study has created an academic vocabulary corpus specifically for the use in applied linguistics from a wide range of professional applied linguistics articles and answers the question of what the most frequent academic and technical vocabulary words in the domain of applied linguistics are.

The relevancy of the ALAWL words to the field of applied linguistics is demonstrated when comparing it to Huong's (2018) ELT, with which the ALAWL shares some overlap. Huong's final list consists of 665 academic words specific to ELT, 38 of which also appear on the ALAWL (see Table 3). This overlap with the ELT demonstrates that the words selected for the ALAWL are pertinent to the field of applied linguistics.

Table 3. ALAWL words on ELT (Huong, 2018).

Word	Word number on ALAWL (Based on frequency)	Word	Word number on ALAWL (Based on frequency)
INSTRUCTOR	13	CONSIDERABLY	244
CONTEXTUAL	45	RELIANCE	293
RELEVANCE	66	COMPLEMENTARY	312
AFFECTIVE	68	COHERENCE	317
REVISION	79	ACKNOWLEDGEMENT	318
INTERACTIVE	80	ISOLATION	343
RETENTION	83	DISTINCTIVE	344
ONGOING	98	DISCRIMINATE	365
BENEFICIAL	106	APPRECIATION	410
CONSISTENTLY	130	CLARITY	412
ACCESSIBLE	138	ADEQUATELY	424
INSTITUTE	147	CONVERSELY	426
CULTURALLY	155	OPTIONAL	436
FORMULATE	188	PREDICTABLE	483
INNOVATIVE	208	HYPOTHETICAL	539
INHERENT	209	COOPERATIVE	540
CLARIFICATION	210	COMPILE	550
ADMINISTRATOR	225	ALTERNATIVELY	558
NEGATIVELY	226	INADEQUATE	563

Possible explanations for the small amount of overlap are that different sources and criteria were used. Huong used nine textbooks for ELT whereas we used twelve academic journals in the field of applied linguistics (for a list of journals used see Appendix B). Also, Huong's blacklist contained the first 2,000 words of the GSL while we removed all words found on the NGSL (2,800

words). These two differences are likely the reason for the small number of overlapping words between the two lists, and they highlight the constraints faced by using a relatively narrow corpus and differing methods.

The ALAWL has immediate applications for language teachers and learners. It can focus student efforts to study independently or in class within the field of applied linguistics. It can be especially useful for upper-level undergraduates and graduate students who enter the field with limited domain knowledge. The domain-specificity and qualitative refinement of the ALAWL address O'Flynn's (2019) argument that general lists do not give students "sufficiently nuanced" vocabulary to guide them (2019). The ALAWL can be used by students in conjunction with the NGSL to make academic writing in their field comprehensible. The use of the NGSL and NAWL as opposed to the GSL and AWL also means that this level of comprehensibility is based on current research in corpus linguistics. The ALAWL can also function like the NAWL to inform course and material designers. Materials may be designed or selected so as to reinforce the learning of these key vocabulary terms. In addition, the list can be helpful to curriculum developers in applied linguistics majors and graduate programs where advisors are eager to ensure their students master an essential set of concepts prior to graduation. As Martinez et al.'s AgroCorpus (2009) did for the field of agriculture, the ALAWL provides the focus for vocabulary instruction that can maximize learning efficiency in its field.

The limitations of this research are both quantitative and qualitative. The quantitative analysis of the AL research journals used twelve recent high-impact journals to address the need for a larger AL corpus that has broader applications than the ELT created by Huong (2018) and the ALC created by Vongpumivitch, et al. (2009). However, the text analyzed was English text. This suits the specific purpose of this study. Applied linguistics, especially TESOL, is a global field with many organizations or journals that publish in languages other than English. If the list created for this study is meant to be used to provide a list for students to learn as they study TESOL and applied linguistics, then it is possible that this list does not address the high-frequency words they may encounter in languages other than English. While some of these words may be synonymous translations of the word list created from English medium articles, it is possible that corresponding frequency lists would include items not on the English list or include the same items but at a different frequency depending on the focus of the foreign language medium articles. Therefore, the list created here is best used for English medium instruction and to prepare students and professionals to better understand applied linguistic academic writing published in English. The development of further corpora would be needed for applied linguistic journals published in languages other than English which may better address the specific vocabulary needs of students and professionals engaging with the applied linguistics community through a different language. Finally, further research is also needed to verify the effectiveness of using the ALAWL in teaching applied linguistics.

ORCID iDs

Ana K. Barraza https://orcid.org/0009-0004-6171-7168

Grant Eckstein https://orcid.org/0000-0002-3667-4571

Zachary M. Lambert https://orcid.org/0009-0009-3310-060X

Carolee Rogers https://orcid.org/0000-0002-6945-5578

Appendix A: Applied Linguistics Academic Word List (ALAWL)

Rank	Word	Frequency	Dispersion	Part of Speech
1	LEARNER	194384	0.4221	NOUN
2	L2 (Language Two)	152640	0.513	NOUN
3	L1 (Language One)	62128	0.632	NOUN
4	PROFICIENCY	46648	0.5482	NOUN
5	LITERACY	14835	0.7749	NOUN
6	EFL (English as a Foreign Language)	14043	0.7393	NOUN
7	FLUENCY	13709	0.7914	NOUN
8	PRONUNCIATION	13549	0.8054	NOUN
9	ESL (English as a Second Language)	12813	0.7496	NOUN
10	MULTILINGUAL	12443	0.7655	ADJ
11	PP (Prepositional Phrase)	12331	0.7072	NOUN
12	PEDAGOGICAL	11874	0.6261	ADJ
13	INSTRUCTOR	10726	0.7811	NOUN
14	SD (Standard Deviation)	10663	0.736	NOUN
15	INSTRUCTIONAL	10116	0.691	ADJ
16	CF (Corrective Feedback)	9860	0.7996	NOUN
17	PEDAGOGY	9736	0.7187	NOUN
18	GENRE	9642	0.8271	NOUN
19	MONOLINGUAL	9549	0.7905	ADJ
20	FL (Foreign Language)	9486	0.9114	NOUN
21	APTITUDE	8954	0.901	NOUN
22	TESOL (Teaching English to Speakers of Other Languages)	8505	0.7972	NOUN
23	EXCERPT	8401	0.8488	NOUN
24	COLLOCATION	8257	0.9067	NOUN
25	NS (Native Speaker)	8101	0.8883	NOUN
26	HERITAGE	8032	0.8428	NOUN
27	COGNITION	7788	0.739	NOUN
28	INTERACTIONAL	7282	0.8123	ADJ
29	ELT (English Language Teaching)	7137	0.8886	NOUN
30	PREDICTOR	7106	0.7882	NOUN
31	ELICIT	6868	0.6564	VERB
32	RATER	6796	0.8475	NOUN
33	ADJECTIVE	6770	0.8279	NOUN
34	PRONOUN	6592	0.8291	NOUN
35	COLLABORATIVE	6226	0.8088	ADJ
36	TOKEN	6041	0.8023	NOUN

37	SE (Standard Error)	5993	0.7835	NOUN
38	INTERCULTURAL	5956	0.8611	ADJ
39	BILINGUALISM	5873	0.8013	NOUN
40	IMMERSION	5853	0.8116	NOUN
41	METHODOLOGICAL	5654	0.6954	ADJ
42	PHONETIC	5560	0.8454	ADJ
43	NON-NATIVE	5539	0.7942	ADJ
44	DESCRIPTIVE	5452	0.6613	ADJ
45	CONTEXTUAL	5215	0.706	ADJ
46	AUTHENTIC	5101	0.7528	ADJ
47	PROFICIENT	5065	0.717	ADJ
48	METALINGUISTIC	5061	0.8181	ADJ
49	ADDITIONALLY	4992	0.6469	ADJ
50	ANOVA	4977	0.7864	NOUN
51	INSTRUCT	4970	0.6656	VERB
52	INTERLOCUTOR	4849	0.7906	NOUN
53	CORRECTIVE	4785	0.8781	ADJ
54	ADMINISTER	4725	0.6735	VERB
55	FOSTER	4630	0.7233	VERB
56	MULTILINGUALISM	4542	0.8718	NOUN
57	SOCIOCULTURAL	4481	0.7729	ADJ
58	RECEPTIVE	4465	0.8261	ADJ
59	INCIDENTAL	4372	0.8706	ADJ
60	FORMULAIC	4346	0.8963	ADJ
61	MOTIVATIONAL	4234	0.9068	ADJ
62	EFFICACY	4220	0.8636	NOUN
63	COMPREHENSIBILITY	4114	0.9308	NOUN
64	COLLABORATION	4103	0.8178	NOUN
65	SA (Study Abroad)	4056	0.9427	NOUN
66	RELEVANCE	4024	0.6971	NOUN
67	AUDITORY	3915	0.8698	ADJ
68	AFFECTIVE	3912	0.8157	ADJ
69	REVIEWER	3860	0.6673	NOUN
70	NP (Noun Phrase)	3793	0.9361	NOUN
71	FAMILIARITY	3789	0.7858	NOUN
72	NNS (Nonnative Speaker)	3777	0.9309	NOUN
73	INTONATION	3737	0.852	NOUN
74	TEXTUAL	3723	0.8383	ADJ
75	LINGUA	3715	0.8708	NOUN
76	LEXICON	3703	0.7982	NOUN
77	ALOUD	3673	0.8155	ADJ
78	MODALITY	3640	0.8236	NOUN

79	REVISION	3635	0.8745	NOUN
80	INTERACTIVE	3635	0.7542	ADJ
81	RESPONDENT	3546	0.8829	NOUN
82	PERCEPTUAL	3508	0.8527	ADJ
83	RETENTION	3469	0.8706	NOUN
84	INTERLANGUAGE	3429	0.8087	ADJ
85	FRANCA	3407	0.8916	ADJ
86	PHONOLOGY	3379	0.8234	NOUN
87	MORPHEME	3372	0.8597	NOUN
88	ORTHOGRAPHIC	3365	0.8966	ADJ
89	TRANSCRIPT	3285	0.8149	NOUN
90	PLACEMENT	3254	0.8097	NOUN
91	ACTIVATION	3251	0.8711	NOUN
92	ANALYTIC	3226	0.8033	ADJ
93	UNGRAMMATICAL	3208	0.9067	ADJ
94	MULTIMODAL	3205	0.9089	ADJ
95	SALIENT	3179	0.7211	ADJ
96	LINGUIST	3178	0.8173	NOUN
97	COGNATE	3152	0.9259	NOUN
98	ONGOING	3140	0.7379	ADJ
99	REFLECTIVE	3094	0.8496	ADJ
100	COMPLETION	3067	0.7693	NOUN
101	REFERENT	3053	0.9061	NOUN
102	IDIOM	3050	0.9274	NOUN
103	L3 (Language Three)	3047	0.9473	NOUN
104	MAINSTREAM	3021	0.8437	ADJ
105	TAKER	3018	0.922	NOUN
106	BENEFICIAL	3008	0.7269	ADJ
107	ALIGN	3001	0.7532	VERB
108	DISCIPLINARY	2987	0.8874	ADJ
109	GRAMMATICALITY	2982	0.8734	NOUN
110	CONCEPTUALIZATION	2903	0.7907	NOUN
111	BEGINNER	2891	0.8273	NOUN
112	DECLARATIVE	2883	0.8753	ADJ
113	DISCURSIVE	2848	0.8659	ADJ
114	HYPOTHESIZE	2844	0.7435	VERB
115	PROCEDURAL	2839	0.8728	ADJ
116	ATTRITION	2839	0.923	NOUN
117	PROSODIC	2830	0.9095	ADJ
118	PREPOSITION	2830	0.8788	NOUN
119	ATTAINMENT	2825	0.8589	NOUN
120	METACOGNITIVE	2801	0.9155	ADJ

121	CONVERSATIONAL	2778	0.8114	ADJ
122	CONVEY	2736	0.7547	VERB
123	ENTAIL	2733	0.7203	VERB
124	PHONEME	2725	0.8931	NOUN
125	EAP (English for Academic Purposes)	2724	0.9409	NOUN
126	ALIGNMENT	2699	0.9028	NOUN
127	OPERATIONALIZE	2662	0.7675	VERB
128	BASELINE	2647	0.8537	NOUN
129	OVERT	2632	0.8464	ADJ
130	CONSISTENTLY	2613	0.7028	ADV
131	MEDIATION	2608	0.9065	NOUN
132	UPTAKE	2583	0.9013	VERB
133	EMERGENT	2571	0.8627	ADJ
143	ENHANCEMENT	2556	0.8886	NOUN
135	GLOSS	2547	0.9254	NOUN
136	ROBUST	2515	0.7538	ADJ
137	MORPHOSYNTACTIC	2506	0.8568	ADJ
138	ACCESSIBLE	2493	0.759	ADJ
139	DUAL	2492	0.8422	ADJ
140	NONWORD	2473	0.9416	NOUN
141	PHRASAL	2459	0.9127	ADJ
142	FLUENT	2442	0.8053	ADJ
143	LINGUISTICALLY	2432	0.7798	ADV
144	CONCEPTUALIZE	2429	0.7769	VERB
145	INTELLIGIBILITY	2416	0.9355	NOUN
146	CURRICULAR	2397	0.8771	ADJ
147	INSTITUTE	2393	0.8204	NOUN
148	ACOUSTIC	2391	0.9209	ADJ
149	CONTINUUM	2363	0.8216	NOUN
150	VIOLATION	2361	0.9069	NOUN
151	FILLER	2358	0.8968	NOUN
152	EXPLORATION	2352	0.7732	NOUN
153	CATEGORIZE	2343	0.766	VERB
154	HOLISTIC	2335	0.8504	ADJ
155	CULTURALLY	2324	0.8248	ADV
156	DETERMINER	2313	0.9068	NOUN
157	FIXATION	2303	0.9397	NOUN
158	PEDAGOGIC	2287	0.8853	ADJ
159	ADVERB	2286	0.8873	NOUN
160	SOCIALIZATION	2263	0.9038	NOUN
161	OUTPERFORM	2260	0.8372	VERB

162	HANDBOOK	2243	0.8305	NOUN
163	CATEGORIZATION	2242	0.8617	NOUN
164	SEMANTICALLY	2239	0.8523	ADV
165	COMPREHEND	2238	0.8132	VERB
166	NATIVELIKE	2234	0.9079	ADJ
167	IRREGULAR	2219	0.9319	ADJ
168	SEMIOTIC	2207	0.9062	ADJ
169	INTERPERSONAL	2196	0.858	ADJ
170	SUPPLEMENTARY	2194	0.8618	ADJ
171	DOCTORAL	2176	0.8372	ADJ
172	RATIONALE	2153	0.7897	NOUN
173	NONVERBAL	2139	0.9054	ADJ
174	SEQUENTIAL	2112	0.8737	ADJ
175	PREDICTIVE	2073	0.8755	ADJ
176	ANONYMOUS	2073	0.7244	ADJ
177	WORKPLACE	2069	0.9176	NOUN
178	ANALYTICAL	2064	0.8208	ADJ
179	CONJUNCTION	2044	0.8463	NOUN
180	TOEFL (Test of English as a Foreign Language)	2043	0.9076	NOUN
181	SUFFIX	2041	0.9278	NOUN
182	NATURALISTIC	2037	0.8395	ADJ
183	COUNTERPART	2031	0.8027	NOUN
184	GENERIC	2027	0.8992	ADJ
185	SCAFFOLDING	2026	0.8985	NOUN
186	ZERO	2015	0.8421	VERB
187	RHETORICAL	2014	0.9044	ADJ
188	FORMULATE	1988	0.7818	VERB
189	UNRELATED	1988	0.8748	ADJ
190	ALPHA	1962	0.8336	NOUN
191	THEMATIC	1960	0.8689	ADJ
192	VARIED	1938	0.7988	ADJ
193	INTERFERENCE	1932	0.8748	NOUN
194	ATTENTIONAL	1912	0.8831	ADJ
195	AFFORDANCE	1909	0.9055	NOUN
196	GUIDANCE	1908	0.8357	NOUN
197	PROSODY	1908	0.9261	NOUN
198	CONSTITUENT	1906	0.8785	NOUN
199	DISTRACTOR	1900	0.9029	NOUN
200	CONSISTENCY	1896	0.8009	NOUN
201	EXPLORATORY	1887	0.8447	ADJ
202	EXEMPLAR	1884	0.892	NOUN

203	SHED	1877	0.7613	VERB
204	INFER	1866	0.8288	VERB
205	CLASSMATE	1865	0.8782	NOUN
206	MASTERY	1858	0.8449	NOUN
207	ARTICULATION	1849	0.8727	NOUN
208	INNOVATIVE	1848	0.8632	ADJ
209	INHERENT	1848	0.786	ADJ
210	CLARIFICATION	1844	0.866	NOUN
211	SYSTEMATICALLY	1837	0.7833	ADV
212	MODAL	1832	0.8961	ADJ
213	CITATION	1831	0.9468	NOUN
214	ETHNOGRAPHIC	1806	0.8926	ADJ
215	IMITATION	1805	0.9274	NOUN
216	AMBIGUOUS	1788	0.8695	ADJ
217	SALIENCE	1783	0.8834	ADJ
218	ELICITATION	1779	0.8759	NOUN
219	MISMATCH	1769	0.8906	VERB
220	ACCEPTANCE	1766	0.8751	ADJ
221	RESPECTIVE	1764	0.7992	ADJ
222	CONTRASTIVE	1763	0.8949	ADJ
223	IDEOLOGICAL	1760	0.8993	ADJ
224	ACCORDINGLY	1741	0.8029	ADV
225	ADMINISTRATOR	1726	0.8943	NOUN
226	NEGATIVELY	1721	0.8177	ADV
227	MANIFEST	1720	0.811	VERB
228	INFLECTION	1715	0.9106	NOUN
229	MIGRANT	1708	0.9316	NOUN
230	R2 (R-Squared)	1698	0.931	NOUN
231	DA (Dynamic Assessment)	1696	0.9493	NOUN
232	EXEMPLIFY	1694	0.8166	VERB
233	SUBORDINATE	1693	0.8975	VERB
234	REALIZATION	1684	0.8564	NOUN
235	AUTOMATIC	1679	0.8431	ADJ
236	AMBIGUITY	1678	0.8708	NOUN
237	COHORT	1670	0.9144	NOUN
238	REGULARITY	1667	0.8942	NOUN
239	REFORMULATION	1666	0.9271	NOUN
240	INTERCEPT	1663	0.9246	VERB
241	AURAL	1649	0.9177	ADJ
242	ORTHOGRAPHY	1647	0.9235	NOUN
243	JUDGEMENT	1643	0.9133	NOUN
244	CONSIDERABLY	1636	0.7807	ADV

245	EMBODY	1624	0.8919	VERB
246	AUTONOMOUS	1622	0.8895	ADJ
247	CAUTION	1620	0.8074	VERB
248	VALIDATION	1615	0.8857	NOUN
249	REFERENTIAL	1612	0.9179	ADJ
250	VALIDATE	1607	0.8311	VERB
251	CONTEXTUALIZE	1603	0.8493	VERB
252	RESIDENCE	1596	0.8989	NOUN
253	POSIT	1593	0.8094	VERB
254	SCAFFOLD	1584	0.9014	VERB
255	EMPIRICALLY	1581	0.8117	ADV
256	FOCAL	1577	0.8969	ADJ
257	UNDERLINE	1574	0.8474	VERB
258	MACRO	1560	0.8968	ADJ
259	ELABORATION	1558	0.887	NOUN
260	WILLINGNESS	1556	0.8619	NOUN
261	STAKEHOLDER	1554	0.9175	NOUN
262	IMPERATIVE	1554	0.894	ADJ
263	INDEFINITE	1552	0.9411	ADJ
264	ATTAIN	1547	0.8342	VERB
265	ENACT	1545	0.8943	VERB
266	ACCEPTABILITY	1544	0.9311	NOUN
267	INCOMPLETE	1535	0.8691	ADJ
268	DISAGREEMENT	1533	0.8708	NOUN
268	DEPENDENCY	1533	0.9277	NOUN
270	APPROPRIATENESS	1525	0.88	NOUN
271	VIRTUAL	1525	0.9318	ADJ
272	MI (Multiple Intelligences)	1502	0.942	NOUN
273	ADVERBIAL	1496	0.9295	ADJ
274	COMPREHENSIBLE	1485	0.8803	ADJ
275	DISCREPANCY	1483	0.8309	NOUN
276	COMPETENCY	1468	0.8909	NOUN
277	USEFULNESS	1460	0.8615	NOUN
278	PERTAIN	1459	0.8435	VERB
279	INDICATIVE	1453	0.8651	ADJ
280	KEYWORD	1451	0.9431	NOUN
281	INCLUSIVE	1445	0.9129	ADJ
282	DEMOGRAPHIC	1445	0.8681	ADJ
283	AUXILIARY	1438	0.9215	ADJ
284	INTRODUCTORY	1437	0.8939	ADJ
285	AUTHENTICITY	1436	0.9245	NOUN
286	RECIPIENT	1435	0.9269	NOUN

287	GENERATIVE	1434	0.9234	ADJ
288	INFORMATIVE	1433	0.8553	ADJ
289	SCHEMA	1433	0.9139	NOUN
290	CAUSAL	1428	0.8879	ADJ
291	SOLELY	1414	0.8039	ADV
292	SUPPORTIVE	1412	0.8728	ADJ
293	RELIANCE	1411	0.8579	NOUN
294	ENCOMPASS	1396	0.8278	VERB
295	SCHOLARLY	1392	0.9108	ADV
296	HIERARCHICAL	1385	0.8766	ADJ
297	INTENTIONAL	1384	0.8943	ADJ
298	INTERVIEWEE	1384	0.946	NOUN
299	MILLISECOND	1383	0.9342	NOUN
300	DEFAULT	1381	0.884	ADJ
301	SPECIFICITY	1379	0.9059	NOUN
302	TRAINER	1370	0.9488	NOUN
303	INFLECTIONAL	1369	0.9286	ADJ
304	CLOZE	1366	0.9252	ADJ
305	TRANSPARENT	1361	0.8869	ADJ
306	THEORETICALLY	1358	0.8377	ADV
307	INTRINSIC	1354	0.9062	ADJ
308	ABSENT	1351	0.843	ADJ
309	ACCOMMODATE	1350	0.8551	VERB
310	SIMULTANEOUS	1346	0.8666	ADJ
311	DEEM	1343	0.82	VERB
312	COMPLEMENTARY	1343	0.8673	ADJ
313	ARGUMENTATIVE	1330	0.9451	ADJ
314	DATASET	1329	0.9024	NOUN
315	CONTRIBUTOR	1329	0.9207	NOUN
316	CATEGORICAL	1322	0.8785	ADJ
317	COHERENCE	1321	0.8984	ADJ
318	ACKNOWLEDGEMENT	1316	0.8103	NOUN
319	IMMIGRATION	1316	0.9255	NOUN
320	SEGMENTATION	1315	0.9427	NOUN
321	NOTABLE	1301	0.8327	ADJ
322	EVALUATIVE	1298	0.9218	ADJ
323	SITUATIONAL	1297	0.8991	ADJ
324	COMPUTATIONAL	1294	0.9105	ADJ
325	ACKNOWLEDGMENT	1290	0.8234	NOUN
326	INSUFFICIENT	1286	0.8381	ADJ
327	DIALOGIC	1283	0.915	ADJ
328	COGNITIVELY	1282	0.8671	ADV

329	CROSSLINGUISTIC	1280	0.9247	ADJ
330	ENDEAVOR	1276	0.8633	VERB
331	OBLIGATORY	1275	0.8972	ADJ
332	EXTENSIVELY	1272	0.8306	ADV
333	DEPLOY	1269	0.8872	VERB
334	IMPLICITLY	1268	0.8502	ADV
335	MULTICULTURAL	1268	0.9078	ADJ
336	INSTRUMENTAL	1268	0.8975	ADJ
337	LEXIS	1267	0.901	NOUN
338	TERTIARY	1249	0.9117	ADJ
339	NOTABLY	1247	0.8339	ADV
340	COHESION	1245	0.9173	NOUN
341	ALBEIT	1239	0.8211	PREP
342	INVENTORY	1234	0.9026	NOUN
343	ISOLATION	1233	0.8596	NOUN
344	DISTINCTIVE	1233	0.8803	ADJ
345	INTUITION	1233	0.9069	NOUN
346	UNCOVER	1230	0.8586	VERB
347	OMIT	1230	0.8678	VERB
348	ATTEST	1222	0.8652	VERB
349	PREDICATE	1222	0.9284	VERB
350	ID (Individual Differences)	1221	0.9467	NOUN
351	RUBRIC	1220	0.9289	NOUN
352	LIKERT	1220	0.8869	NOUN
353	ASSOCIATIVE	1214	0.9304	ADJ
354	DESIRABLE	1213	0.8595	ADJ
355	HYBRID	1208	0.9163	NOUN
356	LATENT	1205	0.9327	ADJ
357	BREADTH	1197	0.9213	NOUN
358	UNDERSCORE	1196	0.8656	VERB
359	GRAPHIC	1196	0.9243	ADJ
360	LEMMA	1192	0.949	NOUN
361	CUMULATIVE	1190	0.9098	ADJ
362	WARRANT	1176	0.8374	VERB
363	OMISSION	1175	0.9142	NOUN
364	INDUCTIVE	1174	0.9169	ADJ
365	DISCRIMINATE	1173	0.9091	VERB
366	KINDERGARTEN	1173	0.9319	NOUN
367	LITERAL	1173	0.9343	ADJ
368	CONTINGENCY	1167	0.9277	NOUN
369	NEWCOMER	1166	0.9499	NOUN
370	DELETION	1161	0.9423	NOUN

371	VERIFY	1160	0.8571	VERB
372	SPITE	1151	0.8814	VERB
373	PHONEMIC	1149	0.9279	ADJ
374	BONFERRONI	1148	0.9078	NOUN
375	REPEATEDLY	1144	0.8444	ADV
376	SOCIETAL	1143	0.9131	ADJ
377	LENS	1139	0.8968	NOUN
378	SUBSTANTIALLY	1138	0.8526	ADV
379	BATTERY	1132	0.9219	NOUN
380	INTERPRETIVE	1119	0.9181	ADJ
381	FORMATIVE	1117	0.9422	ADJ
382	LAG	1117	0.931	VERB
383	SYNONYM	1116	0.9199	NOUN
384	FOREGROUND	1114	0.9002	NOUN
385	APPLICABLE	1113	0.8645	ADJ
386	ETHNICITY	1109	0.9229	NOUN
387	MERIT	1107	0.8686	NOUN
388	PAIRWISE	1102	0.9144	ADJ
389	MORPHOSYNTAX	1097	0.9117	NOUN
390	COLLABORATE	1093	0.9107	VERB
391	INTERDISCIPLINARY	1092	0.9083	ADJ
392	CORROBORATE	1087	0.8507	VERB
393	CONFIRMATION	1087	0.9044	NOUN
394	RICHNESS	1086	0.9214	NOUN
395	POLITENESS	1085	0.9464	NOUN
396	CANONICAL	1079	0.9418	ADJ
397	TARGETLIKE	1072	0.9489	ADJ
398	GRADUAL	1070	0.8781	ADJ
399	MEDIAN	1070	0.9299	NOUN
400	CRUCIALLY	1061	0.8673	ADV
401	ANGLOPHONE	1059	0.9453	NOUN
402	QUERY	1056	0.8542	NOUN
403	EXPLANATORY	1052	0.8879	ADJ
404	ENJOYMENT	1050	0.9463	NOUN
405	COMPATIBLE	1046	0.879	ADJ
406	VIEWPOINT	1044	0.9052	NOUN
407	MISUNDERSTANDING	1041	0.9162	NOUN
408	INFINITIVE	1037	0.9446	ADJ
409	GENERALIZABILITY	1034	0.8889	NOUN
410	APPRECIATION	1034	0.9064	NOUN
411	WORLDWIDE	1034	0.9131	ADJ
412	CLARITY	1029	0.8918	NOUN

413	MODULATE	1026	0.922	VERB
414	LOGISTIC	1024	0.9282	ADJ
415	AVOIDANCE	1019	0.9262	NOUN
416	HINDER	1011	0.8734	VERB
417	NORMATIVE	1009	0.9253	ADJ
418	TECHNOLOGICAL	1009	0.9296	ADJ
419	INVOKE	1006	0.8995	VERB
420	EXCLUSION	1003	0.8984	NOUN
421	HOMOGENEOUS	1003	0.8761	ADJ
422	VISUALLY	1002	0.8875	ADV
423	EXPERIENTIAL	1000	0.9207	ADJ
424	ADEQUATELY	999	0.8716	ADV
425	ASYMMETRY	993	0.9369	NOUN
426	CONVERSELY	992	0.8564	ADV
427	MUTUALLY	992	0.8817	ADV
428	PREPOSITIONAL	992	0.9304	ADJ
429	INTERPLAY	991	0.8815	NOUN
430	OUTLIER	991	0.9193	NOUN
431	RETROSPECTIVE	990	0.9175	ADJ
432	ETHNOGRAPHY	990	0.9402	NOUN
433	TAXONOMY	988	0.9283	NOUN
434	PRIORITIZE	986	0.8938	VERB
435	SEEMINGLY	983	0.8704	ADV
436	OPTIONAL	983	0.9109	ADJ
437	DISTRIBUTIONAL	982	0.9323	ADJ
438	EXCLUSIVE	982	0.8839	ADJ
439	SYSTEMIC	982	0.9263	ADJ
440	DEVELOPER	981	0.9328	NOUN
441	SPECULATE	975	0.8579	VERB
442	THOROUGH	974	0.9006	ADJ
443	ENRICH	971	0.8912	VERB
444	LEXICALLY	971	0.9087	ADV
445	IDIOMATIC	970	0.934	ADJ
446	INTACT	966	0.8916	ADJ
447	RIGOROUS	966	0.9018	ADJ
448	UNFOLD	966	0.8969	VERB
449	ACCOMPLISHMENT	965	0.9262	NOUN
450	GRAMMATICALLY	959	0.8925	ADV
451	CONTINGENT	959	0.8974	NOUN
452	LITERATE	956	0.9326	ADJ
453	PROMINENCE	955	0.917	NOUN
454	INFORMANT	954	0.9479	NOUN

455	REHEARSAL	952	0.9399	NOUN
456	BENCHMARK	951	0.9164	NOUN
457	COVARIATE	950	0.9373	VERB
458	REFINE	949	0.8766	VERB
459	CONFRONT	948	0.8913	VERB
460	FACILITATIVE	946	0.9058	ADJ
461	INTERNALIZE	945	0.9023	VERB
462	ILLUMINATE	945	0.8919	VERB
463	CONTEND	943	0.8957	VERB
464	ATTRIBUTION	943	0.9492	NOUN
465	INTAKE	941	0.9414	VERB
466	TYPOLOGICAL	940	0.9418	ADJ
467	NOTEWORTHY	939	0.8696	ADJ
468	HESITATION	939	0.9241	NOUN
469	DICHOTOMY	937	0.9093	NOUN
470	SOCIOECONOMIC	936	0.9137	ADJ
471	REPRESENTATIONAL	935	0.9379	ADJ
472	RELATIONAL	934	0.9313	ADJ
473	MORPHOLOGICALLY	934	0.9459	ADV
474	EXERT	930	0.8769	VERB
475	ACCORDANCE	930	0.8688	NOUN
476	CONCURRENT	926	0.924	ADJ
477	ASPIRATION	919	0.9309	NOUN
478	SUBGROUP	915	0.9301	NOUN
479	OBSERVABLE	915	0.8925	ADJ
480	QUALITATIVELY	913	0.8813	ADV
481	RELIABLY	910	0.8844	ADV
482	INSIGHTFUL	909	0.8843	ADJ
483	PREDICTABLE	906	0.8943	ADJ
484	PROTOTYPICAL	906	0.9294	ADJ
485	CONFORM	905	0.8874	VERB
486	BLEND	900	0.9295	VERB
487	ACCESSIBILITY	900	0.9366	NOUN
488	COLLABORATIVELY	899	0.9198	ADV
489	DELETE	899	0.9207	VERB
490	PREVALENT	896	0.8844	ADJ
491	DELIBERATE	894	0.9097	VERB
492	UNDERPIN	892	0.9031	VERB
493	SUBJECTIVITY	892	0.9369	NOUN
494	GRADER	890	0.9435	NOUN
495	SYNTHESIZE	889	0.9137	VERB
496	IMPAIRMENT	888	0.9437	NOUN

497	RAPPORT	885	0.9449	NOUN
498	ERRONEOUS	883	0.9238	ADJ
499	PEDAGOGICALLY	879	0.9059	ADV
500	RESPONSIVE	878	0.9335	ADJ
501	ALTERNATION	877	0.9437	NOUN
502	INTEGRATIVE	876	0.9292	ADJ
503	AUTHORSHIP	867	0.9279	NOUN
504	FACILITATION	866	0.9454	NOUN
505	TRANSFORMATIVE	862	0.9407	ADJ
506	ORGANIZATIONAL	862	0.927	ADJ
507	DISPOSITION	861	0.9321	NOUN
508	DESIGNATE	860	0.8931	VERB
509	INHERENTLY	857	0.8853	ADV
510	NARRATION	857	0.9421	NOUN
511	AVENUE	857	0.9019	NOUN
512	INFREQUENT	856	0.9136	ADJ
513	ADOPTION	853	0.9163	NPUN
514	MITIGATE	853	0.912	VERB
515	ALIKE	851	0.8954	ADJ
516	CONSCIOUSLY	850	0.8889	ADV
517	NUANCED	847	0.8914	ADJ
518	EVOKE	847	0.9211	VERB
519	PSEUDONYM	847	0.9022	NOUN
520	COURSEWORK	846	0.9485	NOUN
521	EPISTEMOLOGICAL	846	0.9465	ADJ
522	THEORIZE	845	0.9162	VERB
523	IRRESPECTIVE	844	0.8912	ADJ
524	ROTE	839	0.938	NOUN
525	BROADEN	836	0.8993	VERB
526	MONOGRAPH	836	0.9497	NOUN
527	EMPOWER	833	0.9239	VERB
528	ASCERTAIN	832	0.8867	VERB
529	OPERATIONALIZATION	830	0.9034	NOUN
530	OVERLOOK	828	0.8905	VERB
531	SHORTCOMING	828	0.9059	NOUN
532	REFLEXIVE	828	0.9339	ADJ
533	INFERENTIAL	826	0.9228	ADJ
534	QUANTIFY	824	0.9103	VERB
535	UTTER	824	0.926	VERB
536	STRIVE	822	0.9016	VERB
537	ASSERTION	819	0.9054	NOUN
538	MULTIVARIATE	817	0.9243	ADJ

539	HYPOTHETICAL	814	0.9282	ADJ
540	COOPERATIVE	813	0.935	ADJ
541	EQUIP	812	0.9008	VERB
542	NOTICEABLE	812	0.8822	ADJ
543	IDIOSYNCRATIC	812	0.9257	ADJ
544	NAVIGATE	807	0.9247	VERB
545	COMPULSORY	807	0.9282	ADJ
546	INCREMENTAL	804	0.9362	ADJ
547	HUMANITY	802	0.9344	NOUN
548	BOLD	799	0.9082	ADJ
549	AFOREMENTIONED	797	0.9038	ADJ
550	COMPILE	795	0.9125	VERB
551	CONSTRUCTIVE	794	0.903	ADJ
552	PERSIST	794	0.8936	VERB
553	CODER	792	0.9406	NOUN
554	OBSTACLE	790	0.913	NOUN
555	ORIGINATE	787	0.8902	VERB
556	NONSIGNIFICANT	784	0.9085	ADJ
557	RESIDE	778	0.9003	VERB
558	ALTERNATIVELY	774	0.8729	ADV
559	CORRECTNESS	774	0.9273	NOUN
560	INTENT	774	0.9154	NOUN
561	NATIONALITY	772	0.9383	NOUN
562	PUBLICLY	771	0.9025	ADV
563	INADEQUATE	767	0.9021	ADJ
564	UNCONSCIOUS	766	0.934	ADJ
565	RECURRENT	763	0.9228	ADJ
566	NON-LINGUISTIC	759	0.9305	ADJ
567	INDIRECTLY	758	0.8924	ADV
568	ATTITUDINAL	757	0.9493	ADJ
569	IMPLICATE	756	0.9069	VERB
570	IDEALLY	752	0.8884	ADV
571	TYPOLOGY	752	0.9346	NOUN
572	PARSING	751	0.9474	NOUN
573	CONCEPTUALLY	750	0.911	ADV
574	JOINTLY	750	0.9076	ADV
575	AFFILIATION	749	0.9334	NOUN
576	DIMINISH	748	0.8952	VERB
577	MULTIDIMENSIONAL	745	0.9201	ADJ
578	CONSTRUE	745	0.9233	VERB
579	DEVISE	745	0.9078	VERB
580	RECUR	745	0.9181	VERB

581	ORALLY	743	0.9011	ADV
582	CEILING	741	0.9086	NOUN
583	TEMPLATE	740	0.946	NOUN
584	FRUSTRATION	739	0.9234	NOUN
585	AMPLE	738	0.897	ADJ
586	GEOGRAPHICAL	735	0.9284	ADJ
587	RESTRUCTURING	729	0.937	NOUN
588	SYNTACTICALLY	727	0.9177	ADV
589	PERTINENT	726	0.9069	ADJ
590	PHONOLOGICALLY	720	0.9279	ADV
591	MEMORIZATION	720	0.9369	NOUN
592	INTERSECTION	716	0.934	NOUN
593	NARRATE	716	0.9344	VERB
594	ARGUMENTATION	715	0.9493	NOUN
595	METRIC	709	0.9493	NOUN
596	ENTHUSIASM	707	0.9362	NOUN
597	ENROLLMENT	705	0.9392	NOUN
598	CERTAINTY	704	0.9388	NOUN
599	EQUATE	702	0.8972	VERB
600	PROSPECTIVE	696	0.9417	ADJ
601	NORMALITY	693	0.9258	NOUN
602	FEASIBLE	689	0.9092	ADJ
603	OVERARCHING	689	0.914	ADJ
604	MANIFESTATION	688	0.9132	NOUN
605	ADMINISTRATIVE	687	0.9382	ADJ
606	AUTOMATIZE	684	0.9409	VERB
607	MULTIFACETED	682	0.9122	ADJ
608	READINESS	681	0.9294	NOUN
609	COUNTERBALANCE	680	0.9037	VERB
610	INTUITIVE	680	0.9253	ADJ
611	SOLIDARITY	675	0.9465	NOUN
612	VIOLATE	674	0.9133	VERB
613	INSPECTION	673	0.9098	NOUN
614	INABILITY	673	0.9077	NOUN
615	NORMALIZE	671	0.9417	VERB
616	OUTSET	667	0.912	NOUN
617	INTERRATER	665	0.9162	ADJ
618	HETEROGENEOUS	660	0.9212	ADJ
619	ADHERE	659	0.9103	VERB
620	INCIDENTALLY	654	0.9331	ADV
621	COMPELLING	654	0.9136	ADJ
622	ASCRIBE	653	0.917	VERB

623	TRANSITIONAL	652	0.9476	ADJ
624	TENET	651	0.9262	NOUN
625	CORRELATIONAL	649	0.9275	ADJ
626	INTERNATIONALLY	649	0.9361	ADV
627	DECODE	647	0.9382	VERB
628	IMMERSE	647	0.9353	VERB
629	CONDUCIVE	645	0.9159	VERB
630	DIVERGE	644	0.9144	VERB
631	PERVASIVE	644	0.9132	ADJ
632	WORTHWHILE	644	0.9159	ADJ
633	UPCOMING	642	0.9384	ADJ
634	EFFICIENTLY	638	0.9082	ADV
635	PROBLEMATIZE	638	0.938	VERB
636	DIVERGENCE	635	0.9367	NOUN
637	DIVERGENT	631	0.921	ADJ
638	IMPEDE	630	0.918	VERB
639	CUTOFF	628	0.9379	NOUN
640	RECIPROCAL	628	0.9352	ADJ
641	CONSECUTIVE	625	0.9156	ADJ
642	SEMINAL	622	0.9106	ADJ
643	EXPERIMENTER	621	0.9453	NOUN
644	CONSOLIDATE	617	0.9262	VERB
645	DIFFERENTIALLY	616	0.9157	ADV
646	CONTEXTUALLY	615	0.924	ADV
647	GLOBALLY	613	0.9313	ADV
648	PRECISION	613	0.9311	NOUN
649	HEIGHTEN	612	0.9181	VERB
650	RECRUITMENT	611	0.9432	NOUN
651	PURSUIT	611	0.9338	NOUN
652	DECONTEXTUALIZE	610	0.9361	VERB
653	SOLE	609	0.9114	ADJ
654	FUNCTIONALITY	607	0.8738	NOUN
655	CONTINUATION	607	0.9438	NOUN
656	PERSISTENT	604	0.921	ADJ
657	COLLECTIVELY	598	0.9226	ADV
658	FRUITFUL	596	0.9131	ADJ
659	CENTRALITY	594	0.9441	ADV
660	PRESCRIBE	589	0.938	VERB
661	INTENTIONALLY	582	0.9176	ADV
662	LASTLY	579	0.9365	ADV
663	TEASE	577	0.923	VERB
664	FAMILIARIZE	562	0.9251	VERB

Appendix B: Academic journals used to create the corpus

Journal
Annual Review of Applied Linguistics
Applied Linguistics
ELT Journal
International Journal of Applied Linguistics
Language Learning
Language Teaching
Language Teaching Research
Studies in Second Language Acquisition
Second Language Research
TESOL Journal
TESOL Quarterly
The Modern Language Journal

References

Abushariah, M. A., Ainon, R. N., Zainuddin, R., Elshafei, M., & Khalifa, O. O. (2012). Phonetically rich and balanced text and speech corpora for Arabic language. *Language Resources and Evaluation, 46*(4), 601-634. http://dx.doi.org/10.1007/s10579-011-9166-8

Barcroft, J. (2009). Effects of synonym generation on incidental and intentional L2 vocabulary learning during reading. *TESOL Quarterly, 43*, 79–103.

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253–279. https://doi.org/10.1093/ijl/6.4.253

Bauman, J., & Culligan, B. (1995). *About the general service list.* Enterprise Training Group. http://jbauman.com/aboutgsl.html

Bos, J., & Spenader, J. (2011). An annotated corpus for the analysis of VP ellipsis. *Language Resources and Evaluation, 45*(4), 463–494. http://dx.doi.org/10.1007/s10579-011-9142-3

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, C. (2014). A new General Service List: The better mousetrap we've been looking for? *Vocabulary Learning and Instruction, 3*(2), 1-10.

Browne, C., Culligan, B., & Phillips, J. (2013). *The New General Service List*. http://www.newgeneralservicelist.org/

Browne, C., Culligan, B., & Phillips, J. (2023). *The New Academic Word List. (NAWL).* Charlie Browne Company Inc. https://www.newgeneralservicelist.com/new-general-service-list-1

Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language, 15*(2), 103–116.

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly, 34*(2), 213–238. https://doi.org/10.2307/3587951

Eldridge, J. (2008). "No, there isn't an 'academic vocabulary,' but..." *TESOL Quarterly*, *42*(1), 109–113.

Fang, A. C. (2012). Creating an interoperable language resource for interoperable linguistic studies. *Language Resources and Evaluation*, *46*(2), 327–340. http://dx.doi.org/10.1007/s10579-012-9189-9

Fois, E. (2021), Translator training, English language teaching and corpora: *Scenarios and applications. International Journal of Language Studies*, *15*, 59–78.

Gardner, D. (2013). *Exploring vocabulary: Language in action*. Routledge.

Gardner, D., & Davies, M. (2013). A New Academic Vocabulary List. *Applied Linguistics*, *35*(3), 305–327. http://dx.doi.org/10.1093/applin/amt015

Gilner, L. (2011). A primer on the General Service List. *Reading in a Foreign Language*, *23*(1), 65–83.

Heuberger, R. (2016). Corpora as game changers: The growing impact of corpus tools for dictionary makers and users: Corpus tools have created a paradigm shift in English lexicography. *English Today*, *32*(2), 24–30. http://dx.doi.org/10.1017/S0266078415000474

Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, *23*, 403–430.

Huong, L. P. H. (2018). An academic word list for English language teaching. *Asian EFL Journal*, *20*(9), 117. https://www.elejournals.com/asian-efl-journal/the-asian-efl-journal-quarterly-volume-20-issue-9-2-september-2018/

Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, *18*(1), 41–62. https://doi.org/10.1111/j.1473-4192.2008.00178.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. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans to thinking machines* (pp. 316–323). Multilingual Matters.

Leng, H. A., & Tan, K. H. (2019). From Lexical Bundles to Lexical Frames: Uncovering the Extent of Phraseological Variation in Academic Writing. *3L, Language, Linguistics, Literature*, *25*(2), 99–102. http://dx.doi.org/10.17576/3L-2019-2502-08

Li, Y., & Qian, D. (2010). Profiling the Academic Word List (AWL) in a financial corpus. *System*, *38*, 402–411.

Malisz, Z., Wodarczak, M., Buschmeier, H., Skubisz, J., Kopp, S., & Wagner, P. (2016). The ALICO corpus: analysing the active listener. *Language Resources and Evaluation*, *50*(2), 411–442. http://dx.doi.org/10.1007/s10579-016-9355-6

Martinez, I. A., Beck, S.C., & Panza, C.B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *Journal of English for Specific Purposes, 28*, 183–198.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*, 59–82.

O'Flynn, J. A. (2019). An Economics Academic Word List (EAWL): Using online resources to develop a subject-specific word list and associated teaching-learning materials. *Journal of Academic Language and Learning*, *13*(1), A28–A87. https://journal.aall.org.au/index.php/jall/article/view/592

Prieto Velasco, J. A. (2013). A corpus-based approach to the multimodal analysis of specialized knowledge. *Language Resources and Evaluation, 47*(2), 399–423. http://dx.doi.org/10.1007/s10579-012-9204-1

Rubinstein, A. (2019). Historical corpora meet the digital humanities: the Jerusalem Corpus of Emergent Modern Hebrew. *Language Resources and Evaluation, 53*(4), 807–835. http://dx.doi.org/10.1007/s10579-019-09458-4

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*(1), 26–43.

Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics, 31*(4), 487–512.

Soruç, A., Tekin, B., & Zych, I. (2017), Vocabulary learning through data-driven learning in an English as a second language setting. *Educational Sciences: Theory & Practice, 17*, 1811–1832.

Snaith, M., Conway, N., Tessa, B., De, F. D., Pease, A., Reshmashree, K., Mathilde, J., Gerwin, H., Catherine, P., & Op Den, A. H. (2021). A multimodal corpus of simulated consultations between a patient and multiple healthcare professionals. *Language Resources and Evaluation*, *55*(4), 1077-1092. http://dx.doi.org/10.1007/s10579-020-09526-0

Tahir, B. & Mehmood, M. A. (2021). Corpulyzer: A Novel Framework for Building Low Resource Language Corpora. *IEEE Access, 9*, 8546-8563.

Vongpumivitch, V., Huang, J.H., & 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. *Journal of English for Specific Purposes, 28*, 33–41.

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes, 27*(4), 442–458.

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