Check for updates

ORIGINAL RESEARCH

The word profile of a Global Online Course for English language teachers: A corpus-based research project

Sebnem Kurt iD, Jeanne Beck, Abdulrahman Alharthi

Iowa State University, Iowa, United States

Abstract

Providing useful reference materials for online course participants is an important aspect for online courses. To aid a course designed to provide professional development to English language teachers from around the world, this corpus-based study investigated the frequency and coverage of Academic Word List (Coxhead, 1998), General Service List (GSL) first 1000 words (1K), and GSL second 1000 words (2K) lists. Gathering course materials and participant discussion board posts into two corpora for this study, frequency and coverage of the three base lists were calculated using AntWordProfiler and AntConc, resulting in a coverage of 9.56 % for AWL words and over 80% coverage for the two GSL lists combined in the first corpus. The high percentage of off-list words (9.75%) in the first corpus and low percentage of AWL words in the second corpus (5.23%) motivated the creation of a new word list that contains the most frequently used words outside of the AWL, GSL 1K, and GSL 2K words from the first corpus to supplement future course participants with technical words that are required to successfully complete the course.

Keywords AWL; English teachers; learner corpus; off-list words; online teacher education

1. Introduction

As the popularity of English as a second, foreign, or additional language worldwide increases, the need for highly trained English language teachers, who are versed in connecting the latest practices to their classrooms, is of utmost importance to prepare learners for the future. The British Council (2013) estimates that there are 12 million English teachers worldwide, many of whom teach in their home countries. While English teachers have historically been limited to teacher education within their situated contexts, due to the vast improvements in internet capabilities worldwide, online teacher education programs can help connect teachers to quality instructional content, as well as to other teachers and trainers.

The Online Professional English Network (OPEN) offers virtual learning opportunities to foreign English language educators, professionals and learners worldwide. Sponsored by the U.S. Department of State, OPEN professional development opportunities are developed by U.S. academic institutions and experts in the field of Teaching English to Speakers of Other Languages (TESOL). One of the virtual learning opportunities is the course *Using Educational Technology in the English Language Classroom*. Referred to as the Global Online Course (GOC) by the course developers, TAs, and course mentors, which inspired the names of these corpora, the course is an eight-week, eight-module teacher training course that focuses on integrating technology into each of the major areas that English teachers typically provide instruction (e.g., vocabulary, grammar, writing). Within this online course, the participants are English teachers in their situated contexts who are nominated to join the course by their local U.S. embassies in cooperation with FHI360, the organization that oversees this course.

Given that this course's potential to positively impact countless classrooms and learners, using a corpus-based approach to analyzing the course and making empirically-based recommendations for improving its content is most desired. Specifically, one area of improvement within the course concerns the glossary pages provided in each unit. While it is hoped that the glossaries provide scaffolding for course participants, who may be exposed to unfamiliar vocabulary words and technical English concepts, these lists have yet to be checked using corpus methods.

Moreover, as course participants are granted access to introductory course materials prior to the start of the course, participants could be provided with a list of key glossary terms of relevance to the whole course, which has the potential to improve their studies.

2. Literature review

There is a great wealth of vocabulary studies within corpus linguistics to provide insight and best practices for informing our work with the GOC. Specifically, research surrounding the General Service List (GSL), Academic Word Lists (AWL), and Off-List Words, will be discussed in the following section.

2.1. General Service List (GSL)

The General Service List (GSL) is a frequency-range-based word list, originally of approximately 2,000 headwords, that represents the most common words in the English language. This list provides learners and instructors alike with the most impactful words to study first, which can help them make the largest gains in comprehending the language. Originally published in a report called the *Interim report on vocabulary selection* (Faucett et al., 1936 as cited in Gilner, 2011) the list was later published as the General Service List (West 1953, as cited in Gilner, 2011). This list has since been updated as the New General Service List (NGSL) (Browne, et al., 2013) and the new-GSL (Brezina & Gablasova, 2013), each boasting additional coverage of the English language.

Corpus-based word lists have been developed for a wide range of subject areas and specific purposes, each returning meaningful results within their context. Whether GSL, NGSL, or new-GSL, researchers have used these word lists in multiple studies as an indicator of the prevalence of basic English words in a given corpus. GSL is often employed in tandem with the Academic Word List (AWL), which will be described in the next section. Despite the prevalent use of the GSL, Ward (2009) highlights issues with assuming low level learners know all of the GSL words. Moreover, issues concerning polysemy in GSL words have been noted as problematic for language learners (Clemmons, 2008).

2.2. Academic Word List (AWL)

Since its creation more than two decades ago, the AWL has been integral in areas of corpus research, providing guidelines for developing and evaluating corpora with the purpose of helping native and non-native learners learn crucial vocabulary for their studies. In this landmark study, Coxhead (2000) compiled the AWL based on different academic sources including academic articles, textbooks, course books, and laboratory manuals. These sources stem from 28 subject areas in four major disciplines: science, law, commerce, and arts. The total number of word families in the AWL is 570, each with different numbers of words. Word families had to be outside the GSL to be included. Some of these word families branch into more than 15 words, such as the word *analyze* while others have only one family member such as the word *job.*

With the popularity and respect of the AWL, some criticism has been placed on its reliability and validity. One of the major critiques underscores the focus of the AWL on particular fields, and subsequently, on how the list does not cover the full range of academic disciplines (e.g., Chen & Ge, 2007; Hyland & Tse, 2007). As a result, additional studies have emerged with the aim of filling the gap between different academic fields and their word coverage lists (e.g., Lei & Liu, 2016; Ward, 2009; Yang, 2015).

In order to better understand how research studies have contributed to the bridge between what words learners empirically need to know for their specific disciplines and what learning materials are available to them, a methodical search for empirical articles that included written corpora and incorporated the AWL were evaluated for appropriateness based on the selection criteria. Coverage and frequency analysis are the main research methods used in the selected articles with frequency analysis used to identify the most common word families in the selected corpora. Through frequency analysis, it is possible to evaluate the coverage of a corpus, with the aim of providing pedagogical implications.

Corpus studies showed a range of approaches in assessing the presence and accuracy of the AWL using frequency analysis. While some studies claim the AWL lacks representation of the most frequent words in all the academic fields (Chen & Ge, 2007), others trust the reliability of the AWL and look for ways to teach the AWL words more effectively (e.g., Li & Qian, 2010). Frequency thresholds for the inclusion of words vary greatly in studies, from that of 13.31 times per million words (Bi, 2020) to 28.57 times per million words (Lei & Liu, 2016). While most studies exclude high frequency words, a tendency among AWL studies, Lei & Liu (2016) chose to include them in their study when they convey special meaning in a particular context. Based on their results, the authors advocate that this approach is more impactful than the commonly accepted, exclude-the-high-frequency-words approach. Moreover, when AWL is tested in new contexts, such as Pathan et al.'s (2018) study on Ph.D. theses in Pakistan, the percentage of the coverage of the AWL is similar to what Coxhead (2000) generally proposed.

2.3. Off-list words

Off-list words are specialized vocabulary words that are commonly used in a "particular topic, field or discipline" (Nation, 2001, p. 198). Also referred to using other terms such as *technical vocabulary*, the prevalence of these words will vary by discipline and can cover up to 5% of a given text (Hyland & Tse, 2007). As such, these words can be highly beneficial for English for Specific Purposes (ESP) instructors and learners in order to pinpoint vocabulary needed for a given discipline. Some studies have found a relatively

high percentage of off-list words, with implications for teaching these words (e.g., Gustafsson & Malmström, 2013). To determine a field's off-list words, after creating a discipline-specific corpus, researchers remove words from the GSL and AWL, or lists of their own devising, to arrive at the off-list words.

Studies that look specifically at off-word lists tend to employ the GSL and AWL to remove both high frequency words and general academic words to arrive at the specialized words that the given field uses. Several example studies have employed such methods to investigate target language content. In a 2019 study of specialized vocabulary in Thai food menus, Low used corpus methods to compare its GSL, AWL, and off-word lists, drawing upon specific AWL and off-word lists words that would be useful for Thai chefs. To compare the frequency of academic vocabulary in abstracts written by experts, as compared with Chinese undergraduates, Wang (2014) found statistically significant differences in the 1K and Off-list vocabulary usage, drawing pedagogical implications from the results. Beyond looking at expert text, corpus studies have used off-list words to primarily investigate student vocabulary knowledge. Silva et al.'s (2018) study investigated the vocabulary used by Brazilian students in their written assignments, and how it compared and differed to other corpora. Through this analysis, including the students' use of off-list words, the researchers arrived at some pedagogical implications, especially considering different word usage between life sciences and physical sciences and the explicit teaching of lexical bundles used in these disciplines.

2.4. Research questions

As the course contains both expert writing in terms of its course content and student writing from student discussion boards and other deliverables, there are ample resources to investigate the frequency of AWL and off-word lists among the course materials, glossary pages, and student writing to help us with our goals of improving the course's glossary pages. Therefore, the following research questions have been developed:

1. How does the frequency of the AWL words compare across the course instructional materials and participant writing?
2. What is the alignment between the module glossaries and the AWL wordlist? What is the alignment with off-list words?
3. What are the words that GOC course participants need to know in order to successfully interact with the course materials?

3. Methods

This study adopted a corpus-based research approach to evaluate the frequency and coverage of AWL, GSL 1K and GSL 2K words in a specialized online teacher education course for English language teachers from around the world, with the aim of improving the course materials. Two corpora, which were built from the GOC course and participant writings, will be discussed in the following section.

3.1. Data collection

To answer the proposed research questions, two corpora were constructed based on the course. The first corpus, named CyGOC, contains all the course content that learners encountered in the course, including the course syllabus, module content, and video transcripts. The rationale for including all of the course contents in CyGOC is to represent all of the materials that learners may encounter in this course. As such, the corpus includes 12 journal articles that were included in the course as reading assignments and were included in the downloadable packets for participants with lower bandwidth or who prefer their materials in PDF format. The references, appendices, tables, figures, notes, endnotes and footnotes of these articles were removed as part of the corpus cleaning process. CyGOC contains a total of 148,043 words.

The second corpus created for this study, named CyDis, concerns the discussion board posts that participants complete on a weekly basis for the course. The course iteration that participant data came from was the Fall 2020 iteration, which had a total of six sections of approximately 24 students per class. Of these participants, 49 enrolled in the Fall 2020 iteration agreed to make their data available for research purposes by signing a consent form which was provided within the Canvas modules. CyDis totals 77,370 words, which encompasses 179 text files of learners' discussion posts and 329 text files of replies from Module 2 through Module 5. The reasoning for selecting this period of the course for CyDis is due to practical reasons; since Module 1's discussion board posts are more introductory in nature, during the first week participants become acclimated to the posting practices and grading procedures of the course. On the other hand, discussion posts in the later units may decline as participants prioritize completing their final project, which is worth considerably more points than the discussion posts. CyDis is used to sample the writing of a particular group of learners - English language teachers from around the world who are selected to enroll in the GOC.

3.2. Data analysis and processing

Creating the CyGOC and the CyDis resulted in 620 files: 112 text files for the former and 508 text files for the latter. In order to answer the research questions, two computer software programs were utilized. To calculate coverage of the AWL and GSL lists, AntWordProfiler (Anthony, 2009) was used. AntConc (Anthony, 2014) was utilized to calculate the word frequency for the ten AWL sublists. It should be noted here that AntWordProfiler has the target lists AWL, GSL 1K, GSL 2K already built into the software, however, AntWordProfiler does not offer separate searches for the ten AWL sublists. As a workaround, ten

text files that contain one AWL sublist per text file were created and then uploaded to AntConc to acquire the word frequencies for the AWL sublists in each corpus. Running the AntWordProfiler on the corpora generated output on the overall coverage of the three lists: AWL, GSL 1K, GSL 2K and off-list words, as well the number of tokens identified from each list. All outputs were saved as .txt files and were exported to Excel as delimited files to facilitate the data analysis.

4. Results

The current study focused on the frequency and the coverage of the academic words in a GOC, which is provided for English language teachers from around the world with varying English proficiency. This section presents the results in detail along with relevant data displays and tables according to the research questions. This section will also discuss the findings in

relation to previous studies in the literature.

RQ 1. The frequency of the AWL words and the off-lists words across the course instructional materials and participant writing

4.1. CyGOC coverage

CyGOC, which was created from the course content, has a total number of 148,043 words. Table 1 below displays the number of tokens across the three base lists: AWL, GSL 1K, GSL 2K, as well as the off-list words for the course modules, assigned readings, and video transcripts. From Table 1, it is seen that the course readings have the largest number of AWL tokens while the introductory texts, which could be referred to as Module 0 content, have the lowest number. This result might not be surprising considering the course readings are research articles; since AWL is an *academic word list*, it is fitting that the AWL words occur more frequently in more academic texts.

Table 1. Number of tokens for CyGOC across the modules

	AWL	GSL 1K	GSL 2K	Off-list	Total
Introductory texts	207	1,322	69	231	1,829
Module 1	273	2,022	111	308	2,736
Module 2	335	2,670	325	425	3,665
Module 3	471	3,725	331	509	5,036
Module 4	354	2,269	139	309	3,071
Module 5	557	3,369	248	386	4,560
Module 6	236	1,608	123	247	2,214
Module 7	272	2,206	188	365	3,031
Module 8	436	2,749	157	363	3,705
Readings	7,500	45,858	2,912	6,799	63,069
Video Transcripts	3,173	41,466	2,747	4,147	51,533
CyGOC Overall	14,151	112,057	7,403	14,432	148,043

After the number of tokens were acquired per list, the coverage was investigated. It is worth highlighting here that the readings and video transcripts were separated from their respective modules during the

analysis to account for register differences. Table 2 shows the overall coverage of the base lists and the off-list words in the CyGOC corpus.

Table 2. CyGOC coverage across the modules

	AWL %	GSL 1K %	GSL 2K %	Off-list %
CyGOC Overall	9.56	75.69	5	9.75
M0	11.32	72.28	76.05	12.63
M1	9.98	74.71	4.06	11.26
M2	9.14	72.85	6.41	11.6
M3	9.35	73.97	6.57	10.11
M4	11.53	73.88	4.53	10.06
M5	12.21	73.88	5.44	8.46
M6	10.66	72.63	5.56	11.16
M7	12.04	72.78	6.2	12.04
M8	11.77	74.2	4.24	9.8
Readings	11.89	72.71	77.33	10.78
Video Transcripts	6.16	80.46	5.33	8.05

The overall AWL text coverage in the CyGOC is 9.56 %, which is in alignment with the proposed 10 % AWL coverage in all academic texts in literature

(Coxhead, 2000). This result is higher than some other reported studies, such as the 4.66 % in Konstantakis' (2007) business word list and the 4.94 % in Mozaffari

and Moini's (2014) study which investigated the presence of AWL words in education research articles. However, it is lower than some recent studies, as in the example of Gholaminejad and Anani Sarab (2020), in which the AWL coverage in English language teaching textbooks was reported as 11.13 %. This result is impressive considering the size of their corpus - 11 million words.

In terms of AWL coverage, it is highest in Module 5, the module on teaching writing, and lowest with 6.16 % for the video transcripts, which is expected due to the nature of spoken language in the videos. Contrarily, GSL 1K has the highest coverage in the video transcripts with 80.46 %, whereas it is around the same percentage in all other modules. As a whole, the off-list words seem to have a remarkable coverage, ranging from 8.05% to as high as 12.63% across the modules. Seeing the coverage of the AWL and the off-list words around similar percentages calls for a further

analysis to investigate the off-list words in the GOC. GSL 2K has the lowest coverage among the lists in all modules; among the modules, GSL 2K is highest in Module 3, which is on grammar. In terms of off-list words, the coverage is the highest for the Introductory texts (Module 0) and Module 7, the module on teaching speaking, while it is the lowest in the video transcripts. Nonetheless, off-list words have relatively similar coverage compared to the AWL list with less than 0.20% difference between the two lists in the overall CyGOC coverage.

As a further investigation, CyGOC was analyzed with ten individual sublists of AWL to detect which sublist has a higher or lower coverage. Table 3 displays the type and token distribution of AWL sublists in CyGOC. Sublist 1 has the highest numbers of AWL types and tokens whereas Sublist 10 has the lowest numbers, which is also represented graphically in Figure 1.

Table 3. CyGOC type and token distribution across AWL sub lists

Sub list	1	2	3	4	5	6	7	8	9	10
Type	230	189	154	137	123	124	126	88	68	27
Token	3,741	3,064	1,981	1,367	898	1,335	705	632	516	538

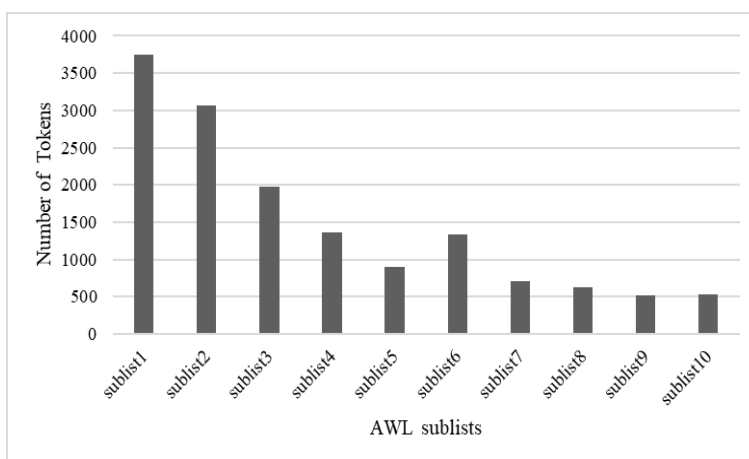**Figure 1.** AWL sub lists in CyGOC by number of tokens

Table 3 and Figure 1 reveal that there is a gradual decline in the usage of the AWL in terms of tokens from the given sublists becoming less frequent. This finding is in line with what other researchers found in the literature, that earlier sublists in the AWL are used more compared to later ones (for example, see Pathan et, al., 2018)

4.2. CyDis coverage

CyDis consists of discussion posts and replies

from the 49 participants in the six Fall 2020 course sections from Module 2 through Module 5. CyDis differs from CyGOC in that CyDis is a learner corpus in the sense that the English teachers are non-native speakers who are learning new content and concepts related to integrating technology into their English courses. The number of tokens for each base list across each module in the CyDis is shown in Table 4. Figure 2 shows that GSL 1K is significantly higher than other lists and that all other lists notably remain under 2,000-word level across all modules.

Table 4. Number of Tokens in CyDis

	AWL	GSL 1K	GSL 2K	Off-list	Total
CyDis Overall	4,045	63,359	4,508	5,458	77,370
Module 2	907	16,403	1,104	1,683	20,097
Module 3	1,152	17,413	1,578	1,575	21,718
Module 4	920	15,474	835	1,018	18,247
Module 5	1,066	14,069	991	1,182	17,308

Figure 2. Number of tokens in CyDis
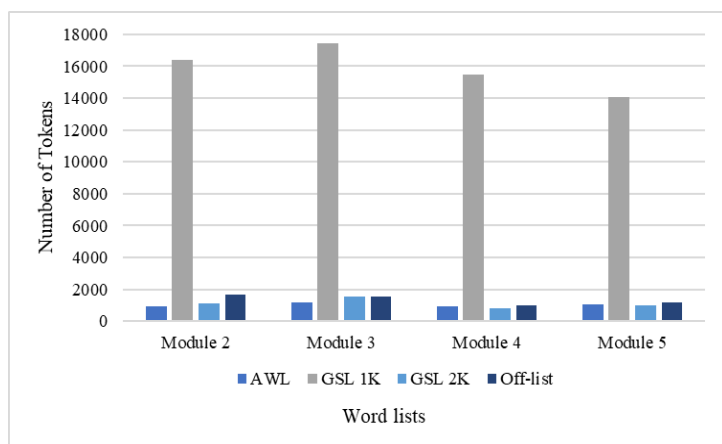
In terms of coverage, CyDis is significantly lower compared to CyGOC when it comes to the AWL words. It has 5.23 % overall AWL coverage compared to CyGOC, which has 9.56%. Moreover, as can be seen in Table 5, GSL 1 and 2K have higher coverage in CyDis compared to CyGOC. However, off-list coverage is lower in CyDis compared to CyGOC. One question that begs answering regards the relatively low coverage of AWL in the CyDis corpus - whether this is due to register differences or participants' English levels. When participants sign up for the course, they do not submit any documents indicating their English proficiency. Therefore, more research is needed to determine the causes of low AWL coverage in this learner corpora, and to what extent including other course assignments would help improve AWL coverage. As discussion boards provide a platform for social interaction in online learning, they carry elements from written and spoken language (Chen et al., 2018). In Rudy et al.'s (2019) study based on a spoken corpus of medical students who were English learners, the researchers found a coverage as low as 1.5 % for AWL words. Considering these two studies' findings, it would be interesting to compare the language of these discussion boards within the course to spoken recordings that the participants submit as part of one major assignment to see how AWL usage differs.

Table 5. CyDis coverage across the modules

	AWL %	GSL 1K %	GSL 2K %	Off-list %
CyDis Overall	5.23	81.89	5.83	7.05
Module 2	4.51	81.62	5.49	8.37
Module 3	5.3	80.18	7.27	7.25
Module 4	5.04	84.8	4.58	5.58
Module 5	6.16	81.29	5.73	6.83

A general increase in the AWL coverage is observed from the earliest modules to the latest ones with the lowest AWL coverage in Module 2, the vocabulary module, and the highest coverage in Module 5, the writing module. This could be due to the participants' gradual adaptation to the course language as the weeks go by and thereby using more appropriate academic vocabulary when writing their discussion posts and replies. However, the highest coverage level of AWL in CyDis is still lower than the lowest AWL coverage level in CyGOC.

Furthermore, to shed more light on the learner corpus, the CyDis, Table 4 shows the breakdown of tokens in the corpus across the word lists and the four sampled modules. It could be seen from Table 4 that the total number of AWL tokens is 4,045, a rather small number compared to 63,359 GSL 1K tokens. The highest number of AWL tokens is in Module 3, the grammar module, whereas the lowest number is in Module 2, the vocabulary module. Moreover, the total number of off-list words is 5,458 compared to 14,432 words in CyGOC.

RQ 2. The alignment between the module glossaries and AWL and off-list words.

Module glossaries are provided for each module with the key words, phrases, or concepts that the course developers believe are necessary for that module. There are eight module glossaries in total. The vocabulary items in these glossaries could be single words (e.g.,

authenticity), two-to-three word phrases (e.g., critical thinking, high-frequency words), acronyms (e.g., COCA, AWE) or websites (e.g., Lingro, Voice of America). The terms in the glossary are introduced with their definitions or explanations. A closer look at the terms in these glossaries reveals that these are highly technical concepts from the field of English language teaching (e.g., incidental vocabulary learning, needs assessment, descriptive grammar). Some terms occur more than once across the glossaries (e.g., authenticity, register, corpus). Additionally, some words describe field-related tests, such as Flesch-Kincaid, a test designed to indicate how difficult a passage in English is to understand, and Gunning Fog, a readability test for English writing to estimate the years of formal education a person needs to understand the text on the first reading (Chapelle & Hegelheimer, 2022).

Learning vocabulary from a list, such as a glossary, has some benefits, especially for low level learners. In the course, glossaries serve to introduce discipline-specific words, which are of crucial importance in vocabulary learning (e.g., Gustafsson & Malmstrom, 2013; Hyland & Tse, 2007). However, the number of words provided in the GOC module glossaries are fairly few in number, ranging between 10-25 words/phrases for each module. When analyzing the alignment between glossary words and AWL, GSL 1K and 2K and off-list words, the frequency and coverage of glossary words were evaluated by including all the

glossary words from the course in a .txt file and running an analysis on this file in AntWordProfiler with the three base-lists. The results displayed a total of 2,897 tokens in the glossary file. Table 6 displays the breakdown of these numbers across the three word lists. As can be seen from the table, approximately 43% of the course glossary tokens were not covered by the three base-lists.

Table 6. Glossary coverage in CyDis

	AWL	GSL 1K	GSL 2K	Off-list
Frequency	383	2,013	177	324
Coverage	21.49%	26.45%	8.26%	43.80%

The high percentage of glossary tokens from the off-list may indicate that course designers tended to choose advanced words specific to the field of English language teaching. Nonetheless, more than 55% of all tokens from the glossary words still come from the three base lists that advanced learners, let alone teachers, should know (Nation, 2013). Therefore, the effectiveness of these glossaries may need further consideration from the course designers. The focus of these glossaries should be on helping GOC participants understand and use more advanced specific words. Consequently, an update to the currently short glossaries, in the form of adding the technical words that this present study has recommended, could supply future learners with more scaffolding to increase their vocabulary range with words relevant to their field.

RQ 3. The words that course participants need to know in order to successfully interact with the course materials.

Nation (2001) proposed that vocabulary is divided into high-frequency words, academic words and technical words, which is represented in this paper as GSL, AWL and the off-list words respectively. The frequency and coverage of the CyGOC and CyDis corpora, as evaluated in this paper, revealed that all three are represented with differing percentages in the GOC. Figure 3 displays the token and type percentages of all the word lists under investigation. It is seen from the table that the type percentage of the off-list words is almost as high as the two GSL lists combined.

Figure 3. CyGOC overall coverage percentages
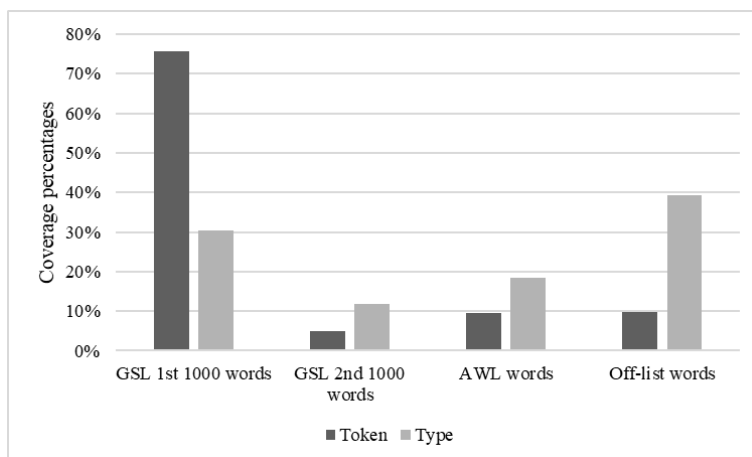
Due to GSL's high coverage, it could be argued that through CyGOC, learners are exposed to high frequency words during the course. The AWL's 9.56 % coverage is close to Coxhead's (2000) reported coverage of 10% AWL words in all academic texts. However, it is apparent that the off-list words cover a higher percentage of the words than the AWL words. This could be explained by the fact that CyGOC is a specialized corpus which requires technical words, as suggested by Lu and Durrant (2017). Therefore, learners could benefit from learning the specialized vocabulary in their field in order to be successful in the course, as suggested by many researchers (e.g., Martínez, et al., 2009; Vongpumivitch, et al., 2009; Yang, 2015). For this reason, a word list containing the most frequently used words from the off-list words were bundled together into a new list named CyVoc to supplement the future participants with the technical vocabulary needed to have a better coverage of course materials. In the creation of this list, only the words which occurred at least 30 times were chosen. This is a comparatively smaller threshold compared to Coxhead's rule of words occurring 100 times when creating the AWL. However, considering the smaller corpus size of this paper, the number could be regarded as frequent. Additionally, expert opinions from the course developers and teaching assistants will be obtained in order to ensure the relevance of the words included in the list.

To provide more insight into the current study, Table 7 was created to display which AWL words are used the most in the materials to help future course designers and corpus researchers understand how often the AWL words appear in the course. However, due to the limitations in time and considering the scope of this paper, we presented here the five most frequent AWL words per module.

Table 7. Five most frequent AWL words per module

	Word 1 (freq)	Word 2 (freq)	Word 3 (freq)	Word 4 (freq)	Word 5 (freq)
Introductory texts	participate (20)	community (13)	contact (13)	instruct (11)	professional (10)
Module 1	technology (34)	task (32)	expert (17)	assess (12)	lecture (12)
Module 2	task (28)	lecture (24)	resource (23)	technology (22)	media (17)
Module 3	function (39)	resource (31)	lecture (27)	task (27)	technology (22)
Module 4	text (79)	task (23)	select (16)	lecture (15)	assign (14)
Module 5	lecture (39)	technology (35)	assign (29)	task (29)	process (25)
Module 6	resource (23)	task (22)	lecture (16)	technology (15)	assign (14)
Module 7	task (26)	technology (23)	resource (20)	assign (14)	create (14)
Module 8	assign (32)	technology (31)	task (27)	project (25)	process (22)
Readings	process (226)	technology (192)	create (169)	text (157)	project (147)
Video Transcripts	technology (183)	text (129)	process (110)	create (95)	context (62)

Words including technology, task, lecture, assign, create, process are among the most frequently occurring AWL words. Given the course objective, to "help teachers integrate pedagogical knowledge and skills with technology to enhance language learning and teaching through course readings, discussions, and assignments, which create new learning activities" (Chapelle & Hegelheimer, 2023 p. 1), the prevalence of the words seems quite fitting.

5. Conclusion

Creation of the CyGOC and CyDis corpora have allowed a data-driven look into the workings of the GOC, offering suggestions for how to better tailor the materials provided in this course for participant needs. Off-list words make up as much frequency as AWL words, therefore both lists are important for the course. However, while the course designers incorporated some technical vocabulary into the glossaries for the course, participants have shown a tendency to use words that are more basic. Therefore, by enhancing the effectiveness of the course glossaries by adding words that are needed in GOC, the course can better encourage the internalization and use of more specified English vocabulary to the course content and learners' English language proficiency level.

Competing Interests
The authors declare none.

Sebnem Kurt (Corresponding Author) is a Ph.D. student in the Applied Linguistics and Technology program, co-majoring in Human-Computer Interaction at Iowa State University. She worked as an EFL instructor at Akdeniz University for over a decade before she started her doctoral studies. She is a Fulbright scholar and a TESOL certificate holder. Her research interests include computer-assisted language learning, second language assessment and project-based learning. Email: skurt@iastate.edu

Jeanne Beck is a Ph.D. student in the Applied Linguistics and Technology Program at Iowa State University. She has an education specialist degree (Ed.S.) in administration and an M.Ed. in TESOL. A certified K-12 teacher and education administrator, she also currently works as an EL coordinator/teacher at the elementary level. Her main interests include second language assessment, teacher training, CALL, and project-based learning.
Email: beckje@iastate.edu

Abdulrahman A. Alharthi is a Ph.D. student in the Applied Linguistics and Technology program at Iowa State University. He has a bachelor's degree in English Language from King Abdulaziz University in Saudi Arabia and a master's degree in Applied Linguistics from the University of Nottingham in the United Kingdom. His research interests include language assessment, L2 listening comprehension, vocabulary acquisition, language motivation, and Directed Motivational Currents (DMC).
Email: aaharthi@iastate.edu

References

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Available from https://www.laurenceanthony.net/software.html

Anthony, L. (2009). AntWordProfiler. [Computer software]. Available from https://www.laurenceanthony.net/software/antwor

dprofiler/

Bi, J. (2020). How large a vocabulary do Chinese computer science undergraduates need to read English-medium specialist textbooks? *English for Specific Purposes, 58*, 77-89. https://doi.org/10.1016/j.esp.2020.01.001

British Council. (2013). The English effect: The impact of English, what it's worth to the UK and why it matters to the world.https://www.britishcouncil.org/sites/default/files/english-effect-report-v2.pdf

Brezina, V., & Gablasova, D. (2013). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics, 34*(5), 1-23. https://doi.org/10.1093/applin/amt018

Browne, C., Culligan, B., & Phillips, F. (2013). A new general service list. http://www.newgeneralservicelist.org/

Chapelle, C. & Hegelheimer, V. (2022). *Using educational technology in the English language classroom.* [Online course]. OPEN. https://classroom.openenglishprograms.org/courses/432

Chen, Q., & Ge, G. C. (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

Chen, B., Chang, Y. H., Ouyang, F., & Zhou, W. (2018). Fostering student engagement in online discussion through social learning analytics. *The Internet and Higher Education*, *37*, 21-30. https://doi.org/10.1016/j.iheduc.2017.12.002

Clemmons, K. (2008). *The problem of polysemy in the first thousand words of the General Service List: A corpus study of secondary chemistry texts.* (Publication No. 3319232.) [Doctoral dissertation, University of Central Florida]. ProQuest Dissertations Publishing.

Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, 34(2), 213-238. https://doi.org/10.2307/3587951

Gholaminejad, R., & Anani Sarab, M. R. (2020). A study of the frequency of academic vocabulary in a corpus of academic textbooks of the English Language Teaching. *Foreign Language Research Journal*, *10*(1), 198-215.

Gilner, L. (2011). A primer on the General Service List. *Reading in a Foreign Language*, *23*(1), 65-83. https://files.eric.ed.gov/fulltext/EJ926367.pdf

Gustafsson, M. & Malmstrom, H. (2013). *Master level writing in engineering and productive vocabulary: What does measuring academic vocabulary tell us?* In N-L. Johannesson, G. Melchers, and B. Bjorkman (eds.), Of butterflies and birds, of dialects and genres: Essays in honour of Philip Shaw. (pp.224-243). Stockholm University.

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

Konstantakis, N. (2007). Creating a business word list for teaching business English. *ELIA, 7, 79-102.*

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

Low, P. (2019). Identifying specialized vocabulary in Thai food menus using computer-based approach. *Journal of Advances in Information Technology*, *10*(3). http://www.jait.us/uploadfile/2019/0823/20190823114233928.pdf

Lu, C. & Durrant, P. (2017). A corpus-based lexical analysis of Chinese medicine research articles. *The Asian Journal of Applied Linguistics*, *4*(1), 3-15. https://caes.hku.hk/ajal/index.php/ajal/article/view/437

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

Mozaffari, A., & Moini, R. (2014). Academic words in education research articles: A corpus study. *Procedia-Social and Behavioral Sciences*, *98*, 1290-1296. https://doi.org/10.1016/j.sbspro.2014.03.545

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge. http://dx.doi.org/10.1017/CBO9781139524759

Nation, P. (2013). Learning vocabulary in another language (2nd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139858656

Pathan, H., Memon, R. A., Memon, S., Shah, S. W. A., & Magsi, A. (2018). Academic vocabulary use in doctoral theses: A corpus-based lexical analysis of Academic Word List (AWL) in major scientific disciplinary groups. *International Journal of English Linguistics*, *8*(4), 282-288. https://doi.org/10.5539/ijel.v8n4p282

Rudy, M., Kristina, D., & Tarjana, S. S. (2019). Measuring spoken vocabulary load on medical English students: A learner corpus evaluation. *International Online Journal of Education and Teaching*, *6*(4), 774-787.

Silva, L. G., Matte, M. L. & Sarmento, S. (2018). Brazilian students' use of English academic vocabulary: An exploratory study. *Linguistica de Corpus*, 509-526.

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, Y.   (2014). A corpus-based contrastive study on applied vocabulary of RA abstracts in applied linguistics. *Applied Mechanics and Materials*, *631-632*, 1442-1446. https://doi.org/10.4028/www.scientific.net/amm.631-632.1442

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

Yang, M. N. (2015). A nursing academic word list. *English for specific purposes*, *37*, 27-38. https://doi.org/10.1016/j.esp.2014.05.003