Research in English Language Pedagogy (2024) 12(4): 652-672

 DOI: 10.30486/RELP.2024.897174

Original Research

A Corpus-Based Study of Academic Vocabulary in Sociology Research Articles: Which Words Should a Teacher Concentrate on?

Reza Bagheri Nevisi[*1], *Mahmood Safari*[2], *Mahnaz Seddighi*[1]

[1]Department of English Language and Literature, University of Qom, Qom, Iran

[2]English Department, Hazrat-e Masoumeh University, Qom, Iran

Submission date: 13-01-2024 Acceptance date: 05-10-2024

Abstract

There have always been too many words in a new language to learn. Therefore, prioritizing important words over the others for learning, i.e., setting vocabulary-learning goals, is of paramount importance. One of the most effective means of prioritizing specific vocabulary items over other words to learn, is the expansion of a technical word list of the most common ones. This research aimed at reporting a corpus-based lexical study of the most frequently-used words within 8 sub-branches of sociology research articles. A technical word list for research articles in Sociology (STWL) was developed based on a corpus of 3,552,900 running words of 508 research articles in 8 subfields of the academic discipline of Sociology, which were compiled from reputable scholarly journals and analyzed via Range. Results indicated that with 1910 words STWL could cover 87.4% of the words running in the Sociology Technical Corpus (STC), while, based on the analysis, GSL-AWL had only 84.19% coverage over the same corpus. Therefore, STWL can be utilized as a vocabulary source for sociology learners and researchers to better understand the concepts of this field, as well as for ESP instructors and syllabus designers who are not familiar with the specialized terms and words of this field.

Keywords: Academic Word List, English for Specific Purposes, Sociology, Technical Words, Word Families

* Corresponding Author's E-mail: re.baghery@gmail.com

1. Introduction

Vocabulary learning has long been viewed as an essential element of language learning (Nation, 2001; Wang et al., 2008) because language learners' vocabulary breadth and depth will directly impact their reading and writing proficiency (Nation, 2001). Due to the increasing instructional and pedagogical merits of technical or specialized word lists and the growing demands to teach EFL learners in their particular fields, considerable care has recently been paid to such vocabularies utilized in academic texts (Yang, 2015). Academic vocabulary, which is of paramount importance in academic texts, is comprised of highly frequent words across a wide range of academic disciplines, which are typically absent in general English texts (Farrell,1990). Given the tremendous vocabulary size of English language, and to boost the efficacy of its teaching, investigators have come up with word lists of the recurring and frequent vocabulary of academic texts across various scientific disciplines. Research into specialized or technical vocabulary has largely concentrated on generating word lists of technical vocabulary in professional fields of expertise in English for Specific Purposes (ESP), such as Nursing (Yang, 2015), Engineering (Hsu, 2014; Ward, 2009; Watson-Todd, 2017), Medicine (see Hsu, 2013), tertiary level research in English for Academic Purposes such as Science (Coxhead & Hirsh, 2007), general academic vocabulary (Coxhead, 2000; Gardner & Davies, 2014), veterinary (Safari, 2018) and Politics (Bagheri Nevisi et al., 2023). Word lists can be employed to recognize the vocabulary required by students of a particular field of study and determine the number of words required to grasp a reading comprehension passage or a listening file (Nation, 2016). Most non-native English students find it challenging to learn academic vocabulary, specifically if the goal is to attain a high literacy level in the target language. Some post-graduate learners are obliged to write their thesis/dissertation in English regardless of their selected field of specialization. Furthermore, some academic institutions even go further than that and necessitate their graduate students to publish research papers in international journals (Yang, 2015). Accordingly, students of sociology who are offered their pertinent courses and texts in English might find it a daunting task to comprehend and write well for academic purposes in the target language since they are not familiarized with such technical vocabularies in English. Although these pupils have learned English for some years in an EFL setting, their command of English remains low-to-intermediate.

A sociology academic word list might play a pivotal role in setting vocabulary

objectives for language courses, boosting learners' autonomy, and assisting course designers to design better pedagogical materials, choose the right passages and create appropriate instructional tasks. The word-forms and word families found in the corpus can be included in EAP writing and reading courses and instructors can direct students' attention to such specialized vocabularies and ask students to apply them in their academic writings. Materials designers can also make academic English textbooks specially-designed to teach political academic vocabulary commonly-used in sociology research papers.

While many word lists of academic words have been developed in other majors, no list has exclusively targeted sociology. Such word list can be helpful to both graduate students of sociology and their instructors. This study attempts to develop such a list to serve as a point of reference for ESP instructors and material developers in English for Sociology Purposes curriculum preparation and English for Academic Purposes (EAP) textbook design, to provide EAP practitioners with more evidence who are interested in producing field-specific academic word lists and to expedite students' learning of academic vocabulary. Such specialized word list aimed exclusively at pupils of sociology can be trained and directly discerned in the similar way as the words from West's (1953) General Service List of English words (GSL) and Coxhead's (2000) Academic Word List (AWL).

2. Literature Review

2.1 Word Lists

Word lists, as an organized list of word items, are generally used for numerous pedagogical purposes, such as setting learning goals, analysis of a text or corpus vocabulary load, recognition of language learners' vocabulary needs, identification of a specific field's specialized vocabulary, material development for language teaching (Nation, 2016). However, choosing the words that are worth focusing on for a special purpose of studying is one of the most perplexing perspectives of learning and teaching vocabulary (Coxhead, 2000).

The basis on which the words are selected from a corpus for creating a word list, i.e. word selection criteria, is also of great importance. For developing a technical word list, the method of extracting the most frequently used terms in the corpora specific to the related field can be adopted (Dudley-Evans & St. John, 2011). Frequency of occurrence is a critical factor in selecting words for a word list and this factor has long been known to be widely different from word to word (Kanzaki, 2019). A word list based on the incidence of word families in a corpus

of texts can give valuable data regarding the actual use of words (Biber et al. 1998).

One of the ways to create a technical wordlist for ESP learners is to analyze discipline-specific corpora to find the words which are frequently-used therein (Dudley-Evans & St. John, 2011). Research articles are of greatest significance for helping the scientists and researchers around the world with the latest developments in various fields of study (Kanzaki, 2019). The vast majority of the abovementioned RAs are published in English (Kanzaki, 2019) so an accurate understanding of the concepts contained in them depends on a relative mastery of their specialized vocabulary. Since many of the researchers who use these research articles are not native speakers of English, they certainly do not know all the words in these papers, so a list of the most normally utilized words in the research articles of each discipline would alleviate this problem to some extent.

Technical vocabulary enjoys a number of essential characteristics and can constitute an enormous proportion of the text. Some of these technical words will occur more frequently than others, which can have far-reaching implications for learning and teaching (Tongpoon-Patanasorn, 2018). Such vocabulary belongs to a narrow domain of language (Nation, 2001). According to Nation (2001), "one person's technical vocabulary is another person's low-frequency words," which reflects the fluctuations of low frequency words (p. 20). Nation (2001) favored mastering frequently-used words first since they were far more advantageous to the learning process and could bring about better learning outcomes technical words can be regarded as words whose meanings are closely associated with one specific discipline, including engineering, medicine, or psychology. Technical words can be words listed in the GSL or the AWL as well. Nevertheless, due to their discipline-specificity, such words might have various meanings in various fields. For instance, the word input is regarded and classified as a highly-frequent word in the GSL, whereas it is also categorized as a specialized vocabulary in Engineering (Tongpoon-Patanasorn, 2018).

2.2 Relevant Studies

ESP learners, the learners who are learning English for specific purposes, need a vocabulary list specific to their field, which most likely have different meanings and function in different academic disciplines. Depending on the discipline, words behave differently and lexical items meaning differ across different registers (Hyland & Tse, 2007). Moreover, the linguistic features are different in different types of disciplines (Biber, 1989). Academic

English has many linguistic features—including vocabulary—that vary from one discipline to another (Bieber et al. 1994). Based on these arguments, studies have recently been conducted that try to find such words, i.e., the most common technical words for some academic disciplines (Coxhead & Demecheleer, 2018; Hsu, 2018: Tongpoon-Patanasorn, 2018). Recent corpus-based studies on academic vocabulary focus mainly on more specific academic areas. There have been plenty of studies developing disciple-specific and subject-specific academic word lists for various academic fields. Capitalizing Coxhead's (2000) AWL, Martinez et al., (2009) determined the academic words in an agricultural paper. Findings revealed that only a small word list from AWL, 92 families existed. The results pointed to the needs to generate field-specific academic word lists.

Hsu (2013) provided a word list of medicine to decrease the divide between non-specialized and specialized vocabularies. The corpus comprised of 155 books with 31 medical themes collected from e-book databases accounting for a total number of 15 million running words. Frequency and range of words beyond the most common 3,000-word word families were completely examined and 595 of the highly frequent word families were finally chosen and constituted the MWL.

Munoz (2015) scrutinized a corpus of 700 agricultural research articles in English. The investigators utilized a mixed-methods approach and the outcomes revealed high lexical variability in the corpus and low word range. Academic words covered merely 6% in comparison to the 10-12% coverage reported for academic articles but the coverage was greater compared to newspapers 4% coverage.

Yang (2015) inspected the most frequently used nursing academic words within various nursing subfields. A 1,006,934-word corpus involving 252 articles of nursing was collected. NAWL including the most used nursing words was extracted. The list comprised 676 word families accounting for approximately 13.64% of the nursing corpus. The outcomes illustrated that it is important to make domain-specific word list at tertiary levels for EFL nursing students to solidify their total academic language proficiency.

Tongpoon-Patanasorn (2018) made a technical vocabulary list in the finance field via taking up a hybrid method. The list entailed 979 finance-related words, that were sub-classified into 569 word families. Such words were listed in both GSL (413) and AWL (291) words.

Safari (2019) delved into the most frequent words in equine veterinary academic scholarly papers. A 3.6 million corpus of running words was inspected by some text analysis

software. The outcomes recommended that 1091 GSL word families and 116 AWL word families were less frequent in the Equine Veterinary Corpus (EVC). Furthermore, 214 non-technical word families were normally applied in the EVC but were non-existent in the GSL and AWL. The resulting list showed a 2.5% higher coverage than the list of GSL and AWL words together, despite containing 993 fewer words.

Heidari et al. (2020), working on a 3.45-million-word corpus of Pharmacy, identified the common words in pharmacy discipline and designed their Pharmacy Academic Word List (PAWL), which consisted of 750 word families and covered 17.69% of the tokens. The outcomes proved the necessity to collect field-specific word lists to deal with the needs of EFL/ESL practitioners and postgraduates over various fields.

Jamalzadeh & Chalak (2019) made their list of physiology academic vocabulary list through scrutinizing a 1.7 million-word corpus. Then, 1450 frequent word families were extracted and made up the Physiotherapy vocabulary list. The outcomes illustrated that AWL might not be totally beneficial to such learners because of its low word coverage and restricted use of frequent physiotherapy vocabularies.

Bagheri Nevisi et al. (2023) delved into a high-frequency wordlist in politics and found the most common vocabulary items in political science. Despite the development of these academic vocabulary lists in above-mentioned fields, no list has specifically addressed sociology. Such word lists can be helpful to both graduate students of sociology and their instructors. This research aims to develop such a list to function as a point of reference for ESP educators and material developers in English for sociology Purposes, to provide EAP practitioners with more evidence to generate domain-specific or specialized vocabulary lists and to expedite learners' learning of academic vocabulary. Such specialized word list aimed primarily at students of sociology can be trained and directly examined in the same way as the words from GSL and AWL. Hence, the researcher formulated the following research questions to achieve the above-stated objectives of the research:

Q1: What high frequency words, across and beyond BNC/COCA 25,000-word list, make a sociology academic word list (SAWL)?

Q2: How do the size and coverage of the SAWL over a sociology academic corpus (SAC) compare to those of lists containing general service and general academic English words (i.e., GSL-AWL and NGSL-NAWL)?

Q3: How many words does the SAWL share with the GSL-AWL and NGSL-NAWL

words? How many words are exclusive to the SAWL?

3.Methodology

3.1 Sociology Academic Corpus (SAC)

Initially, the researchers developed a corpus of sociology research articles in order to identify the highly frequent words in sociology academic texts and develop a *sociology academic word list* (*SAWL*). A sufficient number of research articles from eight sociology subfields (See Table 1) constituted the sociology corpus. The researchers consulted two university professors and the websites of some universities offering MA and PhD programs in sociology disciplines in order to identify sociology subfields and develop a list of subfields which appropriately represent sociology discipline. The research articles which were incorporated into the corpus were downloaded from some leading scholarly journals in sociology discipline; the articles were published between 2012 to 2022. The developed corpus, *Sociology Academic Corpus* (SAC), is composed of over 3.5 million running words and contains eight sub-corpora of identical sizes. Each sub-corpus is relevant to one of the above-mentioned sociology subfields and includes research articles from the subfield. Table 1 demonstrates the information regarding the size of the total corpus and each sub-corpus and the number of the articles in them.

Table 1.

Size of the Sub-Corpora, and Number of Research Articles

Sub-Corpora	Size	Number of Articles
Sociology of deviance	448,412	66
Cultural Sociology	457,430	59
Economic Sociology	440,043	62
Historical Sociology	472,521	63
Political Sociology	437,644	65
Sociology of education	445,660	64
Sociology of knowledge	419,767	60
Sociology of religion	422,516	67
Total	3,543,993	506

3.2 Corpus Analysis Software

In order to identify the high frequency words across sociology sub-fields and develop a sociology academic word list, the researchers employed the Range program, developed by Heatly et al. (2002), freely available at https://www.wgtn.ac.nz/lals/resources/ paul-nations-

resources/vocabulary-analysis-programs. The software lists all the words in a corpus alongside their frequency counts. The words are listed according to the magnitude of their frequency, high frequency words appearing at the top. Moreover, the program can juxtapose the frequencies of a word in several sub-corpora and provide information on frequency of the word in each sub-corpus, as well as in the whole corpus. Thus, researchers can easily scrutinize the frequency of each word in the total corpus and each individual sub-corpus and identify the words which are highly frequent in and evenly spread across the subfields of a discipline. The software version which has the BNC/COCA 25,000 word families as its baseword lists was employed in the present study. The software, initially, lists the words and words families of the 34 BNC/COCA baseword lists alongside their frequency in the analyzed corpus. Researchers can check the frequency and range of the words in each BNC/COCA baseword list and select the words which meet the determined frequency and range criteria. Then the words outside the BNC/COCA list are presented alongside their frequency counts in the corpus, which can be probed in order to find high frequency words beyond the BNC/COCA list.

In addition to the Range program, the researchers used another program (i.e., the Excel) to compare the developed sociology word list to the GSL-AWL and NGSL-NAWL lists in terms of their shared and exclusive words. The program demonstrated which words were shared by the sociology word list and a list combining general English and general academic words (i.e., the General Service List and Academic Word List or the New General Service List and New Academic Word List) and which words were highly frequent in sociology texts but absent in the other lists. The researchers so worked out how much the new list overlaps with the other lists.

3.3 Criteria for Word Selection

To choose the words to be entailed in the sociology academic word list (SAWL), the researchers stipulated three selection yardestics: frequency, range and word family. In many previous academic word list studies, a frequency of around 28 words per one million was considered the required frequency for word selection (e.g., Coxhead, 2000; Valipouri & Nassaji, 2013; Wang et al, 2008; Yang, 2015). As the size of the corpus in the current study was 3.5 million running words, an aggregate frequency of 100 was determined as the first required measure. The second criterion was range, which required the words to repeat at

least 10 times in five of the eight sub-corpora. In Coxhead's study, the words were required to occur 10 times in each of the four sections of the corpus. Similarly, Valipouri and Nassaji selected words which had a minimum frequency of ten in each of the four sub-disciplines of chemistry to include in their academic word list. Finally, the researchers employed word family as the counting unit for the development of the word list. The word families in the Range program are developed according to Bauer and Nation's (1993) level six, which includes inflectional suffixes and major derivational affixes. University students are mainly familiar with theses affixes and do not have much trouble in learning word families. Moreover, lists containing word families are expected to have a higher coverage of a corpus compared to a list containing the same number of word types or lemmas.

3.4. Data Collection and Analysis

The researchers followed several steps in the enhancement and assessment of the sociology academic word list. Initially, two university professors of sociology in Qom university and the websites of some universities offering MA and PhD programs in sociology were consulted in order to identify sociology sub-fields and select a representative list of the subfields. Eight sub-fields were specified as representing the sociology discipline (*sociology of deviance*, *cultural sociology*, *economic sociology*, *historical sociology*, *political sociology*, *sociology of education*, *sociology of knowledge*, and *sociology of religion*). Then some scholarly journals publishing research articles in the above-mentioned subfields of sociology were determined and a sufficient number of research articles published within 2012-2022 in these journals were download to be incorporated into the sociology corpus. The downloaded manuscripts were in various formats such as HTML, PDF, WORD, but since Range program processes only materials in TEXT format, the researchers converted all the files into TEXT format. Also, in order to include only sociology texts in the corpus, the *acknowledgements, affiliations, appendices, biodata* and *references* were removed from the research articles. The cleaned texts of the research papers were added and eight equally-sized sub-corpora (each containing around 450,000 running words) and a total corpus of over 3.5 million running words (the *Sociology Academic Corpus*) were developed. Finally, the corpus was analyzed by the Range program and the words which occurred frequently across the sub-fields of sociology were identified and incorporated into the sociology academic word list (SAWL). The word families in the 34 BNC/COCA baseword lists which had the

total frequency of 100 or above and the minimum frequency of 10 in at least five sub-corpora were identified and included in the SAWL. Then the high frequency words outside the BNC/COCA list were identified and added to the sociology academic word list.

4. Results

4.1 Development of the SAWL

The inquiry of the corpus and the sub-corpora exhibited that 1901 word families in the BNC/COCA list met the determined frequency and range criteria and were sufficiently frequent across the sociology subfields to be included in the SAWL. The majority of these words are general English vocabulary and are included in general service lists of words, such as 3000 BNC/COCA word list. Around 85 percent of the identified words were from the first three BNC/COCA baseword lists. Table 2 reveals information on the baseword lists which contributed the highest number of word families to the sociology word list and the number of words contributed by each baseword list.

Table 2.

BNC/COCA BaseWord Lists Contributing Most Words to SAWL

Baseword List	Number of Words in PAWL
Basewrd list 1	637
Basewrd list 2	414
Basewrd list 3	564
Basewrd list 4	173
Basewrd list 31	68
Basewrd list 5	59
Basewrd list 6	25

Moreover, eight words were identified which met the frequency and range criteria but were absent in the BNC/COCA list. They were: *so-called*, *decision-making*, *long-term*, *postcolonial*, *neoliberal*, *socio-economic*, *large-scale*, *well-known*. As it is evident, some of these words are general English words (e.g., *so-called, well-known, long-term*) and some seem to be relevant to or frequently used in sociology texts (e.g., *socio-economic*, *neoliberal*, *postcolonial*). Adding these words to the BNC/COCA list words which were highly frequent across sociology subfields, the researchers developed a list of 1909 word families for sociology discipline. . Over half of the words, namely 110 words, in the top 200 words of the list were grammatical words (e.g., *pronouns*, *articles*, *conjunctions*). Among the

remaining 90 words, which were lexical words, there were 15 words which seemed relevant to sociology fields. They included: *people*, *public*, *class*, *power*, *life*, *case*, *local*, *crime*, *subject*, *race*, *nature*, *interest*, *problem*, *control*, *position*. Table 3 displays the top 100 words in the SAWL. As it is shown, all of them belong to general service English vocabulary and some are general English words which are highly or partly related to sociology discipline and are expected to be found in sociology texts.

Table 3.

Top 100 Lexical Words in the SAWL

1- One	*35- View*	*69- Local*
2- Work	*36- Change*	*70- Major*
3- New	*37- Make*	*71- White*
4- Only	*38- Second*	*72- Paper*
5- State	*39- Number*	*73- Space*
6- Different	*40- Fact*	*74- Black*
7- Time	*41- Nature*	*75- Past*
8- Study	*42- Level*	*76- Issue*
9- World	*43- Become*	*77- Subject*
10- Power	*44- Question*	*78- Main*
11- See	*45- Crime*	*79- Perhaps*
12- Important	*46- Experience*	*80- Type*
13- Way	*47- Group*	*81- Hand*
14- Science	*48- Need*	*82- Young*
15- People	*49- Law*	*83- Term*
16- Particular	*50- Early*	*84- Kind*
17- Field	*51- Problem*	*85- Far*
18- History	*52- Position*	*86- Clear*
19- Class	*53- Recent*	*87- Long*
20- System	*54- Possible*	*88- Strong*
21- Life	*55- Present*	*89- Say*
22- Public	*56- Interest*	*90- Race*
23- Market	*57- Place*	*91- Book*
24- Use	*58- Certain*	*92- Large*
25- Human	*59- Set*	*93- Little*
26- Own	*60- Especially*	*94- Explain*
27- Party	*61- Understood*	*95- War*
28- School	*62- Control*	*96- Find*
29- General	*63- Take*	*97- Degree*
30- Like	*64- Course*	*98- Good*
31- Order	*65- Turn*	*99- Come*
32- Case	*66- Support*	*100-High*
33- Sense	*67- Idea*	
34- Point	*68- Key*	

4.3 Coverage of the SAWL versus GSL-AWL and NGSL-NAWL lists

After the improvement of the sociology word list, the researchers embarked on evaluating the list by comparing the size and coverage of the newly developed word list, the SAWL, over the sociology corpus, with those of the lists containing general service and general academic English words, GSL-AWL and NGSL-NAWL lists (i.e., the list containing the words in General Service List (West, 1953) and Academic Word List (Coxhead, 2000) and the list including New General Service List (Browne et al., 2013) and New Academic Word List Browne et al.(2013).

To compute the coverage of a word list over a corpus, the researchers divided the total frequency of the list by the number of the running words in the corpus (i.e., 3.5 million). Table 4 reveals the coverage of the word lists (i.e., SAWL, GSL-AWL, and NGSL-NAWL) over the Sociology Academic Corpus (SAC). As it is shown, the SAWL accounted for 87.4% of the running words in the corpus, and the coverage of the GSL-AWL and NGSL-NAWL were 84.19% and 85.1% respectively. The SAWL's coverage of the SAC was 3.21 percent higher than that of the GSL-AWL, though the former contained 646 fewer word families. Also, its coverage over the corpus was 2.19% higher than that of the NGSL-NAWL. The NGSL and NAWL are composed of lemmas instead of word families, therefore the size of the two lists (SAWL and NGSL-NAWL lists) are not quite comparable. The NGSL is composed of 2801 lemmas and the NAWL consists of 963 lemmas.

The coverage of the second 1000 GSL words over the sociology corpus was only 4.15%, which is a rather low coverage. Also, the first 1000 GSL words accounted for 68.69% of the running words in the SAC. As structure words are expected to cover over 50% of running words in a corpus (Kucera & Francis, 1967, as cited in Bowen et al, 1985), leaving out the coverage of structure words, the lexical words of the first 1000 GSL words would account for around 18% of the running words in the sociology corpus, which is not an enormous coverage. However, the AWL word families covered 11.35% of the tokens in the SAC, which is larger than its coverage over Coxhead's (2000) Academic Corpus (i.e., 10%).

The words in the NGSL had a rather high coverage of the sociology corpus (81.7%), which is much higher than that of the GSL (72.84%). However, the NAWL words covered only 3.4% of the running words in the SAC, which is much lower than AWL's coverage over the corpus (11.35%). As the GSL and AWL are composed of word families and NGSL and NAWL consist of lemmas, the sizes of the lists are not quite comparable. Nevertheless,

the SAWL demonstrated to have a much higher coverage over the sociology academic texts, despite its much smaller size.

Table 4.

Coverage of the SAWL and GSL-AWL over the SAC

Word List	Number of Word Families/Lemmas	Coverage of the SAC
SAWL	1910	87.4%
GSL-AWL	2556	84.19%
NGSL-NAWL	3764	85.21%
GSL	1986	72.84%
1st 1000 words	998	68.69%
2nd 1000 words	988	4.15%
NGSL	2801	81.7%
AWL	570	11.35%
NAWL	963	3.4%

4.2 Words common/specific to the SAWL and GSL-AWL

The final step in the study was to compare the words in the newly developed sociology word list (SAWL) and the GSL-AWL and NGSL-NAWL lists to ascertain how many words were shared by the SAWL and a list containing general English and general academic words and how many words were specific to the SAWL and not included in the other lists. Table 5 demonstrates the number of the word families shared by the SAWL and the other two lists. As displayed in Table 5, 1467 GSL-AWL word families were shared by the SAWL and 1089 GSL-AWL words were absent in the SAWL. This indicates that only 57.39% of the words in the GSL-AWL list are frequently used in sociology texts and 42.61% of the list's words are of low frequency in sociology. Over half of the GSL words (996 word families) were of less frequency in sociology texts and thus absent in the SAWL. However, 477 AWL words were frequently used in sociology corpus and were included in the SAWL. In fact, around 84% of the AWL words were shared by the SAWL. Furthermore, there were 443 word families which were highly frequent in sociology texts but absent in the GSL and AWL lists. That is, 443 word families in the SAWL (i.e., over 23% of the list) are not shared by the GSL-AWL list.

As for NGSL-NAWL list, 1446 lemmas in the list were highly frequent in the sociology corpus and included in the SAWL, but 2318 NGSL-NAWL words were of low frequency in sociology texts and absent in the SAWL. That is, less than 40% of the NGSL-NAWL words (i.e., 38.41%) were frequently used in sociology research articles and almost two-third of the list words are rarely used in sociology texts. Less than half of the NGSL lemmas (1258 lemmas, i.e., around 45% of the list) were highly frequent in the sociology texts and were included in the SAWL and less than one-fifth of the NAWL lemmas (188 lemmas) were shared by the SAWL. As the NGSL and NAWL lists are composed of lemmas rather than word families, comparing the size of the lists with that of the GSL-AWL and the SAWL is not straightforward and the fact that the lists have employed different word units must be considered; however, the number of word types in the NGSL-NAWL (10682 words) and the GSL-AWL (8766 words) are almost similar and the coverage of the lists over the corpus can be compared by each other. Finally, there were 464 SAWL words which were absent in the NGSL-NAWL list, that is, over 24% of the SAWL list was not covered by the NGSL-NAWL words.

Table 5.

Number of Words Common/Specific to the PAWL and GSL-AWL

Word List	Number of Word Families/lemmas	Number of Words absent in SAWL	Number of Words present in SAWL
General Service List	1986	996	990
Academic Word List	570	93	477
GSL-AWL	2556	1089	1467
New General Service List	2801	1258	1543
New Academic Word List	963	188	775
NGSL-NAWL	3764	1446	2318

Table 6 displays some example words for the low frequency GSL, AWL, NGSL and NAWL words in the sociology corpus (SAC) and the high-frequency words in sociology texts which are non-existent in the GSL-AWL and NGSL-NAWL lists. As the example words indicate, low frequency GSL-AWL, NGSL-NAWL words are expected to occur less frequently in sociology while high frequency non-GSL-AWL, non-NGSL-NAWL words are highly associated with sociology disciplines and are highly expected in sociology manuscripts.

Table 6.

Example Low Frequency GSL-AWL and High Frequency Non-GSL-AWL Words

Word families	Example words
Low frequency GSL words	*bucket, desk, loaf, juice, mat, mile, plough, shell, temple, waist*
Low frequency AWL words	*chemical, clause, estate, erode, export, levy, protocol, subsidy, tape*
Low frequency NGSL words	*clock, dust, flavor, gear, meat, opponent, paint, stupid, tale, wage*
Low frequency NAWL words	*abdominal, clay, cylinder, ion, jazz, kilometer, niche, rotate, weave*
High frequency Non-GSL-AWL words	*assert, collaborate, diffuse, exert, launch, obstacle, resemble*
High frequency Non-NGSL-NAWL words	*bias, cooperate, hostile, prestige, rank, universe, virtual*

5. Discussion

The purpose of the first research question, which was in line with the general purpose of the study, namely developing list of technical words for the field of Sociology, was to find words throughout the BNC/COCA word lists that were very common and equally distributed in the journal articles of different Sociology sub-disciplines. The computer analysis of the corpus consisting of eight sub-branches of Sociology RAs showed 17299 words involved in the BNC/COCA base word lists that were frequently used and well distributed in the articles of this field. In the next analysis that was done manually, 2005 words with the frequency and range required by the present investigation were selected from among the words which were previously obtained by RANGE software and by removing proper nouns and abbreviations from them, 1901 words were finally selected to be included in the target word list of this study. The difference between the number of selected words for the current word list and the total number of BNC/COCA words is a good indication that many words in the BNC/COCA base word lists did not occur frequently in Sociology research articles so learning all of the words in the BNC / COCA while time-consuming, is not fruitful in the specific field of Sociology.

The second research question aimed to find words that had the required frequency and range in the corpus, but were not present in the BNC/COCA base word lists, in response to which, the RANGE software outputted vocabulary outside the BNC/COCA, although lower in number compared to the number of Sociology technical words across the BNC/COCA,

their presence indicates that the BNC/COCA vocabulary does not necessarily meet all vocabulary needs of Sociology students since there still exist some words despite their high frequency and range in a specific academic field like Sociology, are not necessarily included in general service lists such as Nation's BNC/COCA wordlists, in other words, these two wordlists do not include all the words that are widely used in all academic disciplines. Moreover, it suggests that people who work and study in the field of Sociology, especially those who deal with research articles of this field, can benefit from a wordlist independent from BNC/COCA and specific to Sociology field. This word list includes words which occur most often in Sociology texts, so the probability of encountering them while studying or writing in the field of Sociology is much higher than all BNC/COCA words, hence learning the words of this newly developed list will be much more useful than those of BNC/COCA for Sociology students, researchers and other interested individuals in this field.

Regarding the answers to the first two research questions, the outcomes of this research confirm those of prior investigations (e.g., Coxhead et al. 2016; Hsu, 2018; Lu & Coxhead, 2020) which have been conducted with almost the same goal, stating that learning all the BNC/COCA words is neither necessary nor sufficient for success in studying and writing texts of various academic disciplines because not all BNC/COCA words are frequently utilized in different fields' texts. One of the studies whose findings are consistent with the interpretations presented for the present study was the research conducted by Coxhead et al. (2016), according to which many words in the BNC/COCA did not exist in the corpus of Carpentry vocabulary list. In this study, for vocabulary selection of the technical vocabulary list of Carpentry, the researchers first ran 25,000 BNC/COCA Nation lists on the Carpentry corpus to extract words that were present in the corpus but absent in the Nation lists and then classified the output vocabulary into word families, marginal words, abbreviations, proper nouns, and compound nouns. Following the analysis carried out in this study, Coxhead et al. found that only a limited number of words in BNC/COCA, such as *skillion*, *radiata* and *hardfill*, occurred more than 30 times in the corpus of carpentry academic texts, while a large number of words, occurred only few times, for example *scoria, cadastral* and *kahikatea*. Another study with almost the same findings of the current study is Hsu's (2018). In the mentioned study, for Traditional Chinese Medicine (TCM) wordlist to be developed, BNC/COCA words were put into three categories of high frequency, mid frequency and low frequency word families in English-medium traditional Chinese medicine textbooks which

resulted in identification of 605 high frequency BNC/COCA word families. For the study to be conducted, the researcher compiled a corpus of TCM textbooks with 13 million tokens and then measured the books vocabulary levels from within along the scale of word frequency in British National Corpus (BNC) as well as Corpus of Contemporary American English (COCA).

The third research question intended to compare two important characteristics, including size and coverage of the STWL to those of GSL-AWL lists. The size of a vocabulary list is one of its most important features because it plays a key role in efficiency of that wordlist. The smaller and more focused a word list is on a specific topic, the better and more accurate the knowledge and mastery it gives to its users. In the analysis that was done to answer third research question, it was exposed that the size of the Sociology vocabulary list is much smaller than the GSL-AWL wordlists. This list of words is free of useless items in the field of Sociology, learning which is considered waste of time for Sociology students and those interested in studying and researching in this field and even time-consuming in advancing their educational and research goals. Therefore, using STWL will be most effective for users by saving their time and energy. Another important feature of a high-quality word list is the high coverage of that list on the texts of a particular field compared to other lists. As a result of measuring and analyzing the lexical coverage of the present list, it was specified that the coverage of STWL over the corpus of Sociology is 3.21 larger than that of GSL-AWL lists. Therefore, using the current list as a source of learning technical vocabulary for people working in the field of Sociology will definitely be more appropriate than the GSL and AWL lists. Overall, according to the answer to the third research question, the recently developed wordlist, despite being small in size, covers well a wide range of articles in Sociology. It contains the largest number of known words of the field that are most used and evenly spread in Sociology academic texts, and for this reason, it can meet well the vocabulary needs of people who study, work or research in this field. The outcomes of this research regarding the size and coverage comparison of STWL to those of BNC/COCA are also in line with the results obtained in previous related studies as well as similar wordlists. (Jamalzadeh & Chalak, 2019; Wang et al. 2008; Yang, 2015).

6. Conclusion

The current research indicated that general words do not frequently occur in Sociology texts while there are many frequently-used words that often occur in Sociology research articles and Sociology pupils require to master in order to read their academic texts effectively, but are absent in BNC/COCA. In addition, as a result of the size and coverage comparison between STWL and two lists of GSL and AWL, it was found that STWL contains fewer words than GSL and AWL, while it covers a higher percentage of frequently used words in the texts of sociology journal articles.

The Sociology technical word list, consisting of words which in Sociology research articles, having a high frequency but also have a wide range, can be used to help both instructors and learners center on essential Sociology words. It can help ESP instructors discern which words to explicitly concentrate on in the Sociology classroom and also encourages them to have explicit vocabulary instruction in ESP classes rather than teaching the vocabularies in an incidental or unorganized way. Academic vocabulary identification in reading and utilization in writing is very helpful for academic success and the mastery of technical vocabulary accelerates and facilitates this success. Therefore, using the present technical word list, Sociology students can study Sociology technical vocabulary in a more conscious and manageable way, have enough word information to read academic RAs in their own major, comprehend the published academic English texts that they need to read in Sociology, master the concepts of this field and finally accomplish higher levels of vocabulary learning in their specialized fields. Providing some guidelines concerning vocabulary, the list of technical words of Sociology can also be used as a reference by materials developers for developing EAP materials. The word families in the list are also worth special attention when designing English for EAP courses. It is worth mentioning that STWL consists of base words, which means that by learning each of them, its users will also identify and comprehend the meaning of different forms of each base word. Knowledge of a base word can simplify the assimilation of its derivative or inflectional forms (Bauer & Nation, 1993). As mentioned in the previous sections, the language of many scientific research articles is English, but not necessarily all researchers interested in this field are native English speakers. The present technical word list for research articles in Sociology equips such researchers or EFL English learners who are into keeping further study in Sociology with the most frequent technical words in the Sociology RAs so that they can

expand their vocabulary size in a more effective and faster way to be able to read and comprehend Sociology research articles and also utilize this word list whenever they need to write in this field.

Needless to say, Sociology has many sub-disciplines, so similar studies with a larger number of articles or in a larger number of Sociology sub-disciplines can definitely better address the vocabulary needs of the people mentioned above. Moreover, for the subcorpus of the current study, important Sociology subfields were selected, including Sociology of deviance (Criminology), Cultural Sociology, Economic Sociology, Historical sociology, Political Sociology, Sociology of education, Sociology of knowledge, Sociology of religion. Other studies can consider other subfields for similar research. Furthermore, other studies, with almost the same purpose, can also develop the lists of technical or academic vocabulary of each of the abovementioned sub-disciplines or other sub-disciplines of Sociology by following the methodology and data analysis of this study. More detailed studies can also be done in the future that use different corpora such as books related to Sociology, transcribed spoken conversations or university lectures or a combination of them in order to make their corpus and subsequently their word list more comprehensive. In general, the language and in particular the English language is constantly changing due to the factors such as simplification, synchronic variation, semantic bleaching, and migration. Therefore, today's people consider the texts written in the past centuries to be ancient and their language archaic. They can easily recognize the difference between these texts and distinguish them from today's texts. Of course, these changes occur gradually in the language and over time these changes become more significant. It is clear that the lists of words extracted from old texts are affected by the mentioned changes, so it is recommended that the development of the wordlists be revised and updated every few years. Therefore, it is better to replicate the development of STWL with new texts and newly-written articles in the coming years.

References

Bagheri Nevisi, R., Safari, M., Hosseinpur, R. M., & Mousakazemi, R. S. (2023). A High Frequency Word List for Political Sciences. *Journal of Modern Research in English Language Studies*, *10*(4), 21-43. https://doi.org/10.30479/jmrels.2023.18213.2161

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

Biber, D. (1989). A typology of English texts. *Linguistics*, *27*(1), 3-43. http://doi.org/10.1515/ling-2013-0040

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use.* Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511804489.

Bowen, J. D., Madsen, H., & Hilferty, A. (1985). *TESOL techniques and procedures*. Newbury House.

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

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

Coxhead, A. & Demecheleer, M. (2018). Investigating the technical vocabulary of plumbing. *English for Specific Purposes, 51*, 84-97. https://doi.org/10.1016/j.esp.2018.03.006

Coxhead, A., Demecheleer, M., & McLaughlin, E. (2016). The technical vocabulary of Carpentry: Loads, lists and bearings. *TESOLANZ Journal*, *24*, 38-71.

Coxhead, A., & Hirsh, D. (2007). A pilot science word list for EAP. *Revue Française de Linguistique Appliqueé XII*, *2*, 65-78.

Dudley-Evans, T., & St. John, M. (2011). *Developments in English for Specific Purposes*: *A multi-disciplinary approach*. Cambridge University Press.

Farrell, P. (1990). Vocabulary in ESP: A lexical analysis of the English for electronics and a study of semi-technical vocabulary. CLCS Occasional Paper No. 25. Trinity College, Dublin. Centre for Language and Communication Studies.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305-327. https://doi.org/10.1093/applin/amt015

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). Range computer program.

Heidari, F., Jalilifar, A., & Salimi, A. (2020). Developing a corpus-based word list in pharmacy research articles: A focus on academic culture. *International Journal of Society, Culture, & Language*, *8*(1), 1-16.

Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. *Language Teaching Research*, *17*(4), 454-484. https://doi.org/10.1177/1362168813494121

Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks. *English for Specific Purposes, 33*, 54-65. https://doi.org/10.1016/j.esp.2013.07.001

Hsu, W. (2018). The most frequent BNC/COCA mid- and low-frequency word families in English-medium traditional Chinese medicine (TCM) textbooks. *English for Specific Purposes, 51*, 98-110. https://doi.org/ 10.1016/j.esp.2018.04.001

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

Jamalzadeh, M. & Chalak, A. (2019). A corpus-based study of academic vocabulary in physiotherapy research articles. *Language Teaching Research Quarterly*, *9*, 69-82.

Kanzaki, M. (2019). Making a scientific research article word List. *Studies in Linguistics and Language Teaching*, *30*, 73-98.

Lu, Y., & Coxhead, A. (2020). A corpus-based study of vocabulary in the new concept English text book series. *RELC Journal*, *53*(3), 597-611. https://doi.org/10.1177/0033688220964162

Martinez, I. A., Beck, S. & Panza, C.B (2009). Academic vocabulary in Agriculture: A corpus-based study. *English for Specific Purposes, 28*(3), 183-198. https://doi.org/10.1016/j.esp.2009.04.003

Munoz, V. (2015). The vocabulary of agriculture semi-popularization articles in English: A corpus-based study. *English for Specific Purposes, 39*, 26-44. https://doi.org/10.1016/j.esp.2015.04.001

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

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

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

Safari, M. (2019). English vocabulary for equine veterans: How different from GSL and AWL words. *Iranian*

Journal of English for Academic Purposes, 8(2), 51-65.

Tongpoon-Patanasorn, A. (2018). Developing a frequent technical words list for finance: A hybrid approach. *English for Specific Purposes, 51*, 45-54. https://doi.org/10.1016/j.esp.2018.03.002

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

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes, 27*(4), 442-458. https://doi.org/10.1016/j.esp.2008.05.00310.

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

Watson-Todd, R. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes, 45*, 31-39. https://doi.org/10.1016/j.esp.2016.08.003

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

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