RICHTMANN
PUBLISHING

A Corpus-Driven Study of Terms Used to Refer to Articles and Methods in Research Abstracts in the Fields of Economics, Education, English Literature, Nursing, and Political Science

Riyad F. Hussein

Department of English Language and Translation
Applied Science Private University, Amman, 11192, Jordan

Ahmad S. Haider

Department of English Language and Translation
Applied Science Private University, Amman, 11192, Jordan

Sa'ida Walid Al-Sayyed

Department of English Language and Literature
Al-Ahliyya Amman University, Amman, 19328, Jordan

Abstract

The present study attempts to identify the most frequent terms that are used in research abstracts relating to research articles designations, research methods, and research goals and quantify them in various disciplines, namely, economics, education, English literature, nursing, and political science to see whether there is a unique pattern characterizing terms peculiar to each of the five disciplines under investigation. It also explains why specific terms are overused/ underused by researchers. The current study follows a corpus-driven approach. For this purpose, we compiled a corpus of 2500 research abstracts from online refereed journals in the fields mentioned above. The corpus linguistic software program, AntConc (3.5.8), was used to analyze the collected data. The analysis revealed that some terms are more frequently used in some areas than others. For example, the term 'study' was the most common word to designate academic research. The most frequent term to refer to population and subject-related terms was 'sample/s,' while the least frequent was 'interviewee/s.' The words used to designate tools or instruments varied, with 'test' being the most frequent and 'checklist/s' the least. This study is of significant benefit for researchers in various disciplines. It acquaints them with terms used to designate articles in their respective fields, in addition to terms used most frequently to refer to sample- related terms and finally to words used for setting goals such as objectives, aims, and goals. This, in turn, can help researchers and graduate students embarking on writing their theses to opt for the most relevant terms peculiar to their disciplines. Unlike most studies that focused on developing academic word lists (AWL), this study set off with terms previously established and used in research bodies and research abstracts to unveil their popularity and the extent to which they are used in various discipline abstracts.

Keywords: research abstracts, terminologies, frequency, AntConc, academic disciplines, corpus

1. Introduction

The use of collections of text in language study dates back to the thirteenth century when some exercises in concordance compilation were initiated by Hugh of St. Cher "who employed 500 monks to list almost every word in the Bible with the points where it was used" (Aston, 2011, p. 1). The study of the Bible's linguistic aspects is still common as many linguists and theologians have been engaged in word and idiom counts used in the Bible (Pinnavaia, 2012). Some other scholars attempted to unveil word occurrences and word frequencies for every book of the Bible and showed their rankings in terms of the number of words, which were 33,002 in Jeremiah; 30,147 in Psalms; 25,975 in Exodus; 23,008 in Deuteronomy and John was last in the list with 219 words. Besides, they were interested in the number of times a word appears in the Bible.

Despite the diverse interests of scholars and linguists and the multitude of problems facing them in the Bible analytical study, scholars were all utilizing the methodology intrinsic to corpus linguistics. In other words, they were earlier engaged in traditional word and idiom counts, and later in electronic counts of words. Even if they were not familiar with the term 'corpus linguistics,' much of their work was similar to the corpus analysis that researchers do nowadays with one exception - they did not use computers.

According to Biber, Conrad, and Reppen (1998), two main perspectives can be followed to study a language. First, investigating language structure and how small linguistic units can be put together to form larger ones. Second, examining how language is used in natural ways and actual settings. Corpus linguistics as a way of studying language is classified under the second category. Corpus is "a collection of naturally occurring examples of language, consisting of anything from a few sentences to a set of written texts or tape recordings, which have been collected for linguistic study" (Hunston, 2002, p. 2). According to McEnery and Wilson (2001, p. 1), corpus linguistics is "the study of language based on examples of 'real-life' language use." Advances in technology can be rightly considered the stimulating factor in the development of corpus linguistics, namely computers, software programs and interfaces, and scanners' invention. Utilizing the then-new computer technology, it was possible to build and create small corpora. These include The Brown Corpus of Standard American English (One million words) in 1961, The (LOB), Lancaster-Oslo-Bergen Corpus (British English), the Kolhapur Corpus (Indian English), among others. Subsequent to advances in the computer technology and consolidation of cooperation between linguists and computer and software experts, corpora of tens of millions of words, and even more such as the British National Corpus (BNC) and Corpus of Contemporary American English (COCA) were created.

This study adopts a corpus-driven approach (see Esfandiari & Barbary, 2017; Lei & Liu, 2018), in the sense that the data is approached with no prior assumptions (Tognini-Bonelli, 2001). Tognini-Bonelli (2001, p. 86) argues that "the unexpectedness of the findings derived from corpus evidence leads to the conclusion that intuition is not comprehensively reliable as a source of information about language." In the present study, the corpus linguistic technique of frequency is mainly used to generate a list of all words in the corpus under study, and then compare and contrast the results with the frequency lists of other corpora and sub-corpora (Haider, 2017; Haider, 2019b; Haider, 2019c).

While most researchers were mainly concerned with establishing academic word lists (AWLs) in various disciplines, including Health Sciences (Ohashi, Katagiri, Oka, & Hanada, 2020; Shirzadi & Dolatabadi; Yang, 2017), Social Sciences (Chanasattru & Tangkiengsirisin, 2016; Kwary & Artha, 2017), Linguistics (Antes & Beck, 2020; Moini & Islamizadeh, 2016), language learning (Choo, Lin, Singh, & Ganapathy, 2017; Durrant, 2016; Green, 2019; Nelson & Albakry, 2020), and Engineering (Veenstra & Sato, 2018), the current research focuses on previously established terms and their variation in use in various disciplines research abstracts.

This study attempts to quantify a discipline-specific terms used most frequently to refer to *articles*, *methods*, and *goals* in research abstracts. More specifically, it seeks to answer the following questions:
1. What are the most frequent terms used to refer to academic research articles across various disciplines, namely, economics, education, English literature, nursing, and political science?
2. What are the most frequent terms used to refer to research methods in research abstracts

across the disciplines mentioned above?

3. What classes of words are used to express the goals and aims in the research abstracts across the disciplines mentioned above?

2. Literature Review

Corpus linguistics dates back to the time when linguists decided to study the language as an empirical science rather than a theoretical one. Some techniques used in corpus linguistics are older than computers and go back to the late 18th and 19th centuries when lexicographers collected samples of language from daily life situations to determine the accurate meaning of words. Before the emergence of computers, building large corpora was challenging and time-consuming. Pre-electronic corpora had to be analyzed manually, and this took a lot of time. For example, the Bible was used as a corpus from early on; wordlists and concordance lines were compiled manually to show that the various parts of the Bible are consistent with each other. Corpus-assisted studies were popular even before the term "Corpus Linguistics" was coined, and they included different disciplines such as language acquisition, spelling conventions, language pedagogy, comparative linguistics, and semantics.

After the advent and evolution of computers, electronic corpora appeared in the 1960s, such as The Brown Corpus, which is considered the first work utilizing computers in the modern era. Building massive corpora was difficult for humans to achieve in the past, as it required much work and effort in the act of compilation. Generating frequency lists and concordance lines was not easy until the 70s and 80s when computers became more powerful and user-friendly, and cheaper than before. As a result, large collections of texts were compiled and analyzed electronically for different purposes. In that period, the interest in studying the language in use with computers' help was dramatically increased. Electronic corpora provided researchers with an unprecedented way to analyze language. To sum up, after the fast spread of computers and the availability of electronic texts, it was possible to collect vast amounts of data in a short time. Researchers could also use some available statistical methods and even develop and create new tools and techniques to analyze their data linguistically.

In the last 50 years, cooperation between linguists and computer scientists, and the development of computers in terms of their memory storage and data retrieval and language analysis facilitated the emergence of a revolution of corpus linguistics. Anthony (2009) argued that "the huge advances in computer technology over the last 50 years have allowed us to store far more data than could have been conceived in the 1900s" (P.8). Tognini-Bonelli (2010) pointed out that corpus building and availability fall into three main stages; 1960-1980, which witnessed the learning to build and maintain corpora of up to a million words. 1980-1990 the period where scanning twenty million words became a reality. 1990-2000, the first serendipity when the text became available as the by-product of computer typesetting and finally the new millennium. The second serendipity is when the text that never had existence as hard copy became available in unlimited quantities from the internet.

In line with the above, scholars and linguists from diverse disciplines became more interested in applying corpus linguistics to language study (Al-Abbas & Haider, 2020; Biber et al., 1998; Haider, 2019a; Haider & Hussein, 2020; Haider & Olimy, 2018, 2019; Haider, Olimy, & Al-Abbas, 2021; Sayyed, 2019; Stubbs, 1996). An overwhelming number of research studies and hundreds of conferences and conventions and symposia on corpus related topics were convened, and their results and recommendations disseminated through periodicals, publications, and books. Areas of applied linguistics that have benefited most from corpus linguistics' insights and analytical techniques include lexicographic studies, discourse analysis, diachronic study, semantics, stylistics, language development, and others (Laviosa, 2012).

Corpus linguistic work shares many characteristics: the empirical analysis of texts, the use of a principled collection of texts, and computer-assisted analysis techniques. It also incorporates quantitative analysis, with more qualitative, functional interpretations of language use. The corpus approach is characterized by being empirical where actual patterns of language use in natural texts are analyzed, utilizing a large collection of natural texts as the basis for analysis, extensively utilizing

computer and computer software for analysis, and finally combining both quantitative and qualitative analytical techniques (Biber et al., 1998).

In general, researchers were concerned with establishing AWL for learning and teaching purposes or ESP research writing. Gilmore and Millar (2018) followed a corpus-based approach to examine an 8-million-word Civil Engineering Research Articles corpus using a keyword and cluster analyses and find out the words relating to civil engineering research. The researcher argued that this will be helpful for civil engineering students and instructors as it provides a principled approach to ESP course design. Liu and Han (2015) established the first academic word list for environmental science and suggested a set of compilation criteria for field-specific academic word lists.

Yang (2015) explored the most frequently-used words in nursing sub-disciplines and recommended generating more field-specific AWLs for nursing students, which will help improve their academic performance, especially in reading and writing. Todd (2017) highlighted the importance of academic wordlists and recommended instructors to generate a shortlist of words based on opacity by identifying "those words for which the learners would gain the greatest benefit from a teacher's help, since these are the words learners are most likely to have problems dealing with autonomously" (P. 38). Similarly, Pournia (2019) generated a list of the most frequent words in 13 high-impact factor English nursing journals, pointing out that the list can help nursing students understand and comprehend academic research articles' content without consulting many other resources. Heidari, Jalilifar, and Salimi (2020) used a corpus of about 3.5 million words of pharmacy texts to generate a Pharmacy Academic Word List and highlighted the importance of compiling domain-specific academic word lists due to their considerable significance for non-native researchers and postgraduate students.

Wang (2017) compiled a corpus of art research articles collected from 16 journals. The researcher examined the most commonly used words in the corpus to help students majoring in art improve their English and academic reading and writing. The study recommended establishing more discipline-specific word lists as they are essential to students, language educators, and material designers.

Tongpoon-Patanasorn (2018) adopted a combined method, namely, a keyword analysis and a modified rating scale, to develop a list of frequent technical words in finance to help students and teachers learning/teaching English for specific purposes. The generated list, according to the author, is also useful for curriculum designers and material developers. Csomay and Prades (2018) examined the use of academic vocabulary in the writings of university students whose mother tongue is not English and found that there are "significant relationships between academic vocabulary use and essay scores in some text-types, and differences in the way academic vocabulary is used across text-types and levels of instruction, independent of the drafts" (P. 100).

Jemadi and Iku (2019) investigated the specific academic vocabulary terms in 97 English research articles abstracts written by non-native English scholars to broaden academic vocabulary comprehension in English contexts by speakers of other languages. Kwary (2018) compiled a 5.6 million words corpus of journal articles published between 2011 and 2015 and designed a concordance to help researchers generate frequency lists and explore words in their co-texts (KWIC). Such a corpus is important for researchers, teachers, and translators working on academic English.

As pointed out in the introduction, this study attempts to analyze and quantify the words used to designate research (*article, paper, research, stud*) . Besides, it attempts to explore the most frequent words used in the research methods such as *survey, questionnaire, test/s, method, hypothesis, experiment(s), instruments, participants, population, subjects, sample, corpus, respondents, informants, qualitative, quantitativ*e, and *empirical, procedures.* This study also examines the formulation of objectives or goals and the classes of words used in this formulation. It explores whether goal-related words are expressed more in nouns such as 'the goals/objectives of the study are' or verbs such as 'aim to or aim at.'

3. Methodology

This part explores the research method and design. It discusses how this study corpus is compiled, the sample size, and the software used to analyze the data.

3.1 The Corpus

The corpus of the current study contains 2500 research abstracts. It comprises five sub-corpora covering article abstracts in various disciplines: economics, education, English literature, nursing, and political science. More specifically, 500 abstracts from each field were retrieved from online refereed journals.

The economics discipline corpus was retrieved from two journals, namely *Economics*, vol. 1,2007 to Vol. 10, 2016, and *The Economic Journal* vol. 124 issue 581 to vol. 126, issue 595, 2016. The education sub-corpus was compiled from the abstracts of the *American Journal of Educational Research* vol.2, issue seven through vol. 4 issue 17, 2016. The English literature sub-corpus was collected from four different journals, namely, *Asiatic: IIUM Journal of English Language and Literature* vol.1, 2007 through vol.10, 2016, *International Journal of English and Literature* vol. 1, 2010 through vol. 7, 2016; *The English Literature Journal* vol. 1, 2014 thru vol. 3, 2016; *The International Journal on Studies in English Language and Literature* 2016. The nursing sub-corpus was extracted from *Advances in Nursing Science* vol. 25, 2002 through vol. 39 2016. Finally, the political science sub-corpus was retrieved from an online journal, entitled *British Journal of Political Science* vol.33, 2003 through vol. 46, 2016.

3.2 Software

Using the proper software or interfaces, linguists can identify and recognize the patterns of language use and how they are used and the factors responsible for their variability. In this study, AntConc (3.5.8), created by Laurence Anthony of Waseda University (Anthony, 2019), was used to analyze the compiled corpus to identify language use patterns. As indicated by Römer and Wulff (2010, p. 103):

> *"The software is free to download from the author's homepage. There are versions for different platforms available (Windows, Macintosh, Linux). AntConc is sporadically updated. Information about the program and its tools can be found in the Readme file on Anthony's website. AntConc does not require any installation on your computer but can be launched by simply double-clicking on the executable file".*

4. Findings

In this section, results related to the terms used to designate academic research in various disciplines, the words used to refer to subjects and instruments, and finally, the terms used express research goals in research abstracts are reported and discussed.

4.1 Results Related to Question 1

The researchers carried out a frequency analysis using AntConc to determine the most frequently used terms to designate research (Table 1).

Table 1. Terms used to designate academic research across various discipline abstracts

Item	Economics	Education	Literature	Nursing	Political	Total
study	90	763	187	232	129	1401
research	54	443	66	379	116	1058
article	49	66	123	282	383	903
paper	261	177	370	14	9	831
project	4	95	14	19	11	143
essay	2	3	66	6	1	78
report	9	25	5	17	17	73

Table 1 clearly shows that the most frequent word used to designate academic economics research is the word *paper* with a frequency of 261, followed by *study* 90 and *research* 54. The table also shows that the least frequent words used are *project* 4 and *essay* 2. It is clear that a tendency exists for the use of the

words *paper* and *study* more often to designate research in the field of economics, while the words *article* and *research* are used with less frequency. The frequent use of the term *paper* may be due to humbleness on the part of economic experts or their focus on research procedures rather than terms and terminology. On the other hand, the use of the term *study* may indicate the empirical nature of economic research.

Table 1 also shows that the most frequent words used to designate research in education is *study* with a frequency of 763, followed by *research* with a frequency of 443 and *paper* with a frequency of 177. The terms *report* 25 and *essay* 3 were the least used. The use of *study* in education abstracts can serve as an umbrella or generic term for other designations. As to the frequent use of *research,* this may be attributed to the fact that researchers in education wish to emphasize that their research is as scientific, empirical, and valuable as in other disciplines.

Again Table 1 shows that the most frequent word used to designate academic research in English literature is *paper* with a frequency of 370, followed by *study* with a frequency of 187 and *article* with a frequency of 123, while the frequency of the words *essay* and *research* was 66 for each. Table 1 also shows that the least frequent words used are *project* 14 *report* 5. Therefore, English literature seems to use the word paper to designate research. This may be due to literary research theoretical nature based on theoretical speculation, logical reasoning, and inferencing and hardly ever uses population, samples respondents, questionnaires, observation, or checklists.

The most frequent term used to designate academic research in nursing is *research* with a frequency of 379, followed by *article* 282, and *study* with a frequency of 232. The least frequent words used to designate nursing research are *report* 17, *paper*14, and *essay* 6. The use of the word *research* in nursing research abstracts may be due to the empirical nature of research that requires *sample, population*, *quantitative*, and *qualitative* methods, and the use of *article* may be associated more with positive connotations of the word in comparison with other terms.

The most frequent term used to designate academic research in political science is *article* with a frequency of 383, followed by *study* 129 and *research* with a frequency of 116. The least frequent words used are *project* 11, *paper* 9, and *essay* 1. The use of the term *article* here as with nursing may be associated more with positive connotations of the word than other terms. Furthermore, the use of *study* can serve as a cover term that may subsume other terms such as *project*, *essay,* and *report*.

Table 1 clearly shows that the most frequent term used to designate academic research across the abstracts of various disciplines is *study* with a frequency of 1401, followed by *research* with a frequency of 1058, then *article* with a frequency of 933, and *paper* 831. The terms *essay* ranked sixth with a frequency of 78, and *report* ranked last with a frequency of 73. The use of *essay* was highest in the English literature abstracts with a frequency of 66, whereas its frequency was very low in the other abstracts. The relatively high use of *essay* in English literature may be due to the popularity and standing of this term or the fact that some research papers look similar to essay forms.

4.2 Results Related to Question 2

The frequency tool in AntConc was again used to identify the most frequent sample-related words in the five discipline abstracts, as shown in Table 2.

Table 2. Frequency of sample-related words in different research abstracts

Item	Economics	Education	Literature	Nursing	Political	Total
population	33	34	2	37	13	119
Sample/s	35	125	3	21	21	205
Respondent/s	2	49	0	4	11	66
Informant/s	0	11	0	1	0	12
Participant/s	7	64	22	57	20	170
Subject/s	12	30	6	7	3	58
Interviewee/s	0	2	0	0	0	2
Focus group/s	0	17	2	10	1	30

Concerning subject-related words, Table 2 clearly shows that the most frequently used word in economics is *sample*/s with a frequency of 35, followed by *population*/s with a frequency of 33. The least frequently used words are *participant*/s and *respondent*/s with a frequency of 7 and 2, respectively. The words *informant*/s, *interviewee*/s, and *focus group*/s were missing. Table 2 also shows that the most frequently used words in education are *sample*/s with a frequency of 125, followed by *participant*/s with a frequency of 64. The words *population*/s and *subject*/s were less frequently used with a frequency of 34 and 30, respectively. The words *informant*/s 11 and *interviewee*/s 2 were of very low frequency. Table 2 also shows that all sample-related words were used in education abstracts but with variable frequencies.

Table 2 also shows that the most frequently used word in English literature is *participant*/s with a frequency of 22, followed by *subject*/s with a frequency of 6. The least frequently used words are *sample*/s, *population*/s, and *focus group*/s with a frequency of either 3 or less for each word. Concerning nursing, the most frequently used word is *participant*/s with a frequency of 57, followed by *population*/s 37 and the least frequently used words are *respondent*/s 4 and *informant* only 1. The word *interviewee*/s was entirely missing in research abstracts in nursing.

Finally, the most frequent words used in political science are *sample*/s with a frequency of 21 and *participant*/s 20. The least frequently used words are *subject*/s 3 and *focus group*/s only I. The words *informant*/s and *interviewee*/s were entirely missing in political science research abstracts.

If all discipline abstracts were taken together, then the most frequently used words are *sample*/s with a frequency of 205, followed by *participant*/s with a frequency of 170 and *population*/s 119. The least frequently used words in all discipline abstracts are *informant*/s 12 and *interviewee*/s 2.

Words related to the category of tools or instruments used in research abstracts of the five disciplines are shown in Table 3.

Table 3. Terms used to designate research instruments across different research abstracts

Item	Economics	Education	Literature	Nursing	Political	Total
Test/s	62	204	6	13	81	366
Survey/s	28	116	5	20	76	245
questionnaire	0	1 29	3	6	0	138
Experiment/s	21	29	9	4	23	86
Observation	4	38	6	4	3	55
Corpus	1	3	6	3	1	14
Checklist/s	0	5	1	2	0	8
total	116	524	36	52	184	912

The most frequently used word in economics abstracts is *test*/s with a frequency of 62, followed by the word *survey* with a frequency of 28, and the least frequently used words are *observation* 4 and *corpus* only 1. The words *questionnaire* and *checklist* were entirely missing in economics abstracts. Table 3 clearly shows that the most frequent instrument used in education is *test* with a frequency of 204, followed by *questionnaire* with a frequency of 129 and *survey* 116. The least frequently used instruments are *checklist* and *corpus* with a frequency of 5 and 3, respectively. Unlike economics abstracts, all tools or instruments were used with frequencies ranging from 204 to 3 in education.

Table 3 also shows that the most frequently used instrument in English literature is *experiment* with a frequency of 9, followed by *test*, *observation*, and *corpus* with a frequency of 6 for each. The tools *survey*, *questionnaire*, and *checklist* with a frequency of 5, 3, and 1 respectively were the least frequently used tools in English Literature. The total occurrences of instruments or tools were 36 in English Literature, which is the lowest compared to their occurrence in other discipline abstracts.

Concerning nursing, the most frequently used tool is *survey* with a frequency of 20, followed by *test* 13. The frequency of the remaining tools ranged from 6 to 2. In nursing, the total occurrence of instruments or tools was 52, which is the second-lowest after English literature compared with its

occurrence in other discipline abstracts.

Table 3 clearly shows that *test/s* is the most frequently used instrument in research abstracts across the various disciplines with a frequency of 366, followed by *survey/s* with a frequency of 245, then *questionnaires* 138. The research tools *observation*, *corpus,* and *checklist* ranked last with a frequency of 55, 14, and 8, respectively. The low frequency of the word *corpus* can perhaps be attributed to researchers' unfamiliarity with corpora and interfaces or programs used to search corpora. The term *checklist/s*, on the other hand, is believed to be unreliable for research.

If all research instruments were taken together, then it can be seen that education abstracts have made most use of these instruments with a frequency of 524, followed by political science 184, economics 116, nursing 52, and finally, English Literature 36.

4.3   Results Related to Question 3

To answer the 3rd research question pertaining to the classes of words used to state the research goals and aims, AntConc was used to identify language use patterns and frequencies (see Table 4 and 5).

Table 4. Nouns used to express research goals or objectives in different research abstracts

Item	Economics	Education	Literature	Nursing	Political	Total
Objective/s	16	61	15	8	10	110
Goal/s	4	59	13	28	10	114
Aim/s	13	60	25	26	5	129
Total	33	180	53	62	25	353

Table 4 shows that 353 nouns are used to express research goals or objectives. The word *aim(s)* ranked first with a frequency of 129, followed by *goal/s* 114, and finally the word *objective/s* with a frequency of 110. In economics, the word *objective/s* was used with a frequency of 16 while *aim/s* was used with a frequency of 13, which is nearly close to the use of *objective/s* and thus ranked second. The word *goal/s* was used only four times and thus ranked third. In education, the word *objective/s* was used with a frequency of 61 while *aim/s* was used with a frequency of 60, which is very close to the use of *objective/s* and thus ranked second. The word *goal/s* frequency was 59 and thus ranked third. In English literature, the most frequent word was *aim/s* 25, followed by *objective/s* 15, and finally *goal/s* 13. In nursing, the frequency of the word *goal/s* was 28, followed by *aim/s* 26 and finally *objective/s* 8. In political science, the most frequent words are *objective/s* 10 and *goal/s* 10, followed by *aim/s* 5, which ranked third. Research abstracts in nursing ranked second in using and stating goal-related words 62 with the word *goals* ranking first. English literature abstracts ranked third 53, economics fourth 33, and political science 25, ranked fifth. It should be stated at this point that no researcher, whether in English literature, economics, or political sciences, embarks on conducting research or investigating research problems without formerly conceptualizing or identifying the research problem but may overlook using or stating goals or objectives explicitly. This may be attributed to the focus on the empirical research type, which is prevalent in these disciplines.

A frequency list was generated to identify the most frequently used verbs to describe goals and objectives (Table 5).

Table 5. Verb forms used to express research goals or objectives in research abstracts

Item	Economics	Education	Literature	Nursing	Political	Total
Aims to / Aimed to	3	47	33	4	4	91
Aims at / Aimed at	7	27	23	6	1	64
Aiming at	2	4	0	1	0	7
Total	12	78	56	11	5	162

Concerning the use of the verbs (*aim/s to, aim/s at, aimed at, aimed to,* and *aiming at*) to express research objectives, education abstracts ranked first with a frequency of 78, followed by English literature 56, and economics 12. Nursing 11, and political science 5 ranked fourth and fifth, respectively. However, when taking both nouns and verbs which express research objectives, again education abstracts ranked first (258), followed by English literature (108), then nursing (73), economics (45), and finally, political science (30). From Tables 4 and 5, it is evident that economics and political science researchers do not tend to use the terminology related to goals and objectives explicitly. Accordingly, economics ranked fourth, and political science ranked fifth with regard to using goal-related terms.

5. Discussion

In the research on academic word lists, there seem to be three main approaches: academic word lists for use in learning and teaching English for special purposes. In this case, the establishment of these lists is for pedagogical purposes (see It-ngam & Phoocharoensil, 2019; McDonough, Neumann, & Hubert-Smith, 2018). The second approach is linked to knowledge and skill of using and compiling corpora and the way to navigate them (Kwary, 2018; Talalakina, Stukal, & Kamrotov, 2020). Researchers in this group are familiar with different types of corpora and the way to manipulate them as well as program software and interfaces and do not necessarily have language interests. The third approach includes researchers and scholars concerned not with developing academic words lists for English in general, but rather for English in specific fields such as medicine, pharmacy, engineering, art, environment, and business to name some (see Heidari et al., 2020; Todd, 2017; Tongpoon-Patanasorn, 2018). AWL are necessary in these fields because while some graduate students or researchers are competent in subject matter, they may lack the academic words in their respective fields. Thus, exposure to these lists can acquaint them with the most frequent words and consequently enable them to write research and research findings and understand publications in their respective fields.

This research adopts a corpus-driven approach, which means that it is empirical because it analyzes the actual patterns of use and uses a relatively large corpus, which consisted of five sub-corpora, economics, education, English, nursing, and political research abstracts. Besides, it has made use of computers to create the corpus of research abstracts and then analyze the data to obtain frequencies of occurrence of the terms under investigation (Bennett, 2010). Following a corpus-assisted approach, researchers generally utilize quantitative and qualitative techniques. The frequencies and figures generated from the corpus are analyzed qualitatively. So in corpus-assisted approach research, it is not sufficient to come up with numbers and figures without subjecting them to some explanation or interpretation.

In line with this, it was found that the most frequent word used to designate research in all research abstracts is *study*, followed by *research*, next *article*, then *paper*, and finally *essay* and *report*. The use of *essay* was highest in the English literature abstracts, whereas in the other abstracts, its frequency was very low. The relatively high use of the word *essay* in English may be due to the popularity and standing of this term or to the fact that some research papers in English literature look similar to essay forms. As to the frequent use of *paper* in economic abstracts, it may be due to the researchers' humbleness or to their focus on research procedures rather than terms and terminology.

The use of *study,* which was highest across all discipline abstracts, may serve as an umbrella or generic term for other designations; it may also indicate the empirical nature of research. The second most frequent word *research* may be indicative of the fact that researchers wish to emphasize that their research is as scientific, empirical, exact, and rigorous as the case in natural sciences

Concerning the sample-related words, the most frequently used one in all research abstracts is *sample/s*, followed by *participant/s* and *population/s*. The least frequently used terms are *informant/s* and *interviewee/s.* The high frequency of the word *sample* did not come as a surprise because most research in education, nursing, and economics is applied and empirical and heavily utilizes surveys and questionnaires to be administered to samples of participants. This also explains the high frequency of the terms *population* and *participant.* The frequency of the word *informant* was very low across the

board, and this perhaps is due to the loaded and unpleasant associations and negative connotation it has acquired through its meaning as someone who gives information to the police. Likewise, the frequency of the term *interviewee* was very low, and this perhaps is due to the fact the interview technique requires the use of interviewees, which is time-consuming, especially if it is non-structured and if the number of interviewees is very large.

As to the category of research instruments, the most frequently used terms are *test*, followed by *surveys* and *questionnaire*. The least frequently used words are *observation, corpus,* and *checklist.* The term *checklist* was totally missing in economics and political science abstracts. The high frequency of the terms *test*, *survey,* and *questionnaire* was no wonder as different types of tests are used, such as multiple-choice tests and cloze tests. Surprisingly, they were used in all research abstracts, even in English literature, albeit with a low frequency in the latter. Concerning s*urvey* and *questionnaire*, they were sometimes used interchangeably as some researchers refer to survey to mean questionnaire and vice versa. It is unusual that some researchers even use the compound *survey questionnaire* to refer to either instrument. If the two words were taken as one unit, then its frequency far exceeds that of the term *test*. The low frequency of the word *corpus* can perhaps be attributed to the unfamiliarity of some researchers with corpora and interfaces or software used to query corpora. On the other hand, *checklists* and *observation* are believed to be unreliable as research instruments, especially if the observation technique is not accompanied by a checklist, which perhaps explains their low frequency in research abstracts across all disciplines.

Concerning the classes of words used to state research goals and aims, it was shown that the most frequently used word is *aim(s),* followed by *goal/s* and finally the word *objective/s*. The results have shown that researchers in education explicitly use goal-oriented terms more than researchers in other disciplines. This may point to the fact that researchers in education are more conscious of goals and objectives due to the educational conventions and research paradigms, which gives precedence and priority to identifying and setting goals and objectives before conducting research. In education, researchers are conscious of goals and goal-oriented terms and terms related to methods and procedures such as sampling, subjects, instruments, and population. In using goal-related words nursing abstracts ranked second, English literature ranked third, economics fourth, and political science fifth. The most frequently used word in economics, education, and political science is objective/s, whereas the most frequently used words in English and nursing is aim/s and goal/s respectively. Economics and political science abstracts were the least to use noun categories to state goals and it was expected that verb categories aims/at, aims/to stating goals would be higher, but contrary to expectation economics and political science ranked last. It can be said that economics and political science researchers do not tend to explicitly use the terminology related to goals and objectives.

6. Summary and Conclusion

This research does not attempt to establish a discipline-specific academic wordlist as suggested by the pioneering researcher Coxhead (2000), who first developed the Academic Word List (AWL), and other researchers who followed in her footsteps. In their inquiry, they attempted to identify the academic words used in research article corpora in a variety of disciplines such as applied linguistics, education, agriculture and engineering (Chung & Nation, 2003; Martínez, Beck, & Panza, 2009; Mozaffari & Moini, 2014; Mudraya, 2006; Vongpumivitch, Huang, & Chang, 2009). Unlike these studies, this research's focus is on previously established academic words or lexical items, which are essential to academic research and are subsequently inevitable to any researcher intent on investigating research problems. These include but are not limited to words such as *article, objective, project, goal, participant, test, experiment*, and *checklist*. This research also focused on quantifying these academic terms to find out patterns of use peculiar to each sub-corpora.

7. Limitation and Study Forward

This preliminary research investigated the most frequent terms used to designate academic research articles and terms used to refer to research methods and goals across various discipline abstracts. One limitation of this research is the relatively small corpus used, so one recommendation is that further studies be conducted on a larger corpus of 50 thousand or 100 thousand research abstracts. Another recommendation pertains to examining similar academic terms used in natural sciences such as physics, chemistry, mathematics, or applied sciences such as engineering and medicine to consolidate or validate the results reported here.

References

Al-Abbas, L. S., & Haider, A. S. (2020). The representation of homosexuals in Arabic-language news outlets. *Equality, Diversity Inclusion: An International Journal*, 1-29. doi:https://doi.org/10.1108/EDI-05-2020-0130

Antes, T. A., & Beck, E. (2020). Increasing Lexical Bundles in the Learner Lexicon: Binomial Expressions and the Academic Word List. *TESL-EJ, 24*(2).

Anthony, L. (2009). Issues in the design and development of software tools for corpus studies:The case for collaboration. In P. Baker (Ed.), *Contemporary corpus linguistics* (pp. 87-104). London , UK: Continuum Press.

Anthony, L. (2019). AntConc (Version 3.5.8) Tokyo, Japan: Waseda University. Retrieved from http://www.antlab.sci.waseda.ac.jp/

Aston, G. (2011). Applied corpus linguistics and the learning experience. In V. Viana, S. Zyngier, & G. Barnbrook (Eds.), *Perspectives on corpus linguistics* (pp. 3-16). Amsterdam: John Benjamins Publishing.

Bennett, G. R. (2010). An introduction to corpus linguistics. In G. R. Bennett (Ed.), *Using corpora in the language learning classroom: corpus linguistics for teachers* (pp. 4-22). Michigan Michigan Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: investigating language structure and use*. Cambridge; New York: Cambridge University Press.

Chanasattru, S., & Tangkiengsirisin, S. (2016). Developing of a high frequency word list in Social Sciences. *Journal of Studies in the English Language, 11*.

Choo, L. B., Lin, D. T. A., Singh, M. K. M., & Ganapathy, M. (2017). The significance of the academic word list among ESL tertiary students in a Malaysian public university. *3L: Language, Linguistics, Literature®, 23*(4). doi:https://doi.org/10.17576/3l-2017-2304-05

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. doi:10.2307/3587951

Csomay, E., & Prades, A. (2018). Academic vocabulary in ESL student papers: A corpus-based study. *Journal of English for Academic Purposes, 33*, 100-118. doi:https://doi.org/10.1016/j.jeap.2018.02.003

Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for specific purposes, 43*, 49-61. doi:http://dx.doi.org/10.1016/j.esp.2016.01.004

Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes, 29*, 21-42. doi: https://doi.org/10.1016/j.jeap.2017.09.002

Gilmore, A., & Millar, N. (2018). The language of civil engineering research articles: A corpus-based approach. *English for specific purposes, 51*, 1-17. doi:https://doi.org/10.1016/j.esp.2018.02.002

Green, C. (.(2019Enriching the academic wordlist and Secondary Vocabulary Lists with lexicogrammar: Toward a pattern grammar of academic vocabulary. *System, 87*, 1-10. doi:https://doi.org/10.1016/j.system.2019.102158

Haider, A. S. (2017). Frequency Analysis as a Way of Uncovering News Foci: Evidence from the Guardian and the New York Times. *International Journal of English Linguistics, 7*(2), 1-20. doi: http://dx.doi.org/10.5539/ijel.v7n2p1

Haider, A. S. (2019a). The Representation of Al-Megrahi's Release in Arabic and English Newspapers in 2009 and 2010: A Corpus-assisted Discourse Study. *Dirasat, Human Social Sciences, 46*(1 Supplement 2), 297-317. doi:https://journals.ju.edu.jo/DirasatHum/article/download/15527/10121

Haider, A. S. (2019b). Syrian-Lebanese Relations: A Corpus-based Critical Discourse Analysis of Bashar Al-Assad's Speeches and Interviews. *Dirasat, Human Social Sciences, 46*(4), 551-570. doi:https://journals.ju.edu.jo/DirasatHum/article/download/16040/11058

Haider, A. S. (2019c). Using Corpus Linguistic Techniques in (Critical) Discourse Studies Reduces but does not Remove Bias: Evidence from an Arabic Corpus about Refugees. *Poznan Studies in Contemporary Linguistics, 55*(1). doi:https://doi.org/10.1515/psicl-2019-0004

Haider, A. S., & Hussein, R. F. .(2020)Analysing headlines as a way of downsizing news corpora: Evidence from an Arabic–English comparable corpus of newspaper articles. *Digital Scholarship in the Humanities, 35*(4), 826–844. doi:https://doi.org/10.1093/llc/fqz074

Haider, A. S., & Olimy, S. (2018). The Representation of Laji'een (Refugees) and Muhajireen (Migrants) in the Headlines of Jordan News Agency (PETRA). *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique*, 1-32.

Haider, A. S., & Olimy, S. .(2019)The Representation of Laji'een (Refugees) and Muhajireen (Migrants) in the Headlines of Jordan News Agency (PETRA). *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique, 32*, 155–186. doi:https://doi.org/10.10/07s11196-018-9550-4

Haider, A. S., Olimy, S. S., & Al-Abbas, L. S. (2021). Media Coverage of Syrian Female Refugees in Jordan and Lebanon. *SAGE Open, 11*(1). doi:https://doi.org/10.1177/2158244021994811

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), 1-15.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

It-ngam, T., & Phoocharoensil, S. (2019). The development of science academic word list. *Indonesian journal of applied linguistics, 8*(3), 657-667. doi:10.17509/ijal.v8i3.15269

Jemadi, F., & Iku, P. F. (2019). Academic Words in the English Research Article Abstracts: the Coverage and Frequency. *Vision: Journal for Language Foreign Language Learning, 8*(2), 133-140. doi:10.21580/vjv8i23935

Kwary, D. A. (2018). A corpus and a concordancer of academic journal articles. *Data in Brief, 16*, 94-100.

Kwary, D. A., & Artha, A. F. (2017). The academic article word list for social sciences. *Mextesol Journal, 41*(4), 1-11.

Laviosa, S. (2012). Corpora and translation studies. In K. Hyland, C. M. Huat, & M. Handford (Eds.), *Corpus applications in applied linguistics* (pp. 6 .(84-7London: A&C Black.

Lei, L., & Liu, D. (2018). The academic English collocation list: A corpus-driven study. *International Journal of Corpus Linguistics, 23*(2), 216-243. doi:https://doi.org/10.1075/ijcl.16135.lei

Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for specific purposes, 39*, 1-11. doi:http://dx.doi.org/10.1016/j.esp.2015.03.001

Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for specific purposes, 28*(3), 183-198. doi:10.1016/j.esp.2009.04.003

McDonough, K., Neumann, H., & Hubert-Smith, N. (2018). How Accurately do English for Academic Purposes Students use Academic Word List Words? *BC TEAL journal, 3*(1), 77-89.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics*. Edinburgh: Edinburgh University Press.

Moini, R., & Islamizadeh, Z. (2016). Do we need discipline-specific academic word lists? Linguistics Academic Word List (LAWL). *Journal of Teaching Language Skills, 35*(3), 65-90. doi: 10.22099/JTLS.2016.3901

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

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for specific purposes, 25*(2), 235-256. doi:https://doi.org/10.1016/j.esp.2005.05.002

Nelson, T. S., & Albakry, M. (2020). Building a Corpus-based Interdisciplinary Vocabulary Model in University Reading Support Courses. *Journal of College Reading Learning*, 1-17. doi:10.1080/10790195.2020.1725689

Ohashi, Y., Katagiri, N., Oka, K., & Hanada, M. (2020). ESP corpus design: compilation of the Veterinary Nursing Medical Chart Corpus and the Veterinary Nursing Wordlist. *Corpora, 15*(2), 125-140. doi:https://doi.org/10.3366/cor.2020.0191

Pinnavaia, L. (2012). *Yesterday's idioms today: a corpus linguistic analysis of Bible idioms*. Paper presented at the Proceedings of the 15th Euralex International Congress.

Pournia, Y. (2019). A study on the most frequent academic words in high impact factor English nursing journals: A corpus-based study. *Iranian Journal of Nursing Midwifery Research, 24*(1), 11. doi:10.41/03ijnmr.IJNMR_190_17

Römer, U., & Wulff, S. (2010). Applying corpus methods to written academic texts: Explorations of MICUSP. *Journal of Writing research, 2*(2), 99-127. doi:10.17239/jowr-2010.02.02.2

Sayyed, S., & Al-Khanji, R. (2019). A Corpus-based Analysis of Eight English Synonymous Adjectives of Fear. *International Journal of Linguistics,, 11*(1), 111-138. doi: https://doi.org/10.5296/ijl.v11i1.14297

Shirzadi, D., & Dolatabadi, H. R. (2020). Developing an Academic Word List for the Students of Health Information Management: A Corpus Study. *Iranian Journal of English for Academic Purposes, 9*(3), 46-57.

Stubbs, M. (1996). *Text and corpus analysis: computer-assisted studies of language and culture*. Cambridge, Mass., USA; Oxford, OX, UK: Blackwell Publishers.

Talalakina, E., Stukal, D., & Kamrotov, M. (2020). Developing and Validating an Academic Vocabulary List in Russian: A Computational Approach. *The Modern language journal (Boulder, Colo.), 104*(3), 618-646. doi:10.1111/modl.12664

Todd, R. W. (2017) .An opaque engineering word list: Which words should a teacher focus on? *English for specific purposes, 45*, 31-39. doi:http://dx.doi.org/10.1016/j.esp.2016.08.003

Tognini-Bonelli, E. (2001). Corpus Linguistics at Work. *Computational Linguistics, 28*(4), 5 .583-83 doi:10.1162/coli.2002.28.4.583a

Tognini-Bonelli, E. (2010). Theoretical overview of the evolution of corpus linguistics. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 14-27). London: Routledge.

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

Veenstra, J., & Sato, Y. (2018). Creating an Institution-specific Science and Engineering Academic Word List for University Students. *Journal of Asia TEFL, 15*(1), 148. doi:https://doi.org/10.18823/asiatefl.2018.15.1.10.148

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. doi:https://doi.org/10.1016/j.esp.2008.08.003

Wang, P. (2017). A Corpus-based Study of English Vocabulary in Art Research Articles. *Journal of Arts Humanities, 6*(8), 47-53. doi:http://dx.doi.org/10.18533/journal.v6i8.1255

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

Yang, M.-N. (2017). Exploring the Use of AWL in Nursing Research Articles. *Taiwan International ESP Journal, 9*(1), 26-44. doi: 10.6706/TIESPJ.201706_9(1).0002