A CORPUS-BASED ANALYSIS OF VOCABULARY NEEDS OF ENGINEERING STUDENTS AT A STATE UNIVERSITY IN TURKEY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ŞEBNEM ÇİÇEK DEMİRCİ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF ENGLISH LANGUAGE TEACHING

MARCH 2023

Approval of the thesis:

A CORPUS-BASED ANALYSIS OF VOCABULARY NEEDS OF ENGINEERING STUDENTS AT A STATE UNIVERSITY IN TURKEY

submitted by **ŞEBNEM ÇİÇEK DEMİRCİ** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in English Language Teaching, the Graduate School of Social Sciences of Middle East Technical University** by,

Prof. Dr. Sadettin KİRAZCI
Dean
Graduate School of Social Sciences ____________________

Prof. Dr. Nurten BİRLİK
Head of Department
Department of Foreign Language Education ____________________

Prof. Dr. Ayşegül DALOĞLU
Supervisor
Department of Foreign Language Education ____________________

Examining Committee Members:

Assist. Prof. Dr. Zeynep DOYURAN (Head of the Examining Committee)
Hacettepe University
Department of English Linguistics ____________________

Prof. Dr. Ayşegül DALOĞLU (Supervisor)
Middle East Technical University
Department of Foreign Language Education ____________________

Prof. Dr. Kemal Sinan ÖZMEN
Gazi University
Department of Foreign Language Education ____________________

Assoc. Prof. Dr. Duygu ÖZGE
Middle East Technical University
Department of Foreign Language Education ____________________

Assist. Prof. Dr. Müge GÜNDÜZ
Middle East Technical University
Department of Foreign Language Education ____________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Şebnem ÇİÇEK DEMİRCİ

Signature:

ABSTRACT

A CORPUS-BASED ANALYSIS OF VOCABULARY NEEDS OF ENGINEERING STUDENTS AT A STATE UNIVERSITY IN TURKEY

ÇİÇEK DEMİRCİ, Şebnem
Ph.D., The Department of English Language Teaching
Supervisor: Prof. Dr. Ayşegül DALOĞLU

March 2023, 273 pages

Science courses constitute a significant part of engineering faculties' curriculum. This study is prompted by the need to establish the target needs of freshman engineering students at a state university with a focus on the lexical requirements of the science courses, which is believed to be valuable for curriculum or syllabus design, material development, as well as testing and assessment purposes. This study aims to generate a specialised list of lexical items using corpus frequency data derived from the textbooks used in the science courses. To this end, a corpus is compiled from the textbooks used in physics, chemistry, biology and calculus courses taken by the freshman engineering students, and keyness analysis is conducted on the corpus compiled. The corpus-derived list of keywords specific to science textbooks is then subject to expert opinion with regard to the usefulness of these lexical items for the engineering students. Employing subjective, qualitative data from interviews and questionnaires as well as objective, quantitative corpus data, the study offers a fine-grained, pedagogically convenient, corpus-derived specialised word list, comprising of 1195 lemmas, which is considered to be useful for the engineering students taking science courses at tertiary level.

Keywords: corpus, vocabulary, frequency, specialised word list, keywords

ÖZ

TÜRKİYE'DE BİR DEVLET ÜNİVERSİTESİ'NDEKİ MÜHENDİSLİK FAKÜLTESİ ÖĞRENCİLERİNİN SÖZCÜK BİLGİSİ İHTİYAÇLARININ DERLEM TABANLI ANALİZİ

ÇİÇEK DEMİRCİ, Şebnem
Doktora, İngiliz Dili Öğretimi
Tez Yöneticisi: Prof. Dr. Ayşegül Daloğlu

Mart 2023, 273 Sayfa

Fen dersleri, mühendislik fakültesi müfredatının önemli bir bölümünü oluşturmaktadır. Bu çalışma, bir devlet üniversitesindeki birinci sınıf mühendislik öğrencilerinin fen derslerindeki sözcük bilgisi ihtiyaçlarını belirleme gereksiniminden ortaya çıkmıştır. Bu ihtiyaçları belirlemenin müfredat veya izlence geliştirme, materyal tasarlama ve ölçme-değerlendirme alanlarında faydalı olacağı düşünülmektedir. Çalışma, fen derslerinde kullanılan ders kitaplarından oluşturulan, nesnel derlem verilerini kullanarak bir sözcük listesi oluşturmayı hedeflemektedir. Bu amaçla, birinci sınıf mühendislik öğrencileri tarafından alınan fizik, kimya, biyoloji ve matematik derslerinde kullanılan ders kitaplarından bir derlem oluşturulmuş ve bu derlem üzerinde anahtar sözcük analizi yapılmıştır. Derlem verileri ile oluşturulan listedeki sözcüklerin öğrenciler için ne ölçüde faydalı olduğu ile ilgili uzman görüşü alınmıştır. Nesnel, niceliksel derlem verisinin yanı sıra görüşme ve anketler ile öznel, niceliksel verilerden de yararlanan bu çalışma, pedagojik olarak uygun, derleme dayalı ve 1195 sözcükten oluşan bir hedef sözcük listesi sunmaktadır ve bu listenin yüksek öğrenim düzeyinde fen derslerini alan mühendislik öğrencileri için faydalı olacağı düşünülmektedir.

Anahtar sözcükler: derlem, sözcük bilgisi, sıklık, alana özgü sözcük listesi, anahtar sözcük

To my father, who, I believe, is still with me…

ACKLNOWLEDGEMENTS

I would like to express my deepest gratitude to the following people for their support and guidance throughout the preparation of this dissertation.

First and foremost, I would like to express my sincere thanks to my supervisor, Prof. Dr. Ayşegül Daloğlu, for her endless support and encouragement during my process of academic development. Her academic guidance, supportive attitude and unwavering belief in me made me complete this research.

I am also grateful to the committee members, Prof. Dr. Kemal Sinan Özmen, Assoc. Prof. Dr. Duygu Özge, Asst. Prof. Dr. Zeynep Doyuran and Asst. Prof. Dr. Müge Gündüz, for their precious feedback, which helped improve this work.

My special thanks go to my dear friend, Dr. Hakan Cangır, whose knowledge and understanding I have great faith in. He not only motivated me all the way but also provided me with his valuable guidance and helpful feedback at every phase of this work. Without his support, I would not have been able to complete this study.

I am also thankful to my husband, Volkan Demirci, who always encouraged me to realise my ambition and supported me throughout this difficult journey. His trust in me made me achieve my goal.

TABLE OF CONTENTS

PLAGIARISM iii
ABSTRACT iv
ÖZ v
DEDICATION vi
ACKNOWLEDGMENTS vii
TABLE OF CONTENTS viii
LIST OF TABLES xi
LIST OF FIGURES xiii
LIST OF ABBREVIATIONS xiv
CHAPTERS
1. INTRODUCTION 1
1.1. Background of the Study 5
1.2. Statement of the Problem 4
1.3. Purpose of the Study 6
1.4. Significance of the Study 7
2. LITERATURE REVIEW 10
2.1. Needs Analysis 10
2.1.1. Hutchinson and Waters' Needs Analysis Model 11
2.1.1.1. Target Needs 11
2.1.1.2. Learning Needs 12
2.2. Corpus Linguistics 13
2.3. Types of Corpra 18
2.3.1. General Corpora 18
2.3.2. Specialized Corpora 18
2.3.3. Learner Corpora 19
2.3.4. Historical Corpora 19
2.3.5. Parallel Corpora 19
2.3.6. Comparable/ Multilingual Corpora 20

2.4. English for Specific Purposes 20
2.5. Specialised Corpus and ESP 22
2.6. Use of Corpus Data in Teaching 26
2.7. Corpus Terms 28
2.7.1. Frequency 28
2.7.2. Concordance 29
2.7.3. Keyness 31
2.7.4. N-grams 32
2.8. Corpus Statistics 32
2.8.1. Log-likelihood 32
2.8.2. T-score/Mutual information 33
2.8.3. Type-Token Ratio 33
2.9. Vocabulary and Language Use 33
2.9.1. Vocabulary-related terminology 34
2.9.2. Vocabulary and frequency data 35
2.9.3. Formulaic language and corpora 36
2.9.4. Word Lists 38
2.10. Relevant Studies 41
2.10.1. Flowerdew's study (1986) 41
2.10.2. Ward's study (1999) 44
2.10.3. Mudraya's study (2006) 45
2.10.4. Chen and Ge's study (2007) 45
2.10.5. Coxhead and Hirsch's study (2007) 46
2.10.6. Martinez et al.'s study (2009) 47
2.10.7. Brezina and Gablasova's study (2015) 48
2.10.8. It-ngam & Phoocharoensil's study (2019) 48
2.11. A Brief Summary of the Review of Literature 50
3. METHODOLOGY 52
3.1. Introduction 52
3.2. Research Design 53
3.3. Setting 57
3.4. Data Collection Tools 58

3.4.1. Interview 58
3.4.2. Corpus compilation 59
3.4.3. Questionnaire 60
3.5. Data Analysis 62
3.5.1. Content analysis 62
3.5.2. Corpus analysis 63
3.5.2.1 Frequency analysis 63
3.5.2.2 Keyword analysis 66
3.5.3. Coverage analysis 67
3.5.4. Statistical analysis 68
4. FINDINGS AND RESULTS 69
4.1. Introduction 69
4.2. Findings fron Interviews with Course Designers 69
4.2.1. Requirements of the Course 70
4.2.2. Course Content 71
4.2.3. Skills and Subskills Needed 71
4.2.4. Difficulties Faced by the Students 72
4.2.5. Suggestions 73
4.3. Corpus Analysis 74
4.3.1. Compilation of the Science Textbooks Corpus 74
4.3.2. Frequency Analysis 76
4.3.3. Keyness Analysis 78
4.3.4. Coverage Analysis 83
4.4. Findings from the Questionnaire 84
5. DISCUSSION 98
5.1. Introduction 98
5.2. Evaluation of Research Questions 98
6. CONCLUSION 116
6.1. Introduction 116
6.2. Summary of Research 116
6.3. Pedagogical Implications 120
6.4. Limitations and Further Research 123

REFERENCES.. 125
APPENDICES
A. Interview Questions.. 144
B. Frequency List.. 145
C. Keyword List.. 185
D. List of Multi-Word Units.. 212
E. Questionnaire 1 on Keywords.. 221
F. Questionnaire 2 on Keywords.. 227
G. Questionnaire 3 on Keywords.. 232
H. Questionnaire 4 on Keywords.. 237
I. Questionnaire 5 on Keywords.. 242
J. Questionnaire on Multi-word Units.. 247
K. APPROVAL OF RESEARCH BY HUMAN SUBJECTS ETHICS COMMITTEE.. 251
L. CURRICULUM VITAE.. 252
M. TURKISH SUMMARY/ TÜRKÇE ÖZET.. 253
N. TEZ İZİN FORMU/ THESIS PERMISSION FORM.. 273

LIST OF TABLES

Table 1. The number of students in the preparatory programme as per faculties .. 5
Table 2. Hutchinson and Waters' (1987) target needs analysis framework 12
Table 3. Hutchinson and Waters'(1987) learning needs analysis framework 13
Table 4. Some studies on discipline-specific wordlists 41
Table 5. Top 20 nouns in the Cobuild corpus and in a biology corpus 42
Table 6. The number of occurrences of certain connectors in a biology corpus .. 43
Table 7. The coverage of base word lists over the SAJ corpus 49
Table 8. An overview of research design 54
Table 9 The list of engineering departments where medium of instruction is English 58
Table 10. Textbooks used in first-year science courses in the engineering departments 60
Table 11. Questionnaire for course instructors 62
Table 12. An example of frequency results for the word "helps" in the BNC, retrieved from Sketch Engine. 64
Table 13. Content analysis results 70
Table 14. The number of tokens and coverage percentages of the sub-corpora 75
Table 15. The 20 most frequent items in the Science Textbooks Corpus 76
Table 16. The 20 most frequent lexical items in the Science Textbooks Corpus ... 77
Table 17. The first 30 headwords in the keyword list 79
Table 18. The last 30 headwords in the keyword list 81
Table 19. The 30 most frequent multi-word units in the Science Textbooks Corpus 82
Table 20. Coverage values of the wordlists 84
Table 21. Statistical Findings for Questionnaire 1 85
Table 22. Pearson's Correlations for Questionnaire 1 86
Table 23. Items with scores below 3 in Questionnaire 1 87
Table 24. Statistical Findings for Questionnaire 2 88
Table 25. Pearson's Correlations for Questionnaire 2 88

Table 26. Items with scores below 3 in Questionnaire 2 89
Table 27. Statistical Findings for Questionnaire 3 90
Table 28. Pearson's Correlations for Questionnaire 3 90
Table 29. Items with scores below 3 in Questionnaire 3 91
Table 30. Statistical Findings for Questionnaire 4 92
Table 31. Pearson's Correlations for Questionnaire 4 92
Table 32. Items with scores below 3 in Questionnaire 4 93
Table 33. Statistical Findings for Questionnaire 5 94
Table 34. Pearson's Correlations for Questionnaire 5 94
Table 35. Items with scores below 3 in Questionnaire 5 95
Table 36. Statistical Findings for the Questionnaire on Multi-word Units 96
Table 37. Pearson's Correlations for Questionnaire on Multi-word Units 97
Table 38. Items with scores below 3 in Questionnaire on Multi-word Units 97
Table 39. Two examples from the target and the reference corpora 106
Table 40. Summary of Questionnaire Findings 112

LIST OF FIGURES

Figure 1. An example of concordance from BNC 30
Figure 2. Distribution of the wordlist items as per CEFR levels 80

LIST OF ABBREVIATIONS

ARF **:** Average Reduced Frequency

AVL : Academic Vocabulary List

AWL : Academic Word List

BNC : British National Corpus

COCA: Corpus of Contemporary American English

CEFR : Common European Framework of Reference

DOCF **:** Document Frequency

ESP **:** English for Specific Purposes

EAP : English for Academic Purposes

ELT : English Language Teaching

EFL : English as a Foreign Language

GSL : General Service List

L1 : The source language

L2 : The target language

LOB : Lancester-Oslo-Bergen Corpus

STC : Science Textbooks Corpus

STWL: Science Textbooks Word List

CHAPTER I

INTRODUCTION

1.1 Background of the Study

Vocabulary knowledge has been considered to be a key factor in learning a foreign language (e.g., Koizumi & In'nami, 2013) and focusing on high-frequency words is believed to be effective for proficiency development (e.g., Nation, 2006). To meet this need of vocabulary input, vocabulary lists are widely used by teachers to plan course syllabus, design materials or prepare tests. Corpus data plays a major part in developing such lists (Jones and Durrant, 2010).

Corpora are collections of texts available in electronic form which provide access for analysing recurrent patterns in a language. What make corpora different from other types of text collections, according to Bowker and Pearson (2002), are four important characteristics: "authentic", "electronic", "large" and "specific criteria" (p.9). Corpus linguistics as a method of text analysis based on electronic tools can be considered to have started in the 60s–70s with the compilation of the Brown and the LOB (The Lancester-Oslo/Bergen) corpora, two collections of 1 million words and 500 sample-texts each, of American and British English respectively (Gavioli, 2005). At the beginning of the 90s, corpus linguistics was rising in popularity. The Cobuild project led by John Sinclair (Sinclair, 1987) came as a breakthrough in the field. It aimed to produce more realistic descriptions of English language to be taught in the classroom (Gavioli, 2005). Despite the progress it has made in the last decade, use of corpora as a teaching or learning tool is still limited, and as noted by Kennedy (2004) three decades of research on corpora has had very little effect on language curricula. But still, according to Biber and Reppen (2002), in the past twenty years, empirical analyses of corpora have contributed to the description of the actual patterns of language use in English. Language corpora provide systematic access to naturally occurring language, so actual patterns of language use are made available for teachers, learners, researchers, testers and many more. As noted by Nelson (2010), the exponential rise of available electronic corpora over the last twenty years has provided the academic community with an enormous amount of ready-made data that can be accessed easily on-line.

Meunier and Reppen (2015) argue that corpus-based research should inform the development of textbooks. As textbooks are a valuable source of language input for learners, it is crucial that they reflect samples of real, naturally occurring language in their content. Burton (2012) and Gilmore (2015) think that many ELT writers fail to benefit from corpus analysis. According to Gilmore (2015) "textbook authors are not yet habitually checking their materials against relevant corpus data" (p. 517). This results in the fact that many features that characterise natural discourse (e.g. collocations) are not highlighted in reference books and/or textbooks well enough and the language that learners are exposed to differs significantly from the actual usage (Römer, 2011). Some corpus linguists examined the features of the language of textbooks comparing them with authentic language. Gouverneur (2008) explored the phraseological patterns using the TeMa corpus, which is a collection of textbooks used in general English courses. His study showed that although the textbooks included many exercises of verb-noun collocations with 'make' and 'take', at the advanced level little attention was paid to collocations formed with delexical verbs. He also found that the chunks included in the materials were selected inconsistently. Koprowski (2005) also looked at three intermediate level EFL textbooks and analysed the usefulness of the chunks that were presented in them. He found that many useful chunks were not presented in the textbooks. He came to the conclusion that the selection of lexical phrases to be included in textbooks was "an unprincipled and careless selection process" based on "the personal discretion and intuition of writers" (p. 328).

Corpus data is also a powerful tool to gain an understanding of the recurrent features of discipline-specific discourse, which can provide valuable insights for ESP. A specialised corpus is defined by Bowker and Pearson (2002) as "one that focuses on a particular aspect of a language" (p.12). They are designed to create a sample of specialized language either by collecting texts of similar content (e.g. science, medicine, business, philosophy) or of similar text-type or genre (e.g. research papers, letters, book chapters) or both or even texts from other types of specialized categories, such as newspaper language or academic language (Gavioli, 2005). Such corpora can provide valuable data that may establish the basis for the selection and grading of the items to be included in a syllabus or a curriculum. In addition, frequency lists obtained from specialised corpora can play an important role in programme evaluation. The discrepancies between the specialised corpora and the general one might have implications on the programme in use. Flowerdew (1993) suggests that lists of the most frequently occurring words in a specialized corpus together with their concordances can be used to decide on the contents to be included in the ESP syllabus. Flowerdew (1993) analysed a collection of biology texts that students were supposed to read and transcriptions of lectures they attend during the course. He calculated the word types and frequency in his 100,000 word corpus, and found that about 1000 items occur more than 10 times in the corpus and notes that a comparison of the frequency list of the specialised corpus and that of a general corpus provides implications for syllabus design. Specialised corpora are widely used in ESP and EAP contexts particularly because of the need to identify specific lexical needs in different disciplines. For example, a corpus aimed at first-year PhD science and engineering students at the University of Nottingham was compiled to create word lists and concordances on which vocabulary teaching materials could be based (Jones & Durrant, 2010). The corpus consists of 11,624,741 words and covers a wide range of disciplines in the faculties of science and engineering.

Flowerdew (1993), who highlights the importance of corpus data in ESP syllabus design notes that "high face validity is given to an ESP course if the learning materials contain actual examples of use which are drawn from the content area and which the learner is likely to have come across, or will be likely to come across, in his specific studies" (p.239). The inauthentic examples presented to learners always pose a risk of presenting an inaccurate or defective image of language use. Flowerdew (1993) believes that using authentic examples in teaching materials allows learners to access a real representation of real use. Gavioli (2005) also thinks that corpora can be a good resource especially for ESP course design while syllabus design in general English may be too complicated to limit to corpus work only. A corpus of specialised texts can be exploited to produce a set of items that characterise those texts and these items may constitute the basis to select language features which will be included in ESP syllabi.

Computer software developed for corpus analysis makes it possible to generate a list of recurrent lexical items in a set of texts. Such a specialised word list can constitute a lexical basis which the teacher may take into account during the courses. Higgins' study (1967) on teaching English to science students stresses the need for providing students with some frame words that cause problems in comprehension and production and which occurred frequently in the language of science. Some examples are from the domain of medical science "symptoms", "diagnose"/ "diagnosis", "treatment", "relapse", "heal, and "cure".

According to Nation (2001), there are several ways of making lists of academic vocabulary, three of which (as cited in Gavioli 2005, pp. 59-60) are as follows:

1. Word frequency calculation can be useful in identifying the most important specialised words if the domain is very specialised;
2. Corpora of texts from different specialised domains can be compared to show key lexical characteristics in each of the corpora;
3. Corpora of texts in different specialised domains can be compared with corpora of non-academic texts to demonstrate specialised vs. general lexical features.

1.2. Statement of the Problem

Ankara University is a state university with a large number of students in various faculties. Students who successfully finish high school and manage to get the required score in the university entrance exam are entitled to enroll at the department they choose in a university. If the medium of instruction is English in the department they are to study, they are required to certify that they acquire the necessary level of proficiency to perform their studies in English. This is often through the English proficiency exam held by the school of foreign languages. The students who get the required score in the exam can start their major. Those who cannot are required to receive English education in the preparatory programme.

In the academic year 2022/2023, 1722 students were enrolled at Ankara University School of Foreign Languages, and 995 of them were registered in the English programme. For the students of departments where medium of instruction is fully English, preparatory English education is compulsory. Among the students registered in the English preparatory programme, engineering students constitute the largest group, with 554 students, followed by the Faculty of Science, with 161 students. Table 1 shows the numbers of students in the preparatory programme as per faculties:

Table 1

The number of students in the preparatory programme as per faculties

Faculty	**Number of Students**
Faculty of Engineering	554
Faculty of Science	161
Faculty of Language, History and Geography	103
Faculty of Pharmacy	72
Faculty of Veterinary	43
Faculty of Agriculture	40
Faculty of Divinity	22
Total	**995**

The greatest number of students in the English preparatory programme are the students of engineering departments, namely Computer Engineering, Biomedical Engineering, Electrical and Electronics Engineering, Energy Engineering, Energy Systems Engineering, Physics Engineering, Food Engineering, Geology Engineering, and Chemical Engineering. Basic science courses constitute the core subjects in the engineering departments' first year curriculum. The students of the faculty of engineering take physics, chemistry, calculus and biology courses in their first year.

The students who have completed the English preparatory programme successfully are entitled to start their major, while those who fail are required to retake the programme in the following year. Throughout the programme, students get English education in mixed classes with students from various departments. They learn English through coursebooks for General English as well as supplementary materials developed in-house by a team of teachers working on material development. As part of the general English programme, students have to take the exams given at certain intervals throughout the term. The exam content depends on the content of the coursebook. Yet, what the content of the coursebook depends on or whether the target items in the coursebook correspond with what those students truly need is not known. As such, the syllabus designers or test writers make decisions with no sound basis. The proficiency exam content covers the lexical items in the coursebook or in the materials. CEFR levels are the sole criteria to be referred to in preparing assessment content. This purely subjective approach to teaching and testing with no needs-driven principles reveals the need for specifying the students' target lexical needs based on objective data that is specific to the target group.

Identifying recurrent patterns in a specialised corpora composed of scientific texts that the engineering students would be exposed to is believed to yield insights for sound decisions in the selection of items to be included in the syllabus/curriculum, materials and assessment tasks.

1.3. Purpose of the Study

The ultimate purpose of this study is to help first year engineering students become proficient users of the language so that they can perform their studies without linguistic obstacles. The science courses that engineering students take in their first year constitute a significant part of their studies on which they will build other masteries in their specific discipline. The one-year English preparatory programme they take prior to starting their major, therefore, is of paramount importance for the language proficiency they need to develop. Whether the programme caters for the needs of these students in terms of lexical knowledge is unknown. This study aims to identify the target needs of the freshman engineering students in the science courses they take with a focus on their lexical needs. It is intended to specify most frequent lexical items and combinations of words or structures found in the science textbooks used by the freshman engineering students at Ankara University by creating a specialised corpus. The students enrolled in the preparatory programme are mostly engineering students and they all take the must courses: Physics, Chemistry, Biology and Calculus. The written material covered in these courses have been collected and a specialised corpus have been created. Following the compilation of the corpus a set of analyses have been conducted for the purposes of the study. In this respect, the following research questions are to be addressed:

1. What are the freshman engineering students' target lexical needs for the science courses?

    1.1. What are the perceptions of the lecturing staff regarding the freshman engineering students' target needs?

    1.2. What specific vocabulary do the science textbooks used by freshman engineering students feature?

    1.2.1. What are the lexical frequency representations of the science textbooks used by freshman engineering students?

    1.2.2. What keywords and multi-word terms constitute the key vocabulary in the science textbooks of freshman engineering students?

    1.3. To what extent does the content of the English preparatory programme meet the target lexical needs of freshman engineering students for the science courses?

    1.4. How does a keyword list based on a corpus of science textbooks relate to the commonly available wordlists, namely the *New General Service List*, the *New Academic Vocabulary List* and the *Science Word List?*

    1.5. What are the perceptions of the lecturing staff regarding the usefulness of the items in the key word list derived from a corpus of science textbooks?

1.4. Significance of the Study

Many teachers think that helping undergraduate students develop control over a specialist vocabulary is important and attempts have been made to develop lists of key terms to guide materials writers and help students plan their learning more efficiently (Hyland and Tse, 2007). Such lists are developed surmising that they constitute the most frequently used vocabulary in real general or academic contexts, which would assist students in their studies. However, it would be a fallacy to expect that such a repertoire of lexical items is uniform for each and every discipline, domain or genre. In that respect, Hyland and Tse (2007) state that

> *…whether it is useful for learners to possess a general academic vocabulary is more contentious because it may involve considerable learning effort with little return. It is by no means certain that there is a single literacy which university students need to acquire to participate in academic environments, and we believe that a perspective which seeks to identify and teach such a vocabulary fails to engage with current conceptions of literacy and EAP, ignores important differences in the collocational and semantic behavior of words, and does not correspond with the ways language is actually used in academic writing. It is, in other words, an assumption which could seriously mislead students. (p. 236-237).*

This study is based on the assumption that science courses, constituting a major portion of engineering students' academic studies, feature a specific lexis which would not be covered by the generic lists, namely the new GSL, the new Academic Vocabulary List, and the Science Word list. Therefore, the study aims to establish an inventory of target word list that is specific to the scientific texts used in the specific context of the engineering departments at a state university. The rationale behind constructing a word list from scratch rather than making use of the available wordlists such as AWL (Academic Word List), GSL (General Service List), and AVL (Academic Vocabulary List), is that a general vocabulary list is by no means capable of addressing the specific needs of a specific group of learners in a specific context. In this respect, Hyland and Tse (2007), in their study where they question the assumption that EAP students should study a core of high frequency words common in an academic register, undermines the usefulness of a general academic vocabulary which caters for all students' needs regardless of their specific field of study, and provides evidence that it is necessary to identify students' target language needs and developing "a more restricted, discipline-based repertoire" (p. 235). They believe that this involves "introducing, making salient, and practicing the specialized vocabulary of their fields or discipline (p. 249).

The data to be obtained from this study is intended to assist curriculum developers, material designers and test writers in selecting items to be included in the curriculum, syllabus and assessment. A course syllabus can be developed around the language the students are exposed to in their science classes. A comparison of the corpus compiled in this study with a benchmark corpus is indicative of the core of words that might characterise the specific domain where the data is collected, and thus may be considered to be included in the curriculum or syllabus. Collocational frequencies as well as individual words recurring in the corpus data may guide material developers in prioritising the items to be taught and help learners gain naturalness and automaticity in their linguistic production. By means of a specific target corpus, learners can be presented with instances of actual uses they are likely to encounter during their studies.

The study may also be helpful in guiding decision-makers at preparatory schools of other universities in designing their curriculum or syllabi, selecting EFL coursebooks and providing supplementary lexical resources for students. Well-designed, corpus-derived and pedagogically convenient wordlists can be essential resources for effective learning. As Gavioli (2005) observes, the contribution of corpus work in dealing with specialised vocabulary should not be underestimated and lists of academic words have been profitably used in teaching specialised vocabulary. According to Dang (2020), corpus-based wordlists have an important effect in the selection and development of learning materials for meaning-focused input, meaning-focused output, as well as fluency development activities.

CHAPTER II

LITERATURE REVIEW

2.1. Needs Analysis

Needs analysis is a crucial initial step in designing a syllabus or any other decisions to be taken with regard to the content learners will be exposed to, whether it is an ESP course or a general English course. The value of needs analysis is acknowledged by many scholars and experts (Munby, 1978; Richterich and Chancerel, 1980; Hutchinson and Waters, 1987; Berwick, 1989; Robinson, 1991; Johns, 1991; Seedhouse, 1995; Jordan, 1997; Dudley-Evans and St. John, 1998; Iwai et al. 1999). Needs analysis, the focus of which is primarily the goals and content of a course, aims to discover what the learners already know and what they need to know (Nation & Macalister, 2010). Iwai et al. (1999) define needs analysis as the activities that involve collecting information which will set the basis of a curriculum which will cater for the needs of a specific group of students.

A number of needs analysis models have been proposed by scholars: Munby (1978), McDonough (1984), Hutchinson and Waters (1987), Robinson (1991), Jordan (1997), Dudley-Evans and St John (1998). Amongst all of the models, probably the most thorough and commonly known work is John Munby's *Communicative Syllabus Design* (1978). In his model, Munby describes a set of procedures, which he calls the Communication Needs Processor (CNP), to discover target situation needs through a range of questions regarding key communication variables. The target needs and target level performance are determined through examining the target situation. In his model, the variables that affect communication needs are organised as parameters in a dynamic relationship to each other (Munby, 1978). Most subsequent needs analysis models were based on Munby's model on the grounds that it provides exhaustive data banks and target performance (Robinson, 1991).

2.1.1 Hutchinson and Waters' Needs Analysis Model

Hutchinson and Waters (1987) categorise the concept of "needs" into two broad categories of target needs and learning needs. Target needs refers to what the learner needs to do in the target situation, and learning needs refers to what the learner needs to do in order to learn.

2.1.1.1 Target Needs

Hutchinson and Waters (1987) state that "target needs" is an umbrella term and that it involves a number of distinctions such as "necessities", "lacks" and "wants" (p. 55). They define necessities as "the type of need determined by the demands of the target situation; that is, what the learner has to know in order to function effectively in the target situation" (p. 55). Lacks, on the other hand, refer to "what the learner knows already, so that you can decide which of the necessities the learner lacks" (p. 56). Both necessities and lacks are objective needs, yet, Hutchinson and Waters (1987) think that it is also important to consider subjective needs, and state that the learners also have an opinion as to what their needs are. But they also warn that there is the possibility of learners' views' conflicting with the views of other parties such as course designers, teachers, and sponsors. Using the analogy of the ESP course as a journey, they consider the starting point "lacks" and the destination "necessities"; "wants" refer to what the destination should be, as perceived by the learner. Hutchinson and Waters (1987) explain "wants" as "what the learners want or feel they need" (p. 57).

Hutchinson and Waters' (1987) target situation analysis framework, consists of a set of questions to be answered, most of which also are relevant to Munby's model. It mostly addresses such issues as the reasonswhy the language is needed and how, where and when it will be used as well as the content areas. Table 2 below shows Hutchinson and Waters' target needs analysis framework.

Table 2
Hutchinson and Waters' target needs analysis framework. (1987)

1. Why is language needed? • for study; • for work; • for training; • for a combination of these; • for some other purposes, e.g. status, examination, promotion
2. How will the language be used? • Medium: speaking, writing, reading, etc.; • Channel: e.g. telephone, face to face; • Types of text or discourse: e.g. academic text, lectures, catalogues, etc.
3. What will the content areas be? • Subjects: e.g. medicine, biology, commerce, shipping, etc.; • Level: technician, craftsman, postgraduate, etc.
4. Where will the language be used? • Physical setting: e.g. office, lecture theater, hotel, workshop, library; • Human context: alone, meetings, demonstrations, on telephone; • Linguistic context: e.g. in own country, abroad.
5. When will the language be used? • Concurrently with the ESP course or subsequently; • Frequently, seldom, in small amounts, in large chunks.

2.1.1.2 Learning Needs

Different from the target situation needs, which focuses on the knowledge or abilities required for the learners to be able to perform to the required degree of competence in the target situation, learning needs refer to the question "how". Learning needs are pertaining to the aspects of learning what is required as the target needs. In that respect, Hutchinson and Waters (1987) state that

> *"It is naïve to base a course design simply on the target objectives, just as it is naïve to think that a journey can be planned solely in terms of the starting point and the destination. The needs, potential and constraints of the route (i.e. the learning situation) must also be taken into account, if we are going to have any useful analysis of learner needs." (p. 61).*

They suggest a similar framework to examine learning needs, shown in Table 3.

Table 3

Hutchinson and Waters' learning needs analysis framework (1987)

1. Why are the learners taking the course? • compulsory or optional; • apparent need or not; • Are status, money, promotion involved? • What do learners think they will achieve? • What is their attitude towards the ESP course? Do they want to improve their English or do they resent the time they have to spend on it?
2. How do the learners learn? • What is their learning background? • What is their concept of teaching and learning? • What methodology will appeal to them? • What sort of techniques bore/alienate them?
3. What sources are available? • number and professional competence of teachers; • attitude of teachers to ESP; •teachers' knowledge of and attitude to subject content; • materials; • aids; • opportunities for out-of-class activities.
4. Who are the learners? • age/sex/nationality; • What do they know already about English? • What subject knowledge do they have? • What are their interests? • What is their socio-cultural background? • What teaching styles are they used to? • What is their attitude to English or to the cultures of the English-speaking world?

2.2. Corpus Linguistics

Cheng (2012) defines corpus linguistics as the study of "the compilation and analysis of corpora" (p.6), which are, as defined by Sinclair (1991), "large collections of language texts that naturally occur and are chosen to characterise a state or a variety of language" (p. 171). The need for employing quantitative data in language pedagogy dates back at least to the 1920s (Thorndike 1921, quoted in Kennedy 1992). To meet that need, researchers used to conduct manual counting and classification of "corpora" of texts to reach information on the distribution of words and forms in studies regarding vocabulary, syntax, semantics and the development of children's language. The history of corpus linguistics as a means for textual analysis dates back to 1960s and 1970s, when the Brown and the Lancester-Oslo-Bergen (LOB) corpora were compiled (Gavioli, 2005). These two corpora, with 1 million words and 500 texts each, established the basis of modern corpus linguistics. In the 1990s, corpus linguistic was gaining popularity and the Cobuild project led by John Sinclair was a major development in the field. Collins Cobuild English Project, in which a corpus of 20 million words was employed, aimed to provide realistic input for language teaching. Gavioli (2005) states that the Cobuild catch-phrase helping students with real English seemed to imply an equivalence between:

a) a corpus and real language
b) corpus-based descriptions and "more realistic" students' language production.

Some applied linguists were alarmed by the oversimplificiation of such equivalences and raised objections. Gavioli (2005) explains the situation under three main points. One reason for the objections was that, a corpus, however large, is still only a sample of language production and despite the possibility of providing an accurate representation of the language, a corpus is not the real language. The second reason is that being exposed to real samples of language does not ensure that students learn real language. Carter (1998) claims that invented examples reveal the features of the spoken language more clearly than corpus materials do. Third, in language pedagogy reality cannot be limited to using texts that occurred somewhere and sometime in real life. Widdowson (1998) believes that simply exposing students to corpus-based descriptions or to genuine material from corpora does not mean that the learners will be able to authenticate the language they are exposed to, that is, produce the language in a communicative context.

Despite controversies on the uses of corpora in language teaching, corpora still seem to be a powerful enrichment for the language classroom. According to Kennedy (1992):

> *…corpus linguistics has held potential relevance for the teaching of languages because responsible language teaching involves selecting what is worth giving attention to. Since pedagogy attempts to reduce the time that would be necessary to learn a language through exposure alone, potential usefulness*

and likelihood of occurrence have been seen as relevant for deciding what to teach or learn (p. 335).

Today, the situation seems to be changing as a result of a renewed interest in form-focused teaching; also a number of studies now show various aspects of computer-corpus applications to language teaching and learning (Gavioli, 2005).

Corpus linguistics, despite being a relatively new field, has brought about new dimensions to the way language is analysed, described and taught. Hunstun (2002), thinks it has revolutionised language studies and through computerised study of data it is possible to gain insights into the structure and regularity of naturally occurring language. Biber et al. (1998, p. 4) describe some important traits of corpus analysis as follows:

- *it is empirical, analysing the actual patterns of use in natural texts;*
- *it utilises a large and principled collection of natural texts, known as a corpus, as the basis for analysis;*
- *it makes extensive use of computers for analysis, using both automatic and interactive techniques;*
- *it depends on both quantitative and qualitative analytical techniques.*

Flowerdew (2012) points to two remarkable features of corpus linguistics and states that corpus linguistics is an empirical methodology and follows a phraseological approach to language. It is an empirical approach in that through corpus analyses, we obtain data-based descriptions of language, which help us discover the typical patterns and structures of the language. Automatic, computer-assisted nature of the corpora searches makes the results more reliable and objective. Hunston (2002) points out that as speakers' linguistic experience mostly "remains hidden from introspection", corpora are "a more reliable guide to language use than native speaker intuition (p. 20). The fact that it follows a phraseological approach also makes corpus linguistics a valuable and useful resource. Szudarski (2017) notes that:

Corpus studies demonstrate that language exhibits a highly patterned structure and consists of different kinds of phraseological patterns. More specifically, detailed analysis of large amounts of data reveal that grammar and vocabulary are inextricably intertwined, and the notion of lexico-grammar becomes the focal point of corpus analysis. (p. 8)

Regarding the benefit of corpora for studying language, Hunston (2002) points out that corpus work can be used in language teaching and learning, discourse analysis, translation studies, language for specific purposes, pragmatics, sociolinguistics, media discourse, literary linguistics and political linguistics. The exploration of large amounts of data through computer-assisted searches makes it possible to have an understanding of speakers' and writers' linguistic choices and how these choices are affected by context, register, genre, audience, purpose and form of communication.

There are various definitions of corpus but there is an increasing agreement that a corpus is a collection of (1) machine-readable (2) authentic texts (including transcripts of spoken data) which is (3) sampled to be (4) representative of a particular language or language variety (Mc.Enery et al., 2006). In order to build a corpus there are a number of factors which need to be taken into consideration, which include size, balance and representativeness (Evans, 2007).

As for size, Evans (2007) states that bigger corpora are believed to be better but it is also possible to obtain useful data from a small corpus, especially when searching for high frequency items. His following quote explicates his opinion that the size of the corpus varies according to the purpose of the research:

> *The size of the corpus depends very much on the type of questions that are going to be asked of it. As a rule of thumb, bigger is generally considered to be better as the software can be instructed to filter out some of the output. However, it is possible to get much useful data from a small corpus, particularly when investigating high frequency items. In fact, this may be desirable to do this rather than being overwhelmed by too much data from a big corpus. (p.1)*

There has been a debate in the literature regarding optimal sample size in corpora (Nelson, 2010). According to Oosstdijk (1991), "a sample size of 20,000 words would yield samples that are large enough to be representative of a given variety" (p.50). However, the size of the samples may closely be related to the genre studied as well as the purpose of the corpus building. Biber (1993) notes that there is considerable variation in genres in terms of size on the grounds that 20,000 words would provide an adequate sample size for some genres but for others it would not. In the statistical sense, samples are “scaled down versions of a larger population” and a sample is regarded as representative if the findings of a sample also holds for the general population (McEnery et al.,2006, p. 19).

Regarding the use of extracts or full texts as the corpus data, Stubbs (1996, p.32) suggests that “few linguistic features of a text are distributed evenly throughout”, which implies that a sample of a given text would fail to encompass all the linguistic features of that text. Nelson (2010) emphasises that this is particularly important in studying genre. He adds that:

> *Studies into genre have noted how certain linguistic features are typical of certain parts of a text and an approach to corpus creation that only takes extracts at random will fail to gain a representative sample in this respect. Thus, as with other aspects of corpus design, the purpose to which the corpus will be put is critical in deciding whether to use whole texts or not. (p. 59)*

According to Leech (1991), a corpus can be considered to be representative of the language variety it is supposed to represent provided that the findings based on its contents can be generalized to that language variety (p. 27). Biber (1993) defines representativeness as “the extent to which a sample includes the full range of variability in a population” (p. 243). The generalisability of the corpus findings to the language or a specific aspect of the language determines the representativeness of a corpus. Mc Enery et al. (2006) point to two factors which are influential in determining the extent to which a certain corpus is representative: “the range of genres included in a corpus” and “how the text chunks for each genre are selected” (p.13). They believe that it is essential that both general and specialized corpora represent a language or language variety.

Balance is another consideration in constructing a corpus, which suggests that there be a balance among the sample texts chosen for the corpus. As explained by Mc Enery et al., (2006) a balanced corpus usually covers a wide range of text categories which are supposed to be representative of the language or language variety under consideration. Although balance is often considered an important criterion for corpus design, “there are no reliable ways of determining whether a corpus is truly balanced” (Evans, 2007, p. 1). Mc Enery et al. (2006) also state that “any claim of corpus balance is largely an act of faith rather than a statement of fact as, at present, there is no reliable scientific measure of corpus balance. Rather the notion relies heavily on intuition and best estimates.”

2.3. Types of Corpora

There are different types of corpora categorised according to such factors as their content, use, purpose or language. Six types of corpora, namely general, specialised, learner, historical, parallel and comparable corpora will briefly be defined below.

2.3.1. General Corpora

General corpora refer to the collection of a wide array of texts representing natural language used in various contexts. Bowker and Pearson (2002), using the term “general reference corpora”, define it as “one that can be taken as representative of a given language as a whole and can therefore be used to make general observations about that particular language” (p. 12). They involve material-both written and spoken-from a variety of contexts and disciplines (e.g. fiction, newspapers, academic journals, conversations etc.). Hunston (2002) points out that a general corpus contains many types of texts that represent as wide a spread as possible (as cited in Szudarski, 2017, p.10). British National Corpus (BNC) developed by Mark Davies and consisting of about 100 million words and the Corpus of Contemporary American English (COCA), consisting of about 520 million words are two well-known and commonly-used general corpora.

2.3.2. Specialised Corpora

Specialised corpora are smaller collection of texts collected in a specific discipline, register or discourse. They are usually developed in specific contexts with a specific aim, and thus they are smaller in size and do not represent a language as a whole. They, as Lee (2010) observes, “do not aim to comprehensively represent a language as a whole, but only specialised segments of it” (p. 114). Tognini-Bonelli (2010) points out that the texts that are compiled in a corpus do not aim for representativeness; they are rather chosen “for their extraordinariness (p. 22). The Michigan Corpus of Academic Spoken English (MICASE) is a good example of a specialised corpus, consisting of over 1,800,000 words derived from spoken discourse of meetings, dissertation defences and lectures. The British Academic Written English (BAWE), another example to specialised corpus, is a collection of student writings and consists of almost 7 million words. The British Academic Spoken English (BASE) Corpus consists of data of over 1,600,000 words from lectures and seminars. An example of a specialised corpus created to represent the language of a specific discipline is the Honk Kong Engineering Corpus, consisting of more than 9 million words.

2.3.3. Learner Corpora

Learner corpora are specialised collections of language used by learners of English as a foreign or second language (Granger, 2002). The International Corpus of Learner English (ICLE) and the International Corpus of Crosslinguistic Interlanguage are examples of learner corpora.

2.3.4. Historical corpora

Historical corpora, as its name suggests, are collections of language from historical periods which provide data for examining the changes occurring in the language (Szudarski, 2017). The Corpus of Historical American English (COHA), for instance, is a historical corpus consisting of 300 million words from texts from 1800 till the present day. A Representative Corpus of Historical English Registers (ARCHER) is also a historical corpus created upon data from between the years 1600-1900.

2.3.5. Parallel corpora

Parallel corpora can be defined as “two (or more) corpora in different languages, each containing texts that have been translated from one language into the other” (Hunston, 2002, p. 15). With such corpora, it is intended to make comparisons between the same texts produced in different languages, and therefore, they are widely used by translators or scholars in the field of translation studies. The Oslo Multilingual Corpus, for instance, is a parallel corpus, comprising of source texts written in German, French and Finnish, and their translations.

2.3.6. Comparable /Multilingual corpora

Comparable or multilingual corpora are also a similar kind of parallel corpora. They contain data from texts written in different languages, and are employed to conduct cross-linguistic analyses. As Hunston (2002) notes, they can be used "to identify differences and equivalences in each language" (p. 15).

2.4. English for Specific Purposes

Together with the changes that came about following the World War II, various specific needs for learning English appeared. The great increase in scientific and technical activity on a global scale resulted in a need for an international language (Hutchinson & Waters, 1987). Dudley-Evans (1991) thinks that it was understood by the international community that learning English was important not only for transmitting knowledge and communication but also for international communication. The developments in technology, commerce and economics brought about different purposes for learning English. As noted by Hutchinson & Waters (1987), with English becoming the international language, a new generation of learners who knew the reason why they were learning a language emerged. Nunan (2004) explains the emergence of ESP as follows:

> *The basic insight that language can be thought of as a tool for communication rather than as sets of phonological, grammatical and lexical items to be memorized led to the notion of developing learning programs to reflect the different communicative needs of disparate groups of learners. No longer was it necessary to teach an item simply because it is 'there' in the language. (p.7)*

According to Nunan, this perspective resulted in the emergence of ESP as a subcomponent of language teaching which has its “own approaches to curriculum development, material design, pedagogy, testing and research” (p. 7).

English for Specific Purposes is defined by Hyland (2007) as “language research and instruction that focuses on the specific communicative needs and practices of particular social groups” (p.391). Hutchinson and Waters (1987) describe ESP as an approach of language teaching where all of the decisions regarding the content and method are based on the reasons why learners learn the language. Strevens (1998) define ESP as a particular case of the general category of special-purpose language teaching. He makes a distinction between absolute characteristics and variable characteristics of ESP. From his point of view, ESP consists of English language teaching which is:

- designed to meet specified needs of the learner;
- related in content (i.e. in its themes and topics) to particular disciplines, occupations and activities;
- centered on the language appropriate to those activities in syntax, lexis, discourse, semantics, etc., and analysis of this discourse;
- in contrast with General English.

In terms of variable characteristics, he states that ESP may be, but is not necessarily:

- restricted as to the language skills to be learned (e.g. reading only);
- not taught according to any pre-ordained methodology (pp.1-2)

Carver (1983) identified three characteristics of ESP courses; he suggests that the main features that are common to ESP courses are “authentic material”, “purpose-related orientation” and “self-direction”. With authentic material, he refers to content not developed for a teaching but taken from the main field of study or work. In that respect, Gatehouse (2001) also states that “Use of authentic content materials, modified or unmodified in form, are indeed a feature of ESP, particularly in self-directed study and research tasks” (p. 4). She defines “purpose-related orientation” as simulating various tasks that with the purpose of preparing learners for various target situations. The third characteristic "self-direction" means that learners should have some degree of freedom in deciding what, when and how to study (Gatehouse, 2001, p. 5).

Rahman (2015) notes that an important characteristic of ESP is needs analysis that identifies the language skills that the learners mostly needs and according to which the syllabus is designed (p. 24). Dudley-Evans & St John (1998) also state that "needs analysis is the cornerstone of ESP" and when a needs analysis is practiced properly, the findings will result in a "focused course" (p. 121). Strevens (1988) describes the significant components of an ESP course as follows:

1. conducting needs analysis
2. designing an appropriate syllabus
3. preparing suitable materials
4. meeting and getting to know students
5. teaching the course
6. devising and administering appropriate tests

Dudley-Evans and St John (1998) mention that "the main concerns of ESP have always been, and remain, with needs analysis, text analysis, and preparing learners to communicate effectively in the tasks prescribed by their study or work situation" (p.1).

Regarding the target audience of the ESP courses, Dudley-Evans (1998) states that "ESP is likely to be designed for adult learners, either at a tertiary level institution or in a professional work situation" (p.6). As for the proficiency level of students to take the ESP course, he explains that the learners generally have an intermediate or advanced level of English, and that "most ESP courses assume some basic knowledge of the language system" (p. 6).

2.5. Specialised Corpus and ESP

According to Swales (1990), one defining characteristic of a discourse community is "specific lexis" (p.26), and such specific lexis needs to be acquired by the prospective members of that community, like the ESP students. The specificity of the field also entails discipline-specific vocabulary. Widdowson (1998) states that specific lexis can be a problem for ESP language learners.

One criticism levelled at corpus linguistics is that it depends on general-purpose corpora composed of decontexualised data (Szudarski, 2017). Specialised corpora, which are smaller collections of words appear to have the power to close such a gap. Koester (2010) argues that specialised corpora "allow a much closer link between the corpus and the contents in which the texts in the corpus were produced" (p. 67). This implies that specialised corpora can be more suitable for language analysis in specific fields. Gavioli (2005) describes specialised corpus as one "designed in such a way as to collect a sample of a sub-language" (p. 60). It often represents a limited portion of specialised texts such as chemistry lectures, medical articles etc.

According to Paltridge (2012), corpus work has a significant role in exploring the use of language for specific purposes. Coxhead (2012) believes that vocabulary plays a central role in ESP and explains that teachers and learners should know that their classroomtime is directly connected to their needs. He states that "they should be reading material that contains key ideas and the language of their field and writing using those ideas and language" (p. 116). Webb and Nation (2013) think that "identifying the technical vocabulary is useful because it sheds light on the low-frequency words which may be of greater value to learners with specific academic purposes" (p.3). Corpora can play a significant role in identifying specific vocabulary of a specific discipline as "corpus-based studies allow for larger-scale investigations of words in context" (Coxhead, 2012, p.118). Chung (2003, as cited in Chung and Nation, 2004) for example, studied technical vocabulary in anatomy and applied linguistics and compared the frequency of words between two corpora. He found that 31.2% of the words in the anatomy texts were technical while in applied linguistics texts 20% of the words were technical. This shows that academic disciplines also show variance as to their reliance on technical vocabulary.

Szudarski (2017) notes that corpus-based research into ESP context does not only focus on the use of individual words but also phraseological chunks and longer patterns (p.141). On this issue, Gilguin et al. (2007) observe that academic discourse contains highly specific phraseology. Corpus linguists have been concerned with determining the most frequent chunks and exploring their role in the structure of academic texts (Szudarski, 2017). One of this type of chunks are lexical bundles (or n-grams), which can be defined as “two or more words that repeatedly occur consecutively in a corpus” (Cheng, 2012, p. 72). Biber et al. (1999, 2004) define lexical bundles as word sequences that appear together in natural discourse and believe that they are an important feature of academic texts.

Martinez and his colleagues’ study (Martinex et al., 2009) shows how specialised corpora can be used to identify field-specific vocabulary. The study aimed to investigate the frequency of academic words from Coxhead’s (2000) AWL in a corpus of research articles written in the field of agriculture. The results showed that the Agricultural corpus (namely AgroCorpus) contained merely ninety-two word families from the AWL. This indicates that the discipline-specific vocabulary does not always coincide with the academic vocabulary collected from several disciplines. The study also shows that in ESP contexts it is critical to consider the specific lexical needs of learners as you prioritize the types of vocabulary that they should know (Granger, 2015).

According to Gavioli (2005), in the domain of ESP, “what” is taught is a very critical problem for language teachers as they are not specialised in the discipline in question and they have to find a “linguistic path” to reach language conventions and concepts required in that discipline (p. 14). Also, learners may need to realise the salient features of the discourse of the field they are to study. In that respect, a specialised corpora focusing on the special language features of a specific domain, rather than a general one like the BNC, is more convenient. Specialised corpora focus more on a set of specific topics, which makes it a more reliable instrument for describing the language of specific domains. Gavioli (2005) notes that the necessity of focusing on actual language occurrence and analysing it has been considered a main issue in ESP descriptions and this might be stemming from the fact that the “linguist” alone may not have enough specialisation to describe the features of ESP (p. 55).

The degree of specialisation of a corpus is likely to vary and is widely determined by the purpose of the research. Although Sinclair-the father of Corpus Linguistics- thinks that a corpus must be as large as possible, small corpora have their own merits. According to Koester (2012), smaller, more specialised corpora have a distinct advantage: they allow a much closer link between the corpus and the contexts in which the texts in the corpus were produced (p. 67). He also believes that specialised corpora may be in different sizes but such corpora don't have to be as large as more general corpora to produce reliable results. The size of the specialised corpora also depends on the research purpose. Cheng (2012) observes that such corpora "can usually be measured in the thousands or low millions of words", but the purpose for which they are created is more important than the size (p. 166). Handford (2010) points out that some specialised corpora, such as the Cambridge and Nottingham Business English corpus, consist of a million words, yet others are much smaller and consist of less than 100,000 words. Another example is Koester's (2006) corpus of American and British Office Talk, which comprises 60,000 words.

Barber's study (1962, as cited in Gavioli, 2005) has been influential in the foundation and development of ESP studies. His study is based on a corpus of about 23,000 words, formed by research articles in different scientific domains, namely engineering, chemistry and astronomy. Barber manually calculates data such as average sentence length, the number of clauses per sentence, occurrence of most frequent modal verbs and occurrences of different modal verbal tenses and aspects. As noted by Gavioli (2005) many specialised corpora are compiled for teaching purposes rather than research purposes, and thus are more restricted in the variety of the texts they contain (p.62). For example, in his study, Flowerdew (1993) uses a corpus of 100,000 words consisting of biology texts and lecture transcriptions. Gledhill (2000) uses a 500,000 words corpus composed of medical papers on cancer research. The main aim of such studies were not to represent the ESP domain; they rather aimed to provide a sample which would be suitable to the needs of the students (Gavioli, 2005).

It is generally accepted that specialised corpora can reflect the features of the language of a sub-domain more effectively than general corpora do. By way of illustration, the word "transaminase" occurs 61 times in a small corpus of research articles that deal with hepatitis where as it occurs 17 times in the BNC (Gavioli, 2005, p. 63). However, such small-sized specialised corpora can be too small to depend on to make generalisations about that specialised language. They may provide more instances of technical words, yet it is not possible to ensure that those words or expressions characterise the language of the domain represented in the corpus. (Gavioli, 2005). To be able to notice the characteristics of a corpus compiled from texts that belong to specific register, a comparison of the features derived from that corpus with the features of other registers or genres is required. In that respect, Biber et al. (1998) argue that "register analyses require a comparative approach: we need a baseline for comparison to know whether the use of a linguistic feature in a register is rare or common" (pp. 136-7). Such a comparison is possible owing to the increasing availability of large general corpora such as BNC, COCA, and the Cobuild Direct.

2.6. Use of Corpus Data in Language Teaching

Corpus linguistics has direct and indirect applications in language pedagogy (Römer, 2011; Flowerdew, 2009). In indirect applications, corpus data constitute a basis for development of syllabuses, teaching materials and tests. In direct applications, on the other hand, corpus data is used to engage learners actively in the learning process through hands-on activities. This kind of approach, called data-driven learning (DDL), introduced by Tim Johns in 1990 (Boulton, 2012), posits that language learners explore the naturally occurring language themselves and use the data to come up with generalisations about language use. As the focus of this study is not on the learner aspect, indirect applications of corpus data are to be covered more widely.

Kennedy (2004) notes that corpus linguistics can contribute to language teaching in the selection of those features which seem worth teaching in a specific pedagogic context. Flowerdew (1993) also thinks that processing of data obtained from corpora can "provide criteria for: (a) the selection and grading of items for the syllabus, and (b) the authentic contextualization of these items in learning materials." (p.231).

Sinclair and Renouf (1988) introduced the idea of a lexical syllabus which is created based on the frequency of occurrence. They claim that the teaching process should focus on “the commonest word forms in the language, the central patterns of usage and the combinations which they usually form” (p. 148). This does not imply that the grammar aspect should totally be ignored. They suggest that “if the analysis of the words and phrases has been done correctly”, a teaching approach that focuses on the most frequent items will enable students to learn not only lexis but also grammatical structures and their functions. In other words, grammar is also learnt in the form of lexical patterns. Willis (1990) states that “what is traditionally termed grammar can often be called patterns” (p. 51). Michael Lewis, who is considered as the father of lexical approach, rejects the traditional grammar-lexis dichotomy and perceives language as grammaticalised lexis. Meunier and Reppen (2015) think that corpus-informed teaching can focus on the lexico-grammatical and contextual aspects of language use. They state that “corpus information on registers, frequency, and lexical preferences is key to a good understanding and use of grammar” (p. 510). Despite these views in favour of a lexical syllabus, it did not become a widely adopted language teaching approach (Szudarski, 2017). However, it demonstrated that corpus analysis findings should inform decisions in syllabus design. It also paved the way for lexis-oriented teaching materials which contribute to the promotion of vocabulary in applied linguistics as a whole. Willis and Willis (1988) applied the idea of a lexical syllabus and created a language course based on it, for the first time. The course, called the Collins COBUILD English, intended to cover the first 2,500 most frequent words spread across three consecutive proficiency levels (Szudarski, 2017).

As well as helping to determine the lexical content of a course by providing a total number of words to be taught in a course, frequency data can also help to decide which lexical items should be prioritised. As Flowerdew (1993) suggests, frequency data has an important application which is establishing the relative importance of vocabulary items and thus offering criteria for syllabus selection and grading. Additionally, he points out to the role of frequency data in programme evaluation. He believes that the discrepancies that result from the comparison of the frequency list extracted from the specialist corpus with that of the authentic corpus can constitute basis for evaluation and revision of an existing course. Frequent authentic language use can constitute a sound basis for any decisions regarding curriculum/syllabus design and assessment.

2.7. Corpus terms

In the process of corpus analysis, it is important to understand several terms related to the scope of the analysis, what is searched for and how the findings are to be revealed. In the following sections, a few corpus-related terms, namely frequency, concordancing, keyness, and n-grams are explained.

2.7.1. Frequency

Frequency refers to the number of times that a certain linguistic item occurs in a given context. High-frequency words are commonly acknowledged as the vital starting point for L2 vocabulary learning (Nation, 2013; Schmitt, 2010). The idea that teaching of words should be prioritised according to how frequently they occur in the target language is supported by Mackey (1965). He states that "since items occurring the most frequently are those which the learner is more likely to meet, they are the ones which are selected for teaching" (p. 177). Vocabulary items that occur recurrently in a source can be identified by means of a frequency analysis, which is the most basic kind of analysis that can be conducted on corpus data. The results of the analysis can be used to create wordlists, comprising of words or phrases that are ranked according to their frequency. It is possible to search for a specific word or phrase - called node- to obtain the number of its occurrences in the whole corpus. As highlighted by Tognini-Bonelli (2010), frequency "takes pride of place" in corpus research due to the fact that it provides the basis for all kinds of analyses (p. 19). Analysing texts as a whole set of data provides us with insights into the naturally occurring language. Also, as noted by Szudarski (2017), "empirically derived statements about tendencies in language are more accurate than those based on speakers' intuitions" (p. 51).

Szudarski (2017) points to an important point as regards comparing the frequencies of words or phrases across data set. He suggests "it is essential that you use normalized frequency – that is, the frequency of a word per million words- when you compare the occurrence of words in different corpora" (2017, pp. 20-21). The reason is that corpora can vary in terms of size. To illustrate, written corpora are larger than spoken corpora, which means that raw frequencies derived from these two datasets would yield inaccurate results.

Wordlists that are generated through frequency analyses can be used to make comparisons between different genres, modes of communication or discourse. Szudarski (2017) states that comparing wordlists, for example, derived from spoken and written texts “can provide valuable insights into how the use of vocabulary varies depending on the specific modes of communication” (p. 23). Another useful application of such wordlists is in the decision-making process of syllabus designers, material developers and teachers. The contents of a course can be determined and the items to be prioritised can be identified based on the frequency wordlists.

Wordlists usually contain information on type/token ratio which is an indicator of the degree of lexical diversity of texts. *Types* refer to all unique words in a corpus, and *tokens* refer to the same words repeated throughout the corpus.

2.7.2. Concordance

A concordance is a list of occurrences (all or a selected number) of a word or a phrase in a corpus (Gavioli, 2005). The occurrences are shown on the screen by the search word in the middle and other words in the contexts on both sides of it, which is called KWIC format. KWIC stands for “key word in context” and recurrent combinations of a word are highlighted. In the KWIC format of a concordance list, it is possible to see the word that is searched for with various usages. Through the collocates display option, it is possible to see what words occur near other words, which provides insights into the meaning and usage of a certain item. One can also conduct a simple search through the list option which provides frequency values for individual words.

The following is an extract of the concordance of the word “view” in KWIC format in the British National Corpus compiled by Mark Davies, .

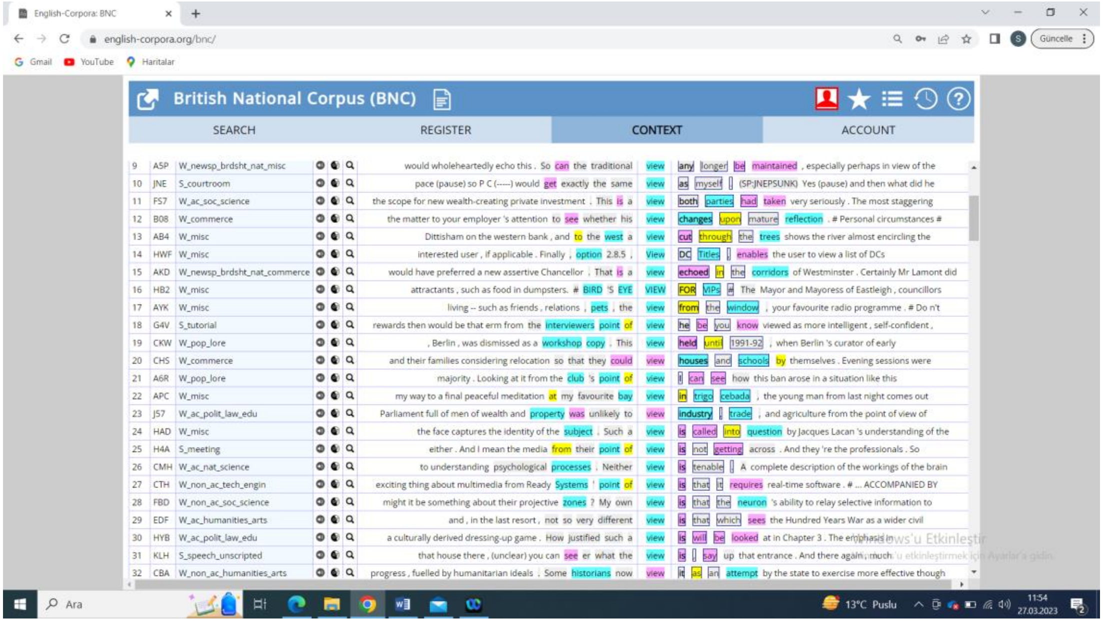

Figure 1. An example of concordance from BNC

Concordances show us how specific lexical items are used in real contexts. Through concordancing, it is possible to determine which uses of items are to be taught in a course. This, as stated by Flowerdew (1993) "reduces those uses of a given item to be presented for learning to those which actually occur in the corpus, eliminating time-consuming attention to other uses pointed to by dictionaries and reference grammars" (p. 238). Furthermore, it is possible to see the syntactic patterns in which lexical items occur, which may be valuable in the decision making processes of a syllabus design. Concordances of specific words in a list reveal how they are used in the corpus, providing information regarding the characteristics of the specialised language, the ways these words are typically used as well as their frequent collocations. Gavioli (2005) states that "grammatical items such as connectors may also be used in characteristic ways in specialised corpora and some frequent verbs may be indicative of typical syntactic structures" (p. 24).

Concordance analysis can also be useful in dealing with problematic areas in language teaching. Gavioli (2005) notes that condordancing has been found to be of great help in supporting teachers in areas which are traditionally considered "difficult to deal with" and where descriptions provided by grammars and/or dictionaries seem inadequate (p. 25). For example, some features of spoken English such as ellipsis or tails (Carter & McCarthy, 1995) and some discourse markers (Zorzi, 2001) which have traditionally received little attention in pedagogic grammars can be delved into through concordancing. A study by Partington (1998) highlights a number of language teaching problems that can be dealt with effectively through corpora-based instruments and concordancing tools. He shows that concordances produced from a corpus of newspaper texts support teachers' intuitions by providing descriptions from dictionaries and grammars, and more satisfactory explanations.

Johns (1994) worked on using concordances in the classroom. He used corpora of texts from scientific and technical magazines such as *New Scientist*, *Nature* and from newspapers such as *The Times, The Guardian* and generated concordances which focused on grammatical points that caused problems for learners. He delivered the data he collected to the students and wanted them to analyse it. As a result, he found out that the students comprehended the meanings and functions of the structures presented in the data better than they did when presented traditionally in list of verb patterns. Working out the grammatical features of the language, students took an active role in the learning process. John's work was the first real attempt to look at corpus concordancing from the point of view of the learner (Gavioli, 2005).

Be it for the benefit of the learner or the teacher/syllabus designer, concordancing data can make a significant contribution to the field of language teaching, which cannot be undervalued. Some concordancing software available on the market are WordSmith Tools, MonoConc Pro, Sketch Engine, and Antconc.

2.7.3. Keyness

Scott (1997) defines keyword as "a word which occurs with unusual frequency in a given text" (p. 236). Evison (2010) states that keywords are "those words which are identified by statistical comparison of a target corpus with another larger corpus, which is referred to as the 'reference' or 'benchmark' corpus" (p. 127). Szudarski (2017) suggests that "keywords provide you with a window into the distinctiveness or uniqueness of data that are found in your target corpus" (p. 236). He also points out that keywords can have either much higher frequency in a target corpus in comparison with a reference corpus or much lower frequency in a target corpus. According to Gabrialetos (2018) keyness analysis is basically a comparison of frequencies and it aims to identify large differences between the frequency of word-forms in two corpora (usually referred to as *study* and *reference* corpus). Scott (1997) posits that a word is considered as "key" if its frequency in a text when compared with its frequency in a reference corpus is such that the statistical probability is smaller than or equal to a *p* value specified by the user.

2.7.4. N-grams

N-gram is a term that is used to refer to combinations of words comprising of "two or more words that repeatedly occur consecutively in a corpus" (Cheng, 2012, p. 72). Greaves and Warren (2010) explain that "N-grams, which have attracted a variety of labels such as 'lexical bundles', 'chunks' and 'clusters', are frequently occurring contiguous words that constitute a phrase or a pattern of use (e.g. you know, in the, there was a, one of the)" (p. 213). N-gram analysis is "a purely frequency-driven approach which explores patterns of lexical co-occurrence without considering semantic and syntactic relationships between particular words" (Szudarski, 2017, p. 25).

2.8. Corpus Statistics

Corpus studies make use of some statistical tests to interpret the data obtained from analyses, due to their quantitative nature. Such tests identify if there is any statistically significant difference between datasets, through which the researcher can come to viable conclusions as to the research questions.

2.8.1 Log-likelihood

Log-likelihood test is used to compare differences in frequency values. Szudarski (2017) explains log-likelihood test as a test that "helps you determine whether differences in the frequency of words are reflective of the actual variation in the language or whether they result from chance occurrences" (p. 27). It can be used as a rapid way of determining if a difference is statistically significant. According to Rayson and Garside (2000) log-likelihood which is higher than 3.84 shows a significant difference between the two sets of data.

2.8.2. T-score/Mutual information

T-score and mutual information tests tell us whether the co-occurrence of words has statistical significance or whether it is a chance occurrence. MI scores higher than 3 and T-scores higher than 2 are regarded as thresholds which show a significant association between two words (Hunston, 2002).

2.8.3. Type-Token Ratio

Type-token ratio (TTR) is a measure of the lexical diversity or richness of texts (Szudarki, 2017). Cobb and Horst (2015) define lexical richness as "the level of development of a learner's lexicon" (p.189). It is possible to calculate a TTR by dividing the number of types (all unique words) by the number of tokens (repetition of the same words) in a given text or corpus and percentage is the usual way of reporting it.

2.9. Vocabulary and language use

Vocabulary is a critical component of language use. Schmitt (2010) points to the role of vocabulary in language proficiency and the high correlations between measures of vocabulary and language proficiency. Meara (1992) also states that learners with larger vocabularies are more proficient in language proficiency compared to those with smaller vocabulary (p.6). Lexical coverage refers to the percentage of known words in a text (Laufer and Ravenhours-Kalovski, 2010; Van Zeeland and Schmitt, 2013). It implies that the extent to which we can comprehend a text is influenced by the number of words we know. For instance, Schmitt et al. (2011) suggest that for a 60% comprehension to be reached, 98% lexical coverage is necessary and this corresponds to 8-9,000 word families (Nation, 2006). In Nation's study (2006) it was found that 8-9,000 word families for written language, and 6-7,000 for spoken language provide sufficient lexical coverage for a good understanding of L2 texts. When it comes to the vocabulary size of native speakers, this number is estimated to be about 20,000 word families or 32,000 vocabulary items.

Milton et al. (2010) show that vocabulary is highly influential on the overall language ability. They found positive correlations between ESL learner's vocabulary and their IELTS scores. Staehr (2008) also found evidence as to the contribution of vocabulary size to language proficiency. He found that EFL learners' vocabulary size was highly correlated with their reading, writing and listening (to a lesser extent) performance.

2.9.1. Vocabulary-related terminology

'Vocabulary', 'words', 'lexis' are widely used in the field of applied linguistics to refer to the same concept. However, it may be important to note the distinctions between these concepts. Carter (2004) defines "word" as a sequence of letters bounded on either side by a space or punctuation mark (p. 35) but 'vocabulary' and 'lexis' cover all the lexical elements in a language. According to Scrivener (2005), lexis can be more specific as it covers not only individual words but also different kinds of combinations between words. The term 'lexicon' also refers to a collection of all words in a given language (Cheng, 2012).

Some vocabulary-related words in the area of corpus linguistics are: 'word form', 'lexeme', 'lemma', and 'word family'. The term 'word form' is often used to refer to different realizations of one 'lexeme' (Szudarski, 2017, p. 6). Carter (2012) defines 'lexeme' as an abstract unit which underlies different grammatical variants of a word. To illustrate, the lexeme 'break' can be realized by different word forms such as 'broke', 'broken' and 'breaking'. Francis and Kucera (1982) define 'lemma' as "a set of lexical forms having the same stem and belonging to the same major word class, differing only in inflection and/or spelling" (p.1). Lemma is the form of a word that appears at the beginning of a dictionary. Szudarski (2017) explains lemma as base forms together with their inflected forms (e.g. all of them are nouns or verbs) (p.36). The term 'word family' has a broader sense than a lemma. It includes a base form of a word as well as its inflected forms and transparent derivatives (Bauer and Nation, 1993; Coxhead, 2000). By way of illustration, a lemma encompasses all the inflected form of the verb ‘employ’: ‘employs’ ‘employed’ and ‘employing’, whereas a word family additionally includes other word classes like nouns (‘employment’, ‘employer’, employee’) and adjectives (‘unemployed’).

2.9.2. Vocabulary and frequency data

As pointed out previously, findings from frequency analyses can provide valuable insights into the language. A frequency analysis can be carried out to compare vocabulary in two modes of communication: spoken and written. Each of these modes of communication have certain characteristics. For example, spoken communication is more spontaneous and interactive while written communication is more planned. Frequency-based corpus analyses allow us to explore the discrepancies and identify the recurrent features of language in these two modes. Also, frequency analysis can provide information on the frequency of different word types. For example, data from COCA was used to create a word list of the most frequent words in English and the words with the highest frequency on the list are function words such as prepositions or articles (Szudarski, 2017).

Corpus data can also provide insights into lexical coverage. According to Nation (2011), high-frequency words are “a relatively small, very useful group of words that are important no matter what use is made of the language” (p. 531). Nation (2006) also observes that high-frequency vocabulary accounts for the largest amount of text. O’Keeffe et al. (2007) suggest, the first 2,000 or so word-forms cover more than 80% of all the words in spoken and written texts.

Schmitt and Schmitt (2014) state that mid-frequency vocabulary consists of approximately 6000 word families that are between the first 3,000 and 9,000 most frequent words in English. Low-frequency vocabulary, which refer to words that are beyond the 9,000 frequency threshold, are considered to be “restricted to certain subject areas” (Nation, 2011, p. 531). Low-frequency vocabulary is often regarded as specialised and technical words required to understand specialised texts (Szudarski, 2017). Such words are usually identified by comparing wordlists from general and specialised corpora.

O'Keeffe et al. (2007) make a distinction between basic and advanced vocabulary, relying on corpus data. Basic vocabulary refers to the first 2,000 most frequent words and they are needed in everyday language use. However, defining advanced vocabulary seems to be more problematic and determining the boundaries of this category might involve some arbitrariness (Szudarski, 2017). To avoid this problem, O'Keeffe et al. (2007) resort to corpus-based estimates of lexical coverage as a benchmark and use a frequency-based criterion for advanced vocabulary. They argue that a receptive mastery of 5-6,000 words seems to be a border between the intermediate and advanced level of proficiency.

Frequency information can be employed by teachers in designing the teaching process. High-frequency words or phrases that account for a high percentage of L2 use can be prioritised. As noted by Szudarski (2017), "by using corpus findings, teachers and language practitioners can ensure that the limited classroom teaching time is devoted to the promotion of those lexical items that provide a good return for the learning effort" (p.66).

2.9.3. Formulaic language and corpora

A large part of language consists of units longer than single words. Szudarski (2017) states that "once we start exploring large amounts of naturally occurring data, we quickly discover that words have a tendency to cluster with one another and form lexical collocations" (p. 72). As Sinclair (1991) notes, "most everyday words do not have an independent meaning, or meanings, but are components of a rich repertoire of multi-word patterns that make up a text" (p. 108). Tognini-Bonelli (2010) also highlights the fact that the patterns of lexical repetition and co-selection are an important aspect of language use. Phraseology deals with analysing the structure and occurrence of multiword units and it constitutes a major stream of research in corpus linguistics (Szudarski, 2017).

Formulaic language can be thought of as an umbrella term that encompasses all types of phraseological units. Wray (2002) introduced the term ‘formulaic language’ in her book on the role of phraseology in language. Wray (2002) points to the importance of formulaic language as:

> *a sequence, continuous or discontinuous, of words or elements, which is, or appears to be, prefabricated, that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. (p.9)*

Moon (1997) defines a multiword unit as “a vocabulary item which consists of a sequence of two or more words which semantically and/or syntactically forms a meaningful unit” (p. 43). There are different types of multiword units. Collocations are pairs of words that are commonly found together. Firth (1957), to whom the term collocation is often attributed to, observed that the meaning of a word was shaped not only by what it possessed itself but by the way it is combined with other items. Halliday (1966) also did remarkable work on word partnerships and he defines collocations as syntagmatic relations between words the probability of occurrence of which can be measured. He suggested using large samples of data to identify the words that co-occur in a regular pattern. While collocations denote lexical partnerships, colligations can be described as the “co-occurrence of grammatical items” (Szudarski, 2017, p. 80). Cheng (2012) states that analysis of colligations “requires the analyst to operate at a level of abstraction” because it is concerned with exploring patterns of co-occurring words in relation to grammatical categories and structural relationships (p.82).

Sinclair’s work on the COBUILD corpus and the COBUILD Dictionary has shed new light on research in language use and placed phraseological patterning at the heart of linguistic analysis (Szudarski, 2017). Lexico-grammar constitutes a core position in corpus-based investigations of phraseology. According to Sinclair (1987) co-selection of words creates the meaning. For Sinclair, studying multi-word units of meaning has a cenral role in linguistics.

Although the importance of collocations was stressed in the 1960s, (Halliday 1966; Sinclair, 1966) the study of multi-word units has become widespread only recently

(Greaves and Warren, 2010). The term n-grams may be used to refer to commonly co-occuring words that form a pattern, and according to Greaves and Warren (2010), most of the studies done on multi-word units in the form of n-grams has an inclusive approach and keep all the recurring co-occuring words in their lists of data. A number of studies have focused on how multi-word units are used specifically in certain registers and genres (e.g. Biber et al., 1999; Carter and McCarthy, 2006; Scott and Tribble 2006; Hyland 2008; Durrant 2015). Hyland (2008) investigated the differences in the use of four-word lexical bundles across four disciplines. His study showed that biology and electrical engineering fields relied more on research-oriented bundles, whereas business and applied linguistics relied more on participant-oriented chunks and text-oriented bundles (Szudarski, 2017). Scott and Tribble (2006) examined the phraseology that is used in different contexts using BNC and three sub-corpora within the BNC. They found that some patterns yield different frequencies across the four sub-corpora and that one specific pattern they investigated -*one of the*- occurred more frequently in the academic discourse. Carter and McCarthy (2006) investigated the use of n-grams in spoken and written discourse and they found that certain n-grams, for instance those expressing time and place relations, are more prevalent in written discourse and that n-grams relecting interpersonal meanings are more frequent in spoken discourse. Durrant (2015) also did research on the disciplinary variation in the use of four-word lexical bundles in university students' writing. He used data from the British Academic Written English Corpus (BAWE) and found distinctive patterns between hard and soft sciences. These studies, looking into the role of phraseology across specific disciplines, show that n-gram analyses can yield valuable insights as to the language features specific to different genres.

2.9.4. Word Lists

A number of corpus-based lists of high- frequency words have been developed such as General Service List (West, 1953), BNC2000 (British National Corpus 2000; Nation, 2006), COCA lists (Corpus of Contemporary American English list; Davies & Gardner, 2013), BNC/COCA2000 (Nation, 2012), New General Service List (Browne, 2014), and New General Service List (Brezina & Gablasova, 2015). Prior to the emergence of an academic word list, West's *General Service List* (GSL) (1953), which included 2000 most frequent word families, was widely used in language teaching and represented an average of "around 82 per cent coverage" of various types of texts (Nation & Waring, 1997, p. 15). Many graded readers and English language teaching materials were developed based on GSL.

In 1984, the *University Word List (UWL)* was published by Xue and Nation (1984), which contained 836 items found commonly in a variety of academic texts but not included in the GSL. The UWL is supposed to provide 8.5% coverage of academic texts (Nation & Waring, 1997). The list is divided into 11 sublists based on frequency. Coxhead's *Academic Word List* (AWL) was created in 2000 as a general-purposes academic word list, especially for reading, based on corpus research. It consists of 570 word families that are not included in the GSL but commonly occur in academic texts, across four disciplines: Arts, Science, Law and Commerce. The AWL has been used widely and has been an influential resource in the field of English language teaching. Gardner and Davies (2014) introduced a new *Academic Vocabulary List (AVL)* on the grounds that 79% of the AWL were also in the GSL. The text coverage of AVL was higher than AWL but Nation (2001) found that 40% of the top 500 words of the AVL are also in the GSL. This means the AVL includes high-frequency words which students most likely know. According to Webb and Nation (2017) the AVL is too big to be used in a language course as it contains about 3,000 academic words.

Given that there are several word lists, which list is the most useful for L2 learners is a question. Dan and Webb (2016) addressed this question by examining their lexical coverage, which refers to the percentage of words covered by items from a particular word lists in a corpus (Nation & Waring, 1997). The results of their study showed that Nation's (2012) BNC/COCA2000 wordlist accounted for the largest coverage; on the other hand, Brezina and Gablasova's (2015) New-GSL has the largest number of frequent items.

Despite the significance of these wordlists, their usefulness was not certain as the audience they appealed to was not homogeneous. Dang et. al (2022), in their latest study, also state that "although lexical coverage is an important criterion to evaluate corpus-based word lists, to make these lists more relevant to L2 learning and teaching, list evaluation should involve their end-users— learners and teachers" (p. 619). Hirsh (2004) found that academic subject areas featuring the highest amount of technical vocabulary use the lowest amount of general service vocabulary. As for the usefulness of the AWL, Lei & Liu (2016) notes that it varies significantly across disciplines in terms of range, frequency, collocation, and meaning. According to Yang, (2015) each discipline has its own conventions. This indicates the need for creating academic word lists for specific disciplines. The motive behind corpus-based studies to generate specialised word lists could be the conception that the available wordlists are far from representing the discourse of a specific discipline, which typically has its own conventions. Durrant (2016), in his study where he investigated the relevance of AVL (Academic Vocabulary List) to university student writing, found wide variance across disciplines, which was in line with previous research.

With the help of corpus linguistics, specialized academic wordlists were created in scientific disciplines. The *Science Word List* (SWL) (Coxhead and Hirsch, 2007) is an apt example of a specialized word list based on corpus data. Data from 14 disciplines, namely agricultural science, biology, chemistry, computer science, ecology, engineering and technology, geography, geology, horticultural science, mathematics, nursing and midwifery, physics, sport and health science, and veterinary and animal science, were compiled in a corpus (Pilot Science Corpus of Written Texts) and the SWL was created consisting of 318 word families which covered 3.79% of the Pilot Science Corpus. However, the list was also too broad. According to Biber (2006), the specialized vocabulary in natural science (i.e., biology, chemistry, mathematics, and physics) is different by nature from other scientific branches, which means that some words in the SWL may not be equally valuable for students from different science disciplines. Business Word List (BWL) created by Konstantakis (2007), the Medical Academic Word List (MAWL) created by Wang, Liang, and Ge (2008), and the Basic Engineering List (BEL) created by Ward (2009) are some other examples to subject-specific academic wordlists.

The following table lists several previous studies on discipline-specific wordlists with information on the inclusion of AWL words.

Table 4
Some studies on discipline-specific wordlists

Past research	Word list	Number of words	Coverage
Konstantakis (2007)	The Business Word List (BWL)	560, No AWL words	2.79 %
Liu and Han (2015)	Environmenal Academic Word List	458 with 318 from AWL	15.43%
Wang et. al. (2008)	Medical Academic Word List (MAWL)	623 with 342 from AWL	12.24%
Yang (2015)	A Nursing Academic Word List (NAWL)	676 with 378 from AWL	13.64%
Hsu (2014)	Engineering Word List (EEWL)	729 with 304 from AWL	14.3 %
Mudraya (2006)	Student Engineering Word List	1200	Not available
Ward (2009)	Basic Engineering List (BEL)	299	16.4 %
Coxhead& Hirsh (2007)	Science Word List	318, no AWL words	3.79 %

2.10. Relevant Studies

Studies that dealt with the issues of vocabulary and corpus data provide insights into the vocabulary behaviour profiles in authentic contexts. Below are a few studies focusing on the aspects of vocabulary and corpus data.

2.10.1. Flowerdew's study (1986)

Flowerdew's (1993) early work presents an application of a corpus concordancing in the field of ESP course design. The rationale behind the study is that computer processing of data obtained from corpus can provide basis for the selection and grading of the items on a syllabus as well as integration of these items into the materials in an authentic way.

Science students taking the foundation course at Sultan Qaboos University (SQU) take science and English courses. The researcher created a corpus from the written and spoken input the science students were exposed to. The corpus was then analysed using an in-house frequency-concordancing programme developed at SQU.

Flowerdew found out that the 10 most frequent items in the specialist corpus and reference corpus, the Cobuild general corpus, were all grammatical words but the difference in their order was remarkable. For example, while "was" was the tenth most frequent item in the general corpus, it was fiftieth in the specialist corpus, which may have implications for the syllabus design. Also, there was significant variation between Cobuild and the specialist biology corpus in terms of vocabulary. The following table displays the 20 top nouns in a general corpus (the Cobuild Corpus) and in Flowerdew's biology corpus (1993, p. 236).

Table 5

Top 20 nouns in the Cobuild corpus and in a biology corpus

Cobuild corpus	**Biology corpus**
time, people, way, man, years, work, world, thing, day, children, life, men, fact, house, kind, year, place, home, sort, end	cell, cells, water, membrane, food, plant, root, molecules, plants, wall, energy, concentration, organism, cytoplasm, animal, stem, structure, body, part, animals

Such a list may constitute a basis for decisions taken in designing a curriculum or syllabus in terms of what to include in the lexical part of the programme. An interesting consequencye of this study is that contrary to the common belief that the specialist corpus would mostly consist of technical words, the majority of the items are neither technical nor general. In other words, they are words in general usage, but they have a special meaning in technical context (Inman, 1978). Some words that fall into this category are as follows: *wall, energy, concentration, structure, body and animal.*

Another remarkable finding of the study is about the frequency of the connectors. A small number of connectors – "*so", "then", "first", "next*"- are frequently used; others are less frequent, and some do not appear at all. This finding is also valuable in rank-ordering the teaching of certain concepts and prioritising the items to be taught. Considering the fact that a great amount of time might be spent on an item that is hardly ever used in the target context, it is reasonable to use such information to constitute the basis for designing the content of a course so that less time and effort is spent attempting to teach a useless item.

Table 6 below shows the number of occurrences of certain connectors in the corpus:

Table 6

The number of occurrences of certain connectors in a biology corpus (1993)

Connector	The number of occurrence in the specialist corpus
So	1183
Then	266
First	103
Next	72
However	13
Therefore	11
Thus	8
Finally	8
As a result	4
What is more, furthermore, nonetheless, nevertheless, hence, consequently, in conclusion, in contrast, after that	-

In Flowerdew's study (1993) concordancing has demonstrated discrepancies between the specialist corpus and the published materials. The first one concerns definitions. Definitions are commonly taught through a formula such as "X is/can be defined as . . ." [(e.g. Allen and Widdowson (1974) and Master (1986)] in a number of coursebooks in the market whereas the specialist corpus presents only one instance of the word "define". Instead, 417 instances of the word "called" are used in a defining function. Another variation appertains to the syntactic patterns. For instance, the connector "then" rarely occurs as sentence initial; it is rather found between subject and verb (1993, p. 238):

> "the viruses then do the same"
>
> "these goblet cells then secrete mucus"
>
> "the liquid is then discharged"

Another example concerns the passive structure. In many published materials passive voice is taught as "subject + auxiliary + past participle". However, the passive uses found in the specialist corpus have an adverbial between the auxiliary and past participle. Below are some examples:

> "Water is actively passed."
>
> "The nerve vells are also linked together."
>
> "The viruses are then released."

2.10.2. Ward's study (1999)

Ward aimed to construct a wordlist for engineering students which would provide them with the sufficient lexical knowledge to read texts in an attempt to question the claim that learners need a vocabulary of 3000 word families to be able to read effectively and that this vocabulary should be based on general words initially and then be built on by an academic and/or technical word list. Expressing reservations about the figure of 95 percent suggested as a threshold for reading comprehension (Laufer, 1989), and also pointing to the evidence (Nurweni and Read, 1999) that many engineering students are not able to reach this 3000-word level, Ward (1999) set out to investigate the lexical resources necessary to achieve this 95 percent by undergraduate engineering students who read their textbooks in English. He believed that developing an engineering wordlist as a short cut to reading fluency and providing the learners with the list at an early stage would be valuable. To this end, Ward used five extended texts from the first year courses that engineering students were required to take, from the subjects of engineering thermodynamics, engineering mechanics, fluid mechanics, statistic & provability and mechanics of materials and compiled an engineering corpus of 1 million running words, from which he extracted a list of 3000 word families. The engineering list which he referred to as *EngList* was run against a variety of texts selected from various academic disciplines. The first two thousand most common word families in EngList were remarkably different from the GSL, with 50 percent of the word types not occurring in GSL. The list provided high coverage of the texts which the list itself was derived from; however, when it was run against a different engineering mechanics text, not included in the engineering corpus, it yielded a coverage value of 96.9. The predictive value of the EngList (the two thousand most common words) was also better in all disciplines than the two thousand most common words of GSL. This difference was reported to be greater in scientific/technical disciplines.

Ward concludes that "a first-year engineering student may know 95 % of the tokens in many basic engineering texts with a vocabulary of only 2000 word families" noting that such a vocabulary will have a "technical flavour" but also include general words

(p. 321). He advises that students with an aim of reading engineering textbooks should start reading materials based on lists like the engineering list he developed.

2.10.3. Mudraya's study (2006)

In her study integrating the lexical approach with a corpus-based methodology, Mudraya (2006) developed a corpus of student engineering lexis, consisting of approximately 2,000,000 running words. The goal of her project was to construct a sound lexical syllabus for English teaching aimed at engineering students at Walailak University, Thailand. Collecting a total of 13 English language textbooks used in basic engineering disciplines, she compiled a corpus, which she called *SEEC* and created a word list of 1260 most frequent word families. The word frequency analysis showed that the most frequently occurring words in SEEC were sub-technical words, which are words with non-technical and technical senses, and non-technical words from the academic register. Mudraya concludes that more attention should be paid to academic English and sub-technical vocabulary and suggests a lexical syllabus with data-driven corpus-based methodology in ESP teaching.

2.10.4. Chen & Ge's study (2007)

Another corpus study conducted in ESP is Chen and Ge's (2007) study on the word frequency and the text coverage of the 570 word families from Coxhead's Academic Word List (AWL) in medical research articles based on a corpus of 50 articles written in English. The corpus was comprised of 190425 running words. To ensure the representativeness, objectivity and manipulability of the samples chosen, articles were selected from two journals from each category by random sampling. Since whole texts would provide more opportunities for words to reoccur and longer texts allow for more frequency of occurrence as well as variety of vocabulary (Coxhead, 1998; Stubbs, 2001), the articles chosen were kept at their original length with their tables, diagrams and bibliographies removed.

The basic word database was formed with the 570 AWL word items and the sub-sections of the research articles were input separately. The data was analysed through a self-designed computer programme. Among the 570 AWL word families, only 292 word families (51.23%) occurred more than 10 times in the corpus of medical research articles, and 111 word families (19.47%) appeared less than 4 times and 99 word families (17.37%) did not appear at all. There are some differences between the frequency of AWL words in Coxhead's corpus and the target corpus. A number of AWL words listed as most frequently used words in Coxhead's list did not occur as frequently in the target corpus and vice versa.

The study's findings that academic vocabulary has a rather high text coverage (around 10%) in medical research articles are in line with the results of Coxhead's study on academic texts across a wide range of subject disciplines. This indicates that academic words are indeed a set of important word items in medical research articles. However, out of the 570 AWL items, only 292 (51.2%) were found to be frequently used in medical research articles, which shows that the AWL list fails to represent an overall picture of the frequently used academic words in medical research articles. Some high frequency-items in Coxhead's corpus do not appear as frequently in medical research articles.

The authors suggest that a medical academic word list be created, which would meet the specific lexical needs of medical students so that they become proficient users of medical language.

2.10.5. Coxhead and Hirsch's study (2007)

Coxhead and Hirsch (2007) conducted a study to investigate whether a science-specific vocabulary outside the GSL and AWL words could exist. Their study is prompted by the need to make for the students' lack of specific lexical knowledge in their field of study, which academic staff often express their disappointment about. With the aim of determining whether there is a core of words occurring outside the GSL and AWL which are specific to scientific content, the researchers updated their existing science corpus by adding texts from agricultural science, ecology, horticultural science, engineering and technology, nursing and midwifery, sport and health sciences, and veterinary and animal sciences, which finally contained 1,761,380 individual words. Then a word list was developed by identifying all words occurring in the corpus outside GSL and AWL, making use of the criteria of range, frequency, and dispersion. They found that the "items in the pilot Science list give a better coverage over a written corpus of Science than items in the AWL", but this is true for "only after sublist 1 of the AWL" (p. 74).

Coxhead and Hirsch, note that, the pilot list is intended as a guide rather than as the only opportunity to come across these items and advise material designers to consider the necessity of providing meaningful contexts and rich learning opportunities when they are working with vocabulary lists. They also acknowledge that the study has some limitations, one of them being the small size of the corpus. They believe that a larger corpus would provide more representative samples of language.

2.10.6. Martinez et al.'s study (2009)

Martinex, Beck and Panza (2009) conducted a study on how specialised corpora can be used to identify field-specific vocabulary in the field of agricultural sciences. Their study integrates corpus-based and genre-based approaches (Flowerdew, 2005), analysing the research articles to uncover specific characteristics of academic vocabulary using the AWL as its point of departure. The study focuses on frequency, coverage and distribution of the words from the Academic Word List in agriculture research articles.

Aiming to investigate the frequency of academic words from Coxhead's (2000) Academic Word list in a corpus of articles written in the field of agriculture, Martinez and his colleagues collected 218 articles written by scholars who work in English-speaking universities. The corpus, which they called *AgroCorpus*, consisted of over 800,000 words. It was compiled based on the criteria of representativeness, use of whole texts, and availability of electronic sources. Some subsections of the articles such as abstracts, numbers, acknowledgements, references and appendices were excluded from the word count. They used the WordSmith tools for the analysis and identified the most frequent academic words in the corpus.

The results of the study showed that the AgroCorpus contained only ninety-two word families from Coxhead's academic word list. This shows that the agricultural vocabulary in the corpus was in contrast with Coxhead's (2000) corpus, which was comprised of more general academic vocabulary retrieved from different disciplines. Another important result of the study was that some of the most frequent words in the AgroCorpus were from West's (1953) General Service List. This can be indicative of the fact that some frequent words in English may be used as specialized words in academic texts. A comparison of the ninety-two word-families were compared with the most frequent word families from Coxhead's corpus revealed that only twenty-six items coincided and these were academic words of a more general nature like 'significant', 'analysis', 'data', 'area' and 'variation'. Martinez et al.'s study provides evidence for the idea that discipline-specific lexical variation is an important factor that must be taken into consideration when trying to identify academic vocabulary.

2.10.7. Brezina and Gablasova's study (2015)

In their recent study, Brezina and Gablasova investigated the overlap among four corpora, namely LOB, BNC, BE06, and EnTenTen12**,** in the top 3,000 words based on the average reduced frequency. They first created wordlists from the four corpora and then compared them pairwise. The comparison of the corpora showed that there is a stable core of 2,122 items among the corpora. Following the identification of a common lexical core among the four wordlists, the researchers extracted the shared items. One of the research purposes being the identification of the lexical items representing a recent development in the English language, the wordlists based on the two most recent corpora BE06- 3000 and EnTenTen12-3000 were compared, and the shared lexical items were extracted. Combining the common core lexical items with the current words reflected in BE06 and EnTenTen12, they compiled the new-GSL, as an up-to-date general service list derived from a large source of corpus data.

2.10.8. It-ngam & Phoocharoensil's study (2019)

It-ngam & Phoocharoensil's (2019) study is a recent one conducted with the purpose of exploring the specialized academic words across 11 sub-disciplines of natural science. They claim that it is necessary to know specialised words to comprehend scientific texts and that available word lists such as the General Service List (GSL) (West, 1953) and the Academic Word List (AWL) (Coxhead, 2000) do not cover all sub-disciplines of natural sciences (It-ngam & Phoocharoensil, 2019). The authors aimed to create a new academic word list for science disciplines, based on the data from journal articles of science disciplines, which would contribute to the design of an appropriate syllabus and constitute self-study material for students. To this end, they created the Corpus of Scientific Academic Journal, which they called SAJ corpus, comprising of 5.5 million running words from 1062 journal articles in science disciplines. The results show that while the GSL covers around 70% to 95% of most text (Gilner, 2011; Nation & Hwang, 1995), it provides 63% coverage in the corpus of scientific academic text. In other words, the SAJ corpus contains fewer general words than corpora of general texts (2019). The General Service List and the Academic Word List together provided a 73% of coverage in the SAJ corpus. The following table shows the coverage of different base word lists over the SAJ corpus.

Table 7

The coverage of base word lists over the SAJ corpus (It-ngam & Phoocharoensil, 2019)

Word lists	**Running words**	**% of SAJ**	**Headwords**
1st GSL	3,239,363	58.23%	994
2nd GSL	285,525	5.13%	898
AWL	531,119	10.09%	568
SAWL	323,611	5.82%	432
Off-list	1,155,034	20.76%	100,888
Total	5,562,996	100.00%	103,780

The authors conclude that the specialized academic word list (SAWL) created from this study provides high coverage of science English in research articles, it should be a good resource for students and teachers of science English, syllabus designers, and material developers. They also suggest that attention should be given to collocations used together with the SAWL words. Teachers should introduce how the SAWL words are used in the correct context. It is also necessary to keep in mind that the SAWL was built on the notion that the science students are familiar with the most commonly used words in GSL (West, 1953) and general academic words in AWL (Coxhead, 2000).

However, for low proficiency students, teachers might design their ESP courses that are accompanied by GSL, AWL, and SAWL. (2019, p. 664).

2.11. A brief summary of the review of literature

It is intended with the review of literature provided above to establish a basis upon which this research study can be built. In light of the pertinent research summarized so far, the following conclusions can be drawn.

Needs analysis is considered to be the starting point for any curriculum or syllabus development endeavor. In order to make reasonable decisions in designing syllabi/curricula or developing materials, it is necessary that learners' target needs, learning needs and lacks be determined comprehensively. One way of identifying the target needs of a group of learners is using corpus data. Corpora, large collections of texts in electronic form, are a valuable tool for language teaching and learning either as a source for data driven learning or as a reliable instrument for developing curricula, syllabi, materials and tests. It is possible to gain insights on the naturally occurring patterns of language by means of retrieving frequently occurring, authentic linguistic samples. Corpus linguistics, which have gained momentum in recent years, allow for employing corpus data to inductively learn patterns of grammar and vocabulary behaviours in certain disciplines, genres and discourses, to identify frequently occurring items either as individual words or phraseology and to make inferences as to specific usages in specific contexts.

A number of studies have been conducted using corpora, particularly after the automated computerized analysis have become possible with the advent of software like Word Smith, AntConc, and Sketch Engine. Lexical studies conducted through corpus data have been influential especially with regard to putting frequency of occurrence to a more central position in the development of lexical syllabi or decisions made regarding any vocabulary-related issues. In addition to frequency, the importance of range and dispersion criteria were also established in corpus research. For the reliability of frequency figures, it is important that the items occur in a wide range of texts or subcorpora and be evenly distributed across texts.

There exist many different types of corpora used for different purposes, including written and spoken corpora, general and special corpora, monolingual and multilingual corpora, and learner corpora.

ESP is claimed to have a specialized vocabulary or specific style which is different from general English. It is one of the fields where corpus data provides valuable insights as regards the typical language encountered in a specific domain. Corpora can be useful resources for understanding the characteristics of a specific context, lexical and grammatical patterns used in that specific context. In ESP studies, small corpora are preferred for the sake of mirroring the specialities of the special context more effectively.

Vocabulary learning is an indispensable part of language learning and it is believed that 95 percent vocabulary coverage is necessary to comprehend a text. Word lists have been developed with an aim to provide an inventory for language learners based on frequency of occurrence. Wordlist development has been one of fields where corpus data have been widely used. In creating wordlists, frequency of occurrence or keyness analyses are performed where the unit of analysis must be specified. Also, frequency and range threshold values must be set for the analyses to generate reliable results. Several wordlists have been constructed based on corpus data. The new GSL (Brezina and Gablasova (2015) is probably the most up-to-date one derived from four language corpora (LOB, BNC, BE06, and EnTenTen12) of the total size of over 12 billion running words. Despite being low in number, there have also been attempts to develop specialised wordlists featuring the core words frequently occurring in a specific discipline or domain. The Basic Engineering Word List (BEL) developed by Ward (2009) and the Science Word List developed by Coxhead and Hirsch (2007) are two examples of specialised wordlists derived from corpus data.

CHAPTER III

METHODOLOGY

3.1. Introduction

The overarching research objective of this study is to analyse the lexical needs of freshman engineering students at a state university in Turkey and construct a corpus-derived word list that is representative of the content that they are exposed to in their science courses. Science courses, namely Physics, Chemistry, Biology and Calculus, constitute the majority of the course work in the first term and are categorised as must courses common to all engineering students. To attain the ultimate goal of the research, it was necessary to determine the course requirements, the course content and materials that the students use. Following that, the content was to be examined through a quantitative approach, to identify the frequently occurring vocabulary. In the pursuit of meeting the research objectives, the following research questions were devised:

1. What are the freshman engineering students' target lexical needs for the science courses?

    1.1. What are the perceptions of the lecturing staff regarding the freshman engineering students' target needs?

    1.2. What specific vocabulary do the science textbooks used by freshman engineering students feature?

    1.2.1. What are the lexical frequency representations of the science textbooks used by freshman engineering students?

    1.2.2. What keywords and multi-word terms constitute the key vocabulary in the science textbooks of freshman engineering students?

1.3. To what extent does the content of the English preparatory programme meet the target lexical needs of freshman engineering students for the science courses?

1.4. How does a keyword list based on a corpus of science textbooks relate to the commonly available wordlists, namely the *New General Service List,* the *New Academic Vocabulary List* and the *Science Word List?*

1.5. What are the perceptions of the lecturing staff regarding the usefulness of the items in the key word list derived from the corpus of science textbooks?

3.2. Research Design

This research is a mixed-method single case study. It features the characteristics of a case study by its nature due to the fact that it aims to describe the instructional needs of a specific group in a particular, real-life setting. Becker (1970) explains that case study is a detailed analysis of an individual case where "one can properly acquire knowledge of the phenomenon from intensive exploration of a single case" (p.75). In the sense that it portrays a real phenomenon about real people in real situations, case study method provides readers with valuable insights on the topic explored. Gall, Borg and Gall (1996) point out that case studies richly describe, explain, or assess and evaluate a phenomenon. According to Merriam (1998), the purpose of the case study research is to choose one or multiple cases regarding the actions or phenomenon within their real life context so as to collect data to understand various aspects regarding the research problem. Yin (2014) also states that case study involves the investigation of one or more real-life cases to capture its complexity and details. This study's particular focus being describing the needs of a specific group of learners in a specific real life setting, the research methodology fits into the category of single case study. It aims to capture the specific lexical needs of a group of learners through the investigation of their case. The study adopts a mixed-method research design as a combination of quantitative and qualitative approaches are employed in the study. A mixed-method study is one where the researcher uses at least one quantitative method and one qualitative method to collect, analyze, and report findings in a single study (Fielding & Fielding, 1986; Greene et al., 1989). Creswell (1999) also defines a mixed-method study as "one in which the researcher incorporates both qualitative and quantitative methods of data collection and analysis in a single study" (p. 455). He explains that this type of study enables the researcher to understand or explain a complex phenomenon not only qualitatively but also using numbers, charts and statistical analyses.

The current study employs a combination of qualitative (interview and questionnaire) and quantitative (corpus analysis) approaches. Table 7 below provides an overview of the research design adopted in this study.

Table 8

An Overview of Research Design

<table>
<tr><th></th><th>Data collection and analysis steps</th><th>Research questions addressed</th></tr>
<tr><td>Phase 1</td><td>Interview with course instructors</td><td>1.1. What are the perceptions of the lecturing staff regarding the freshman engineering students' target needs?</td></tr>
<tr><td>Phase 2</td><td>Corpus compilation</td><td rowspan="4">1.2. What specific vocabulary do the science textbooks used by freshman engineering students feature?
1.2.1. What are the lexical frequency representations of the science textbooks used by freshman engineering students?
1.2.2. What keywords and multi-word terms constitute the key vocabulary in the science textbooks of freshman engineering students?</td></tr>
<tr><td>Phase 3</td><td>Developing a frequency list</td></tr>
<tr><td>Phase 4</td><td>Keyness analysis and development of a key word list</td></tr>
<tr><td>Phase 5</td><td>CEFR level categorisation and PoS tagging</td></tr>
<tr><td>Phase 6</td><td>Comparison of the target wordlist with the EFL wordlist used in the preparatory programme</td><td>1.3. To what extent does the content of the English preparatory programme meet the target lexical needs of freshman engineering students for the science courses?</td></tr>
<tr><td>Phase 7</td><td>Comparison of the target wordlist with commonly used wordlists, the new GSL, the new Academic Vocabulary List and the Science Word List</td><td>1.4. How does a keyword list based on a corpus of science textbooks relate to the commonly available wordlists, namely the New General Service List, the New Academic Vocabulary List and the Science Word List?</td></tr>
<tr><td>Phase 8</td><td>Integration of qualitative data from teachers for the target word list</td><td>1.5. What are the perceptions of the lecturing staff regarding the usefulness of the items in the key word list derived from the corpus of science textbooks?</td></tr>
</table>

Phase 1

The first phase involves the process of needs analysis. Semi-structured interviews are conducted with the instructors delivering the science courses at the engineering departments. It is intended with the needs analysis to set forth the freshman students' target needs and lacks. Expressed another way, through the needs analysis, the researcher aims to specify the requirements of the science courses and the extent to which students can fulfil these requirements. The primary focus of this study being the identification of target lexical items required in the science courses, analysis of learning needs was considered to be irrelevant, as students would not have an opinion regarding what vocabulary they would need in these courses. Within the framework of Hutchinson and Waters' learning needs analysis (1987), it is necessary to investigate who the learners are, why the learners are taking the course, how they learn, and the available sources. Nevertheless, this aspect of needs analysis is beyond the scope of this study, due to the fact that the overarching goal of the study is to determine the lexical target needs of the learners; as such, collecting data from the learners as to the "how" of learning process would not relate directly to the objective of this study.

Phase 2

In the second phase of the study, the course content is analysed quantitatively. The science textbooks constitute the core content of the courses; hence, the textbooks used in each course is collected. The texts are converted into *txt*. format for computerised analysis. Recurrent and irrelevant data such as the titles, table of contents, figures, visuals and appendices are removed from the texts manually prior to the analysis. By means of the Sketch Engine software programme, a corpus is compiled, which will hereinafter be referred to as the "*Science Textbooks Corpus*" for the purposes of this study.

Phase 3

The third phase of the study involves constructing a word list from the corpus data based on frequency information. By means of a frequency analysis conducted within the Sketch Engine tool, a word list is created according to the items' frequency of occurrence in the corpus, taking into consideration the criteria of range and dispersion. The list is then revised by eliminating the grammatical words and other frequent non-lexical items such as abbreviations, letters or figures. There is no need to lemmatize the list as the unit of analysis in the frequency list is chosen as "lemma" and thus it is already built in lemmas.

Phase 4

The next phase of the study is keyness analysis. In order to investigate the specialised vocabulary that is peculiar to the science text books, the items that occur with significantly higher frequency in the *Science Textbooks Corpus* in comparison to the *British National Corpus,* which is the reference corpus, are identified. The list is then revised by removing the irrelevant items, as well as the items that are beyond the set threshold levels for frequency and dispersion. Also, a list of multi-word terms is extracted through the analysis. In a similar vein, the multi-word terms list is cleared off irrelevant data and items that are not within the set limits.

Phase 5

Following the fine-tuning of the list, the list is reorganised according to the CEFR levels of the words. The items on the list are categorised according to CEFR levels. The items are tagged as A1, A2, B1, B2, C1, and C2; however, there were also items that were not categorised under any of the levels. The A1 level items, being very simple words which can be assumed to have been learnt at lower levels within a general English programme, are excluded from the list. The items on the list are also tagged with part of speech information.

Phase 6

In this phase, the corpus-derived keyword list is compared with the target vocabulary list taught at the EFL preparatory programme in order to discover to what extent the items on the keyword list are covered. The coverage values are obtained through AntWordProfiler programme.

Phase 7

In this phase of the study, the *Science Textbooks Wordlist* is compared with the New GSL, Brezina and Gablasova (2015), the new Academic Vocabulary List (Gardner and Davies, 2014) and the Science Word List (Coxhead and Hirsch, 2007), with the aim of determining the extent to which these lists overlap. The coverage values are obtained through AntWordProfiler programme.

Phase 8

In the last phase of the study, teachers' intuitions are explored regarding the usefulness of the items on the *Science Textbooks Word List,* constructed upon the corpus data with the ultimate goal of suggesting a fine-grained, pedagogically convenient target word list for the students.

3.3. Setting

The study was conducted within the context of Ankara University, Faculty of Engineering, which has the highest number of students enrolled at the English preparatory programme. The faculty is divided into nine departments each of which provides English-medium instruction. To start their majors, students need to complete the compulsory English preparatory programme or pass the proficiency exam if they already possess the linguistic competence necessary to perform their studies. Thıse who cannot prove that they possess the required level of English receive a one-year English education at the School of Foreing Languages before starting their majors. At the end of the programme, they sit the proficiency exam. If they pass the exam, they have the right to start their studies.

Table 8 presents the list of these engineering departments with the number of students enrolled at the preparatory programme.

Table 9

The list of engineering departments where medium of instruction is English.

Engineering Departments	***The number of students enrolled at the preparatory programme***
1. Food Engineering	*109*
2. Chemical Engineering	*106*
3. Electrical & Electronics Engineering	*84*
4. Biomedical Engineering	*76*
5. Computer Engineering	*75*
6. Physics Engineering	*59*
7. Geological Engineering	*59*
8. Energy Systems Engineering	*34*
9. Energy Engineering	*3*

The students enrolled at the preparatory programme are offered one-year general English education. The students learn general English in classes together with the students from other departments or faculties; in other words, they are not exposed to a specific tailored programme designed as per their specific needs. Having completed the preparatory programme, the students start their majors. Science courses, namely Physics, Chemistry, Biology and Calculus are offered to the engineering students in their freshman year.

3.4. Data collection tools

Qualitative and quantitative data collection tools have been employed for the purposes of this study. Qualitative data was collected through the instruments of interview and questionnaire; quantitative data is collected through corpus compilation.

3.4.1. Interview

It is intended with the needs analysis to specify the necessities and lacks of the freshman students at the engineering departments regarding the science courses they take. In Nation and Macalister's terms, (2010) necessities refer to what the learners need to do when they start their studies at the department such as listening to lectures, writing assignments and exams whereas lacks indicate the learners' present level. In this respect, it is supposed that the course instructors can provide a vivid picture of the necessities for the courses and lacks of their learners. To this end, semi-structured interviews were conducted with the course instructors. A total of 7 instructors giving the science courses at the Faculty of Engineering were interviewed. The interviews were conducted face-to-face and online. The questions were devised as open-ended items so that the interviewees could offer elaborated answers and express their views, which would constitute valuable data for the research. The questions in the interview addressed the requirements of the course, what the students would do with the language they have learnt, the skills and the lexical knowledge required for the course, which would point to the target needs. Also, the interviewees were asked about the challenges and difficulties the students face throughout the course, which was assumed to provide insights into what the students "lack".

The interviews were held in English and lasted approximately 40 minutes and were recorded with the purpose of transcription and content analysis.

3.4.2. Corpus compilation

The results of the interview data, a detailed account of which is presented in the "Results" section of the study, indicated that the core materials used in the courses are the textbooks, supported with presentations and lectures. The presentations used in the courses and the lectures were reported to be based on textbook information; therefore, the data extracted from the textbooks would portray the lexical needs of the students taking these courses. In order to investigate the lexical features of the content of the science courses, the textbooks used in the courses were collected and a corpus was compiled. All of the science textbooks used in courses were used for corpus compilation in order to ensure representativeness and balance. The corpus content was comprised of whole texts, not samples, so as to encompass all the features of the material. The whole content of each textbook used in each course was compiled into a sub-corpus, which would then constitute the specialized corpus.

The textbooks used in the science courses in the first year of studies in the engineering departments are shown in Table 10 below.

Table 10

Textbooks used in first-year science courses in the engineering departments

Must Courses	Textbooks
Physics	*Physics for scientists and engineers*. R. A., & Jewett, J. W. (2018). Cengage learning. (6th Edition)
Calculus	*Thomas' Calculus*. Thomas, G. B., Weir, M. D., Hass, J., & Giordano, F. R. (2005). Addison-Wesley.
Chemistry	*General Chemistry: Principles and Modern Applications* Petrucci, R. H., Herring, F. G., & Madura, J. D. (2010). Pearson Prentice Hall.
Biology	*Biology: Life on Earth*. Audesirk, T., Audesirk, G., & Byers, B. E. (2001). Pearson Educación.

The textbooks used in the courses were made available in PDF format and converted into *txt*. format to match the input requirements of the software programme *Sketch Engine,* a corpus analysis toolkit, which hosts a comprehensive set of tools such as concordancer, word frequency generator, keyword and multi-word terms analysis and so forth. Following this, the texts were subject to standardisation, where the visuals, graphs, tables, figures and the sections such as the table of contents, appendices, preface and references were removed. The process was carried out manually by the researcher. The texts were then uploaded into the *Sketch Engine* programme for corpus compilation and analysis. The target corpus compiled was comprised of a total of 2,303,096 tokens and 1,898,324 words. The collection of the texts consisting of the reading in the science classes compiled as a corpus within the Sketch Engine programme would serve as the database for frequency and keyword analysis. The keyword list derived from the corpus data would serve the ultimate goal of this study.

3.4.3. Questionnaire

High-frequency word lists based on objective corpus data are doubtlessly valuable sources for L2 learners and teachers. However, the extent to which the words in a word list are relevant to learners in a specific context may vary (Milton, 2009). Nation (2016) also thinks that a word list that is based purely on corpus data bears the risk of missing the items that occur with low frequency in a corpus but are valuable for L2 learning. In that respect, Stein (2017) points out that some items in the New-GSL compiled by Brezina and Gablasova (2015) may not be relevant to EFL beginners, and as a result of that teachers and learners may not fully understand how corpus-based word lists can contribute to their teaching and learning. Therefore, it was considered necessary to resort to expert opinion to generate a list based not only objective, quantitative corpus data but also on the intuitive ratings of teachers regarding the usefulness of the lexical items presupposed to be key concepts for the courses in question. For the purposes of triangulation and obtaining a more fine-grained, pedagogically convenient word list, the items extracted from the keyness analysis are presented to the lecturing staff to receive their opinion. To this end, a questionnaire was designed to examine the teachers' perceptions of the usefulness of the words in the keyword list developed.

The final keyword list derived from the corpus data was subject to revision by the researcher. Irrelevant items and A1 level words were removed from the list and the final list consisted of 1195 items. In order to be able to receive expert opinion on each of the items on the list, it was necessary to make the list more manageable in terms of the number of the items the list consisted of. Expert opinion is required to determine the usefulness of the key vocabulary that is more discipline-specific and thus the items that are lower level are not included in the questionnaires. To this end, A2 level items were also excluded from the list for expert opinion, and the list was reduced to a total of 1103 items so that the questionnaires could be designed in a way that the participants would address all of the items on the list. A total of 5 questionnaires were constructed, one consisting of 223 items and the rest 220 items each. Dividing the list into five, it was aimed to have the participants rate a separate set of words, and thus each item on the list could be addressed. A five-point Likert scale, which measures respondents' attitudes to a particular question or statement, was employed for the questionnaire. Each participant rated the usefulness of each word in helping their students to perform their studies. Point 5 on the scale was coded as "extremely useful", and Point 1 as "not useful at all".

Table 11 below shows the design of the questionnaire, with a sample of 10 items. (The questionnaires can be found in Appendices E-I)

Table 11. Questionnaire for Course Instructors

Questionnaire for Course Instructors						
Please give your answer to the question: **To what extent is the word useful for your students in the science courses?** by choosing a number from 1 to 5 in the degree of usefulness column (1 is the LEAST useful and 5 is the MOST useful.						
		Degree of Usefulness				
N	***Headword***	***5***	***4***	***3***	***2***	***1***
1.	*energy*					
2.	*figure*					
3.	*water*					
4.	*equation*					
5.	*point*					
6.	*example*					
7.	*mass*					
8.	*force*					
9.	*cells*					
10.	*reaction*					

3.5. Data Analysis

The data collected through interviews was analysed by means of content analysis method, and the questionnaire data was subject to statistical analysis. The corpus data was used for frequency and keyness analysis. The details of the procedures are explicated in the following sections.

3.5.1. Content Analysis

Content analysis, as defined by Krippendorff (2003), "is a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use" (p. 18). To interpret the interview data collected from the teachers, content analysis was employed. The content analysis procedure involved transcription of the interviews, coding of the data and revealing the common themes and patterns occurring within the data. The responses of seven participants, who are the course instructors giving the science courses at the engineering departments, were analysed and recurrent concepts and themes were identified, which were then categorised. The results of the content analysis are displayed in the following chapter of the study.

3.5.2. Corpus Analysis

In this study, *Sketch Engine* software was used as the corpus analysis tool. Sketch Engine is an online tool used to explore how language works. Its algorithms analyse authentic texts of billions of words (text corpora) to identify what is typical in language and what is rare, unusual or emerging usage. It is used by linguists, lexicographers, translators, students, and teachers. Its functions are based on mathematical and statistical computations which enable users to accurately search and filter queries in language corpora.

Prior to running a frequency analysis and keyword analysis, the corpus was compared against a reference corpus to ensure that it is different from a general corpus. If the target corpus compiled with a purpose of creating a specialised database were highly similar to a general corpus, then the research objective would be reconsidered.

3.5.2.1. Frequency Analysis

Following the compilation and analysis of the corpus, a frequency analysis was conducted which listed the most frequently occurring items within the corpus. In addition to the absolute frequency values, the Sketch Engine programme generates results with a range of frequency values, namely *relative frequency (frequency per million), document frequency, relative document frequency and average reduced frequency*, which are taken into account for the purposes of this study.

Absolute frequency values show how many times an item occurred in the corpus. To illustrate, if an item has a frequency of 10, it means that it appeared 10 times. *Relative frequency,* on the other hand, refers to the number of occurrences of an item per million tokens; therefore, it is also called frequency per million. It is used to compare frequencies between corpora of different sizes. Relative frequency is always related to the whole corpus or sub-corpus, not to a text type. Restricting the query to one or more text types will affect the number of hits but the frequency per million will still be calculated using the number of tokens in the whole (sub)corpus.

The absolute frequency and relative frequency values are illustrated through an example within the Skecth Engine programme, shown in Table 12. Looking up the frequency of the word *helps* in the British National Corpus (112,181,015 tokens), first in the spoken text type and then in the spoken subcorpus will produce these results. The results show how many times the word is used within the spoken subcorpus and spoken texts and its frequency in relation to the number of tokens in the whole corpus and subcorpus.

Table 12. An example of frequency results for the word "helps" in the BNC, retrieved from Sketch Engine

SUBCORPUS SELECTED	none	none	Spoken 11,787,138 tokens
TEXT TYPE SELECTED	None	spoken	None
HITS	3,116	302	302
FREQUENCY PER MILLION	27.75 in relation to the number of tokens in the whole corpus	2.69 in relation to the number of tokens in the whole corpus	25.62 in relation to the number of tokens in the subcorpus
POSSIBLE INTERPRETATION	**helps** appears 27.5 times per million tokens in BNC	'spoken' helps appears 2.69 times per million tokens in BNC	**helps** appears 25.62 times per million tokens in the spoken part of BNC

Document frequency refers to the number of documents where the word or phrase appears. In other words, it shows the "range" of texts the lexical item appears in. For example, if the corpus has 100 documents and 2 documents contain a word, the document frequency of that word is 2. *Relative document frequency* is the percentage of documents that contain the word or item and it is used to compare document frequencies between corpora of different sizes.

For the analysis to yield sound results, a number of choices were made on the tool prior to running the analysis. The frequency threshold is specified as fifty, which means that the occurrences with a lower frequency than fifty will not be shown in the list. As Nation (2013) suggests, it is especially important to determine which words that L2 learners should learn first, in that it helps them get the best return for their learning effort. Given the importance of frequency, the items that appeared at least fifty times within the corpus were to be involved in the list. Also, a cut-off value for document frequency was also set as two; the items to be included in the list must occur for at least in two sub-corpora out of four comprising the corpus. The selection criteria were based on Coxhead's principle of selecting AWL words. Coxhead's corpus for the AWL consisted of 3.5 million words and to select the words for her list, she determined that each word form in the AWL needed to occur at least 100 times in the entire corpus and at least 10 times in each of the four disciplines. Regarding the criteria in the selection of the Academic wordlist, Coxhead and Hirsch (2007), note that "principles used in the selection of words for the AWL were range (the word families occurred in more than 15 of the 28 subject areas), frequency (the word families occurred more than 100 times in the corpus), and uniformity (the words occurred at least ten times in each of the four disciplines)" (p.66). The Science Textbook Corpus being half size of Coxhead's corpus, the frequency threshold value was set as fifty.

Another important choice to be made before the analysis is the unit of analysis. The unit of analysis was chosen as *lemma,* which is the basic form of a word. For instance, the lemma *go* encompasses the inflicted forms of the verb such as *goes, went, going, gone.* In a frequency list of lemmas, the forms of the same basic form are counted together and listed under one item.

Having set these criteria, the frequency analysis was conducted and a list of 2954 items were generated. The list produced was based on absolute frequency values. Following the analysis, the researcher revised the list for several criteria to be met. Firstly, the items with a document frequency (DOCF) value of one were removed from the list as they appeared in only one of the sub-corpora. Next, for the list to comprise "lexical" content only, all the non-lexical and irrelevant items, such as function words (the, is etc.) symbols (x, y, t etc.), abbreviations (cm, rna, dna etc.), prepositions (at, on, in etc.), conjunctions (therefore, besides etc.), proper nouns (Einstein, Kelvin etc.), were removed from the list manually by the researcher.

3.5.2.2. Keyword Analysis

The next step was to check the extent to which the words were distinctive in the scientific domain, which would show keyness. A keyword list provides a measure of statistical saliency that is based on chi-square or log-likelihood measures (Anthony, 2019) but a word list is built based on merely frequency figures (Baker, 2006). Evison (2010) states that keywords are "those words which are identified by statistical comparison of a target corpus with another larger corpus, which is referred to as the 'reference' or 'benchmark' corpus" (p. 127). In this respect, the target corpus, *Science Textbooks Corpus*, was compared to the British National Corpus and keyness values were obtained. For the analysis, the frequency threshold was specified as thirty, which means that the occurrences with a lower frequency than thirty would not be shown in the list. The unit of analysis is chosen as lemma so that the different forms of the same lemma would be treated as the same item.

The keyword list extracted through the analysis was then manually revised. The items that appeared in fewer than two sub-corpora were removed. The list was filtered to eliminate the irrelevant items such as abbreviations, non-words, symbols, grammar words, cognates etc. as well as erroneous entries. Average reduced frequency (ARF), a variant on a frequency list that 'discounts' multiple occurrences of a word that occur close to each other, e.g. in the same document, was used to meet the criteria of dispersion. The items that have a lower ARF value of ten were removed from the list. The list consisted of a high number of technical terms and substance names. Considering that not all of such items can meet the criteria of teachability, particularly within an English syllabus, the technical terms and substance names occurring with an ARF value lower than twenty were eliminated. The rationale for setting the ARF threshold value as twenty, a figure higher than ten, was that some items can be worth including in the list because of their high frequency of occurrence despite being a technical word.

In a similar vein, the list of the multi-word terms extracted from the corpus analysis was subject to manual filtering. The document frequency is set as two; the items appearing at least in two of the subcorpora are included in the list. The items occurring in solely one of the subcorpora are removed from the list, which assures that the items worth including in the list appear in a range of texts in a balanced way.

3.5.3. Coverage Analysis

One of the purposes of this study is investigating to what extent the corpus-derived keyword list cover the vocabulary items taught in the preparatory programme. It was aimed to find out how much of the vocabulary that the students need for their first year courses overlap with the vocabulary that is taught in the preparatory programme before starting their studies. In pursuit of this objective, the keyword list (*Science Textbooks Word List*) was compared with the word list of the intermediate level coursebook used in the preparatory programme, through AntWord Profiler. The analysis generated coverage results in percentages showing how much of the *Science Textbooks Wordlist* the coursebook wordlist covered.

Using the same programme, the Science Textbooks Word List (STWL) was compared to the New GSL, the new Academic Vocabulary List and the Science wordlist, in order to investigate how much overlap existed among these wordlists. The rationale for choosing these wordlists for comparison is basically owing to their being the most up-to-date lists developed. The New General Service List was developed by Brezina and Gablasova (2015) as a result of a robust comparison of four language corpora (LOB, BNC, BE06, and EnTenTen12) of the total size of over 12 billion running word and contains 2494 headwords. The General Service List published by Michael West in 1953 was considered to be old and dated, and thus, a new list was necessary. the New General Service List developed by Brezina and Gablasova (2015) was based on a wider set of corpora and developed more recently, and consisted of 2494 lemmas. The New Academic Vocabulary List, developed by Gardner and Davies (2014) is an improvement on the commonly known Academic Word List (Coxhead, 2000) in that the AWL was based on 3.5 million words from the 1990s whereas the new Academic Vocabulary list was derived from a 120-million-word academic sub-corpus of the 425-

million-word Corpus of Contemporary American English (COCA) (Davies, 2012) and consisted of 1991 word families. The Science Word List (Coxhead and Hirsch, 2007) which was based on 1,761,380 tokens from 14 subject areas consisted of 318 word families that did not occur in GSL and AWL.

The analysis yielded coverage percentages for each wordlist.

3.5.4. Statistical Analysis

The data provided via questionnaires were analysed by means of *JASP* programme, a practical and valuable tool to deal with quantitative data in research. Such data as descriptive statistics, arithmentic mean, frequency and percentages were obtained from the statistical calculation. Also, the correlation between the objective corpus frequency measures and the subjective intuitive ratings of teachers was investigated. The results are displayed in the following chapter of the study.

CHAPTER IV

FINDINGS AND RESULTS

4.1. Introduction

In this chapter, the findings obtained from the interviews, frequency and keyword analyses, coverage analysis and questionnaires are presented in an elaborate way. The common themes that emerged from the interviews, the results of the frequency and keyword, the coverage values obtained as a result of the comparison of the wordlists and the questionnaire results are explained in this chapter

4.2. Findings from Interviews with Course Instructors

Research Question 1.1. What are the perceptions of the lecturing staff regarding the freshman engineering students' target needs?

The first research question of the study addressed the target lexical needs of freshman engineering students in the science courses they take, which are physics, chemistry, biology and calculus. In order to discover the necessities and lacks, expert opinion was received through interviews with course instructors (*n=7*) who have been delivering the science courses at the faculty of engineering. The interviews were semi-structured and included open-ended questions for the participants to be able to elaborate on the responses they give. The interviews were transcribed and content analysis was conducted on the data. The results were classified through content analysis are summarised in Table 13, where common themes and categories that emerged are presented.

Table 13

Content Analysis Results

	Themes	Sub-categories	N
Necessities	Requirements of the course	▪ understanding and answering exam questions	7
		▪ comprehension of written materials & lectures	7
		▪ giving presentations	2
		▪ reading equations & theorems	2
	Course content	▪ textbooks	7
		▪ lectures	7
		▪ specific lexis	4
	Skills and sub-skills needed	▪ general proficiency	7
		▪ vocabulary	4
		▪ listening	4
		▪ speaking	4
Lacks	Difficulties faced by students	▪ understanding long sentences & vocabulary	7
		▪ understanding the exam questions	4
		▪ expressing themselves	3
		▪ presentation skills	2
Suggestions		▪ better overall proficiency	5
		▪ familiarity with specific lexis	4
		▪ scientific reading	2
		▪ presentation skills	1
		▪ ESP syllabus	6
		▪ general English syllabus	1

4.2.1 Requirements of the course

In order to establish the target lexical needs of the students, the course instructors were posed questions as to the course requirements, that is, what the students needed to achieve throughout the course, as well as what skills and subskills were needed to meet the course objectives. One outstanding concept commonly referred to was regarding the assessment component of the courses. All the interviewees pointed to the necessity of a thorough comprehension and answering of the exam questions. Students were required to understand fully what is expected of them and provide a satisfactory response to the exam questions. It was deemed necessary to be proficient enough and have necessary lexical knowledge to meet this objective Another important requirement of the course was reported as the full comprehension of the written materials and the lectures. All of the interviewees stressed the importance of understanding the written content in the textbooks as well as the lectures and explanations given by the instructors, which was possible through a good command of language skills, including knowledge of grammar and vocabulary. They mentioned that students needed to prepare for the class by reading the assigned chapters from the book or consolidate their learning by doing the practice activities on the relevant pages. During the classes, they needed to engage in active listening and be able to understand the lectures.

Presentations were another aspect mentioned by the interviewees as to the requirements of the course. Two course instructors stated that students must be able to deliver presentations regarding the course content. It was deemed necessary to be able to prepare a proper presentation and show the skills of good delivery.

Two of the course instructors noted the importance of being able to read the equations and theorems effectively. Equations and theorems being one of the salient features of science courses, it was necessary for the students to read and understand them, which is essential for scientific reasoning. This entailed, besides a wide repertoire of vocabulary, some knowledge of discipline-specific usages.

4.2.2. Course content

As regards the content of the course, all of the interviewees stated that the courses were mostly based on textbook information and the lectures they give. Four interviewees pointed out that the courses feature some specific lexis, and that there are some phrases or chunks that are widely used in the discipline. Familiarity with such items are thought to be of help in comprehension of the content matter.

4.2.3. Skills and sub-skills needed

To be able to achieve the requirements of the course, students need to possess some skills and sub-skills in terms of linguistic competence. In that respect, all of the interviewees stated that students needed to have a good command of English and have a certain level of proficiency so that the goals of the course could be achieved. Four of them mentioned the significance of vocabulary knowledge for understanding the course content, as well as performing the assessment tasks. They also expressed their belief that being familiar with the vocabulary they will come across during the courses can have a positive influence on their comprehension and production, and hence success in the course. Similarly, four instructors pointed to the necessity of good listening skills as most classes are conducted through lectures, the content of which is based on the textbook information. In terms of oral skills, the importance of being able to express their ideas, ask questions and communicate what they have in mind, was noted by four instructors.

4.2.4. Difficulties faced by students

In pursuit of discovering what the students lack, and thus specifying their needs, the course instructors were posed a question about the kind of difficulties students face in terms of language, throughout the courses they take.

The most recurrent theme in the responses was the failure to understand particularly complicated sentence structures and vocabulary, which causes an overall lack of understanding of the target content. All of the course instructors stated that the students had difficulty understanding the texts particularly when they are composed of lengthy sentences with less familiar vocabulary.

Some of the instructors (*n=4*) also noted that the students sometimes failed to understand the exam questions, and hence, asked for explanations in their native language. This may be resulting from the low level of language proficiency in addition to the lack of knowledge of the concepts specific to the discipline.

In addition to the difficulty they have in comprehension, the students were reported to have problems in productive skills, particularly in speaking, as well. Four instructors said that the students cannot express themselves and thus prefer to stay silent instead of inquiring. It is possible that they do not understand something or need clarification or elaboration on the issue but abstain from communicating their inquiry in the target language. This results in the fact that students, all too often, resort to their native language when they feel the urge to speak.

Two of the instructors also mentioned that students do not have the necessary presentation skills. One instructor explained that they do not know how to do research on a topic, outline the main points, and prepare and deliver a Powerpoint presentation. The problems students face in the delivery of a presentation might be resulting from their low level of language proficiency and lack of speaking skills, yet they also seem to lack the knowledge they need to do the preliminary work, to specify the main points of the topic and integrate them into their presentation.

4.2.5. Suggestions

The course instructors' opinions' on what the students need in order to close the gap between what they are required to do and what they are able to do were also explored. The majority of the instructors (*n=5*) stated that the students need to have a better proficiency of English to get a good grasp of the written and spoken content they are exposed to during the course. They expressed their belief that the students would not have much difficulty in comprehending new concepts and topics in their discipline providing that they improved their target language skills.

Another theme that is worth mentioning is the specific vocabulary inherent in the scientific texts. Four instructors hold the opinion that students comprehension level can increase if they have a better vocabulary knowledge. In that respect, they noted that familiarity with scientific terms could have a positive impact on their success. The instructors expressed that they did not expect students to know the scientific concepts which they would learn in the course of the classes they take during their study, but rather they would appreciate if the students were familiar with less technical discipline-specific lexical knowledge on the grounds that this would contribute to their overall understanding and internalising of the key concepts.

One of the course instructors stated that it is necessary that students learn the essentials of a good presentation. The instructor believed that having good presentation skills must be a prerequisite for any course at tertiary level. The students need to know how to search for a topic, how to specify the main ideas worth including in the presentation, how to paraphrase the information, and how to prepare the content of the slides without writing each and every piece of information on the slide.

Another suggestion from the instructors (*n=2*) was that the students could be exposed to reading passages that are relevant to their field of study. They believed that dealing with scientific texts appropriate to their level of proficiency would be valuable in that they would help the learners familiarise with the style and content of such texts. Also, one of them stated that reading such texts could help them learn different vocabulary.

When asked for their opinions on the idea of developing an ESP syllabus for engineering students within the preparatory programme, six of the instructors said it would be useful for the students. They expressed their belief that getting familiar with more discipline-specific content would have a good effect on their performance. One of the instructors responded negatively, saying that there is no need for such a specific syllabus, and that improving overall proficiency would suffice.

4.3. Corpus Analysis

A target corpus (*Science Textbooks Corpus*) was generated from the textbooks used as the core content in Physics, Chemistry, Biology and Calculus courses in the engineering departments in the first year of studies, meeting the criteria of balance and representativeness. The corpus provided the basis for development of a frequency based keyword list.

4.3.1. Compilation of the *Science Textbooks Corpus*

The corpus created from the science textbooks used in the engineering departments at a state university consisted of 2,303,096 tokens, and 1,898,324 words. Using samples from four different fields of science and keeping the samples as whole texts in the corpus, it was intended to meet the criteria of representativeness. The target corpus is divided into four sub-corpora according to the text types. That is, the sub-corpora were created according to the specific disciplines the data were obtained from so that a further analysis between them could be possible when necessary. In terms of token coverage, the physics textbook covered 34 %, the biology textbook 25.8 % percent, the chemistry textbook 25.1% and the calculus textbook 15.2% of the whole corpus. These findings indicate that an acceptable balance was achieved in the compilation of the corpus. The following table shows the number of tokens and coverage percentages of each sub-corpus within the main corpus.

Table 14

The number of tokens and coverage percentages of the sub-corpora

Name of the subcorpus	***Tokens***	***Percentage (%)***
Physics subcorpus	783,425	34
Biology subcorpus	591,391	25.7
Chemistry subcorpus	577,433	25.1
Calculus subcorpus	350,847	15.2

Following the analysis of basic corpus information such as the number of tokens, words and sentences in the main corpus and the coverage values of the sub-corpora in the main corpus, text type analysis was conducted. Text type analysis show that 30.5% percent of the data include function words like *the, of, a, in, and, is* and *to,* and the remaining 69.5 % consist of other items.

In order to establish that the target corpus is a specialised one different from a general corpus, a comparison was conducted. The programme used for corpus analysis, Sketch Engine, allows for the comparison of a number of corpora including the corpora compiled by the user. It compares the corpora through the comparison of word forms or lemmas in the corpora. As a result of this comparison, a score indicating the extent to which the corpora are similar or different is obtained. A score of 1 indicates identical corpora; the higher the score, the more different the corpora are. Since the comparison is done on tokens, the score is not affected by sentence length, number of documents, corpus size or grammatical features. The comparison of the *Science Textbooks Corpus* to the *British National Corpus (BNC)* yielded a value of *3.96*, which indicates a significant difference between the two corpora, confirming that the former is of a specialised nature. Obtaining this value confirms that the *Science Textbooks Corpus* is different from the *British National Corpus*, which is a general reference corpus, and that the *Science Textbooks Corpus* is a specialized one.

4.3.2. Frequency analysis

Research Question 1.2. What specific vocabulary do the science textbooks used by freshman engineering students feature?

1.2.1. What are the lexical frequency representations of the science textbooks used by freshman engineering students?

The second sub-question addressed the frequency of occurrence of the lexical content in the science textbooks used by freshman engineering students. To find out the number of lexical occurrences within the corpus, a frequency analysis was conducted through the Sketch Engine programme. The analysis generated a total of 2954 items. The most frequently appearing items are mostly function words, which were subsequently removed from the list to obtain the frequently occurring lexical items. Table 15 shows the most frequent 20 items, all of which are function words.

Table 15.
The 20 most frequent items in the Science Textbooks Corpus

N	**Item**	**Frequency**	**Relative freq.**	**Document Freq.**	**ARF**
1.	*the*	164292	71335.28	4	107231.01
2.	*of*	79935	34707.62	4	51343.98
3.	*be*	71955	31242.72	4	47122.85
4.	*a*	70339	30541.06	4	42147.58
5.	*in*	44448	19299.23	4	27631.53
6.	*and*	43626	18942.327	4	27785.12
7.	*to*	38266	16615.02	4	23591.74
8.	*that*	21586	9372.60	4	12994.11
9.	*for*	15220	6608.49	4	8629.27
10.	*as*	14550	6317.58	4	8390.39
11.	*by*	13478	5852.12	4	7785.23
12.	*at*	11741	5097.92	4	5685.00
13.	*with*	11434	4964.62	4	6712.23
14.	*we*	10923	4742.74	4	4588.92
15.	*have*	10838	4705.83	4	6107.43
16.	*from*	10708	4649.39	4	6098.39
17.	*on*	10580	4593.81	4	5712.64
18.	*this*	10415	4522.17	4	6077.61
19.	*can*	7745	3362.86	4	4348.72
20.	*it*	7368	3199.17	4	4135.45

The frequency-based wordlist extracted from the corpus was then subject to qualitative analysis by the researcher. That is, the items that appeared in only one sub-corpora and the non-lexical items such as function words, symbols, abbreviations, prepositions, conjunctions, proper nouns and erroneous entries were cleared off the list. All of the items on the list were checked for consistency carefully so that a sound and effective word list could be obtained. Then, the list was reorganised according to average reduced frequency value. The final list was comprised of 1688 items. The full list can be found in Appendix B. The table below shows the most frequent twenty lexical items arranged as per the value of average reduced frequency.

Table 16.

The 20 most frequent lexical items in the Science Textbooks Corpus

N	HEADWORD	Frequency	Relative Freq.	DOCF	ARF
1.	*show*	4082	1772.39	4	2149.39
2.	*find*	4459	1936.08	4	2048.18
3.	*example*	3650	1584.82	4	1904.98
4.	*give*	3358	1458.03	4	1751.37
5.	*point*	4730	2053.75	4	1647.80
6.	*equation*	4524	1964.31	4	1547.45
7.	*form*	3177	1379.44	4	1485.96
8.	*value*	3925	1704.22	4	1455.36
9.	*see*	2498	1084.62	4	1401.08
10	*energy*	5630	2444.53	4	1385.48
11	*time*	3395	1474.10	4	1382.68
12	*change*	3293	1429.81	4	1252.27
13	*result*	2170	942.20	4	1202.46
14	*call*	2409	1045.98	4	1200.77
15	*make*	2082	904	4	1161.09
16	*water*	4074	1768.92	4	1155.20
17	*produce*	2639	1145.84	4	1140.43
18	*small*	2262	982.15	4	1130.25
19	*number*	2896	1257.43	4	1057.48
20	*function*	3888	1688.16	4	1056.58

The item with highest frequency in the list is ranked on the top of the list and appears 4082 times in the target corpus of 2,303,096, and would occur 1772.39 times per million tokens, indicated by the relative frequency value, while it would occur 2149.39 times in a homogenous corpus. The document frequency value also shows that the item occurs in all of the four sub-corpora.

4.3.3. Keyness analysis

Research Question 1.2.2. What keywords and multi-word terms constitute the key vocabulary in the science textbooks of freshman students?

Having obtained the frequency-based results, the next step is to examine keyness to address the research question: "What keywords and multi-word terms constitute the key vocabulary in the science textbooks of freshman students?" In order to identify the specialised key vocabulary in the science textbooks occurring with reasonable frequency and range, a comparative analysis was conducted. The most common way of determining the specific occurrences in a corpus is to compare the discipline-specific corpus with a general, representative one. The items that do not appear or appear only with low frequency in the reference corpus, but appear in the discipline-specific corpus with a higher frequency and range can be considered keywords. With the purpose of constructing a specialised keyword list for the science textbooks, a keyword analysis is conducted on the British National Corpus (BNC), a one million written general English corpus. After obtaining the results, the data was examined and the items with the document frequency value below two were removed from the lists manually. Also, the list was reviewed for non-lexical items such as abbreviations, function words, proper names and etc. The final list was comprised of 1249 lemmas.

The subsequent steps were tagging the items with part of speech information and specifying their CEFR levels. Each word in the list was checked for their CEFR level by means of the website "Text Inspector", which is a web-based language analysis tool developed by Stephen Bax. It uses Cambridge Learner Corpus (CLC), which is a large collection of examination scripts from English language learners across the world to help analyse texts in terms of their CEFR level. The words were labelled as A1, A2, B1, B2, C1 and C2. A high number of words (*n=450*) were categorised as "unlisted" in the tool; therefore, these words were not labelled with any level information.

Having identified the CEFR levels of the items, the researcher decided to omit the A1 level items with the rationale that A1-level words were too simple to be included in a discipline-specific keyword list. It can be assumed that these basic words have already been learnt at the earlier stages of language learning. Additionally, removing the A1-level items (*n=53*) would reduce the list to a more manageable size. As a consequence, the A1 level items were eliminated and the final list comprised of 1195 lemmas. Table 17 shows the first 30 words ranked according to the average reduced frequency (ARF) value. The full list can be found in Appendix C.

Table 17
The first 30 headwords in the keyword list.

	Item	***PoS***	***CEFR level***	**Frequency (focus)**	**Relative frequency (focus)**	**Document Frequency**	**Average Reduced Frequency (focus)**
1.	***point***	*n*	*A2*	4730	2053.75	4	**1647.80**
2.	***equation***	*n*	*C1*	4524	1964.31	4	**1547.45**
3.	***form***	*v/n*	*A2*	3177	1379.44	4	**1485.96**
4.	***value***	*v/n*	*B1*	3925	1704.22	4	**1455.36**
5.	***energy***	*n*	*B1*	5630	2444.53	4	**1385.48**
6.	***result***	*v/n*	*B1*	2170	942.209	4	**1202.46**
7.	***call***	*v/n*	*A2*	2409	1045.98	4	**1200.78**
8.	***produce***	*v*	*B1*	2639	1145.84	4	**1140.43**
9.	***function***	*n*	*B2*	3888	1688.16	4	**1056.58**
10.	***move***	*v*	*A2*	2666	1157.57	4	**1027.43**
11.	***increase***	*v/n*	*B1*	2469	1072.03	4	**1017.93**
12.	***follow***	*v*	*A2*	1957	849.72	4	**979.89**
13.	***constant***	*adj*	*B2*	2514	1091.57	4	**955.75**
14.	***large***	*adj*	*A2*	1951	847.12	4	**946.56**
15.	***system***	*n*	*B1*	3056	1326.91	4	**907.43**
16.	***cell***	*n*	*B2*	5311	2306.02	4	**879.46**
17.	***determine***	*v*	*C1*	1876	814.55	4	**874.27**
18.	***describe***	*v*	*A2*	1646	714.69	4	**864.24**
19.	***mass***	*n*	*B2*	3384	1469.32	4	**859.34**
20.	***force***	*v/n*	*B2*	4023	1746.77	4	**859.17**
21.	***occur***	*v*	*B2*	1834	796.31	4	**852.65**
22.	***solution***	*n*	*B1*	3048	1323.43	4	**830.46**
23.	***high***	*adj*	*A2*	1682	730.32	4	**765.28**
24.	***contain***	*v*	*B1*	1583	687.33	4	**752.13**
25.	***line***	*n*	*A2*	2327	1010.37	4	**749.05**
26.	***molecule***	*n*		3143	1364.68	4	**741.08**
27.	***unit***	*n*	*B1*	1682	730.32	4	**728.97**
28.	***surface***	*n*	*B2*	2469	1072.03	4	**726.46**
29.	***section***	*n*	*B1*	1381	599.62	4	**723.97**
30.	***consider***	*v*	*B1*	1315	570.97	4	**709.20**

The Science Textbooks Wordlist is constructed by rank-ordering the words according to the average reduced frequency value (ARF), which combines frequency and dispersion into a single measure (Savický & Hlaváčová 2002), to highlight the most frequent and evenly dispersed items. At the top of the list is the word "point", labelled as a noun of A2 level, which occurred 4730 times in the corpus of science textbooks, and would occur 2053.75 times per million words. With an ARF value of 1647.80, it is the most frequent and most evenly dispersed item in the list, appearing in all of the subcorpora (Rel. DOCF=100).

Of the total 1195 items, apart from the 450 unlisted words, which were not categorised under any CEFR level, B2 level words (n=269) constitute the majority of the list, followed by B1 level words (n=193), C1 level words (n=119), A2 level words (n=93), and C2 level words (n=71). Below is a graphic illustration of the distribution of the items according to their CEFR levels.

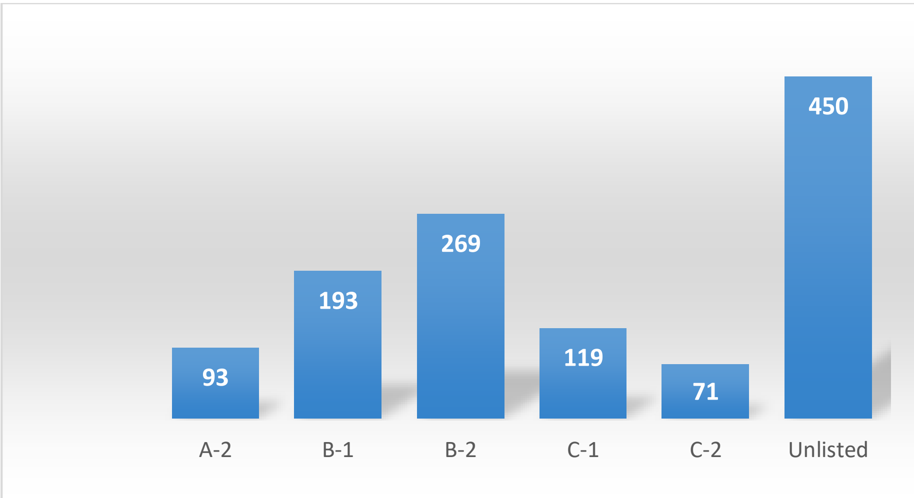

Figure 2. Distribution of the wordlist items as per CEFR levels

The top 30 items in the list include words of various levels, mostly being A2, B1 and B2 levels. Within these 30 items, there are two C1 level words (*equation, determine*) and one uncategorised word (*molecule*). The items with lower average reduced frequency values are mostly higher level items or those that could not be categorised as per CEFR levels. Below are the last 30 items on the list, where it is possible to notice that the less frequent items turn out to be more specialised.

Table 18

The last 30 headwords on the keyword list

N	Item	*PoS*	*CEFR level*	Frequency (focus)	Relative frequency (focus)	DOCF (focus)	ARF (focus)
1.	**micrograph**	*n*	*Unlisted*	42	18.23	2	**10.87**
2.	**watery**	*adj*	*Unlisted*	45	19.53	2	**10.83**
3.	**rupture**	*v/n*	*Unlisted*	30	13.02	3	**10.82**
4.	**parabolic**	*adj*	*Unlisted*	40	17.36	2	**10.78**
5.	**reactor**	*n*	*Unlisted*	58	25.18	3	**10.75**
6.	**superposition**	*n*	*Unlisted*	48	20.84	2	**10.65**
7.	**outermost**	*adj*	*Unlisted*	37	16.06	4	**10.63**
8.	**elementary**	*adj*	*B1*	52	22.57	3	**10.51**
9.	**buoyant**	*adj*	*Unlisted*	63	27.35	4	**10.50**
10.	**conductivity**	*n*	*Unlisted*	45	19.53	3	**10.48**
11.	**subunits**	*n*	*Unlisted*	69	29.95	2	**10.44**
12.	**ellipse**	*n*	*Unlisted*	121	52.53	2	**10.41**
13.	**nutrition**	*n*	*C1*	36	15.63	2	**10.40**
14.	**lightbulb**	*n*	*Unlisted*	66	28.65	2	**10.33**
15.	**fetus**	*n*	*Unlisted*	62	26.92	2	**10.31**
16.	**endpoint**	*n*	*Unlisted*	77	33.43	2	**10.28**
17.	**nucleic**	*adj*	*Unlisted*	44	19.10	2	**10.28**
18.	**algebra**	*n*	*Unlisted*	30	13.02	3	**10.27**
19.	**dissociate**	*v*	*Unlisted*	51	22.14	2	**10.22**
20.	**continuity**	*n*	*C2*	55	23.88	3	**10.21**
21.	**logarithmic**	*adj*	*Unlisted*	47	20.40	4	**10.20**
22.	**magnification**	*n*	*Unlisted*	79	34.30	2	**10.18**
23.	**endangered**	*adj*	*B2*	45	19.53	2	**10.17**
24.	**prefix**	*n*	*B2*	156	67.73	3	**10.16**
25.	**recycle**	*v*	*B1*	40	17.36	3	**10.13**
26.	**arctic**	*adj*	*Unlisted*	49	21.27	3	**10.09**
27.	**pea**	*n*	*B1*	50	21.70	3	**10.08**
28.	**spacing**	*v/n*	*Unlisted*	34	14.76	3	**10.04**
29.	**semicircle**	*n*	*Unlisted*	45	19.53	2	**10.04**
30.	**predatory**	*adj*	*Unlisted*	41	17.80	2	**10.00**

Multi-word terms were also extracted using the same tool in the Sketch Engine software. The list, based on a keyness score obtained through a mathematical method for identifying keywords of one corpus vs another, yielded 892 items. The items that are considered irrelevant, such as *function f, x axis, etc.*, as well as those items appearing in fewer than two sub-corpora were removed from the list. The final list was comprised of 379 multi-word items.

Table 15 below demonstrates the first 30 multi-word terms appearing on the list. The full list of multi-word units is available in Appendix D.

Table 19

The 30 most frequent multi-word units in the Science Textbooks Corpus.

Item	Frequency	Relative frequency	Document Frequency	Average Reduced Frequency
1. time interval	464	201.46	4	121.37
2. kinetic energy	564	244.88	4	99.75
3. electric field	904	392.51	4	96.29
4. magnetic field	912	395.98	4	77.94
5. straight line	169	73.37	4	69.96
6. potential energy	389	168.90	3	58.30
7. chemical reaction	195	84.66	4	54.80
8. hydrogen atom	269	116.79	3	49
9. surface area	150	65.12	4	46.66
10. internal energy	269	116.79	3	45.26
11. maximum value	167	72.51	3	43.93
12. water molecule	211	91.61	3	40.63
13. rate of change	155	67.30	4	40.43
14. amino acid	269	116.79	2	38.33
15. force act	174	75.55	2	37.38
16. carbon atom	285	123.74	3	36.77
17. center of mass	302	131.12	2	36
18. gravitational force	227	98.56	2	33.55
19. positive charge	154	66.86	3	33.30
20. total energy	147	63.82	3	32.94
21. same value	66	28.65	3	32.88
22. amount of energy	109	47.32	4	32.78
23. constant speed	123	53.40	4	32.56
24. blood cell	207	89.87	2	32.55
25. boiling point	171	74.24	4	32.52
26. negative sign	77	33.43	3	30.72
27. function of time	127	55.14	3	30.70
28. numerical value	62	26.92	3	29.32
29. high temperature	81	35.17	3	29.24
30. potential difference	317	137.64	2	28.28

As can be seen from the list, most multi-word units found frequently in the Science Textbooks Corpus appear to be specific to the scientific domain; for instance, *time* *interval, kinetic energy, electrif field* etc. are probably not found widely in a general corpus.

4.3.4. Coverage Analysis

Research Question 1.3. To what extent does the content of the English preparatory programme meet the target lexical needs of freshman engineering students in the science courses?

Once a keyword list was obtained based on the criteria of frequency, keyness and dispersion, it was then necessary to explore the extent to which the items on the list overlap with those taught at the English preparatory programme. To this end, the *Science Textbooks Keyword List* is compared with the target vocabulary used in the preparatory programme through AntWord Profiler, a tool for profiling vocabulary levels and text complexity. The analysis yielded a 12.60 % coverage value. To make it more explicit, 12.60 percent of the words in the *Science Textbooks Wordlist* (STWL) also occur in the wordlist taught in the preparatory programme. Of the 1195 items in the STWL, 151 items overlap, that is they occur in both wordlists, whereas 1044 items do not appear in the list of the words taught at the preparatory programme.

Research Question 1.4. How does a keyword list based on a corpus of science textbooks relate to the commonly available wordlists, namely the New General Service List, the Academic Vocabulary List and the Science Word List?

The keyword list is assumed to be a specialised wordlist consisting of items peculiar to the textbook content of natural sciences- Physics, Chemistry, Biology and Calculus. In order to find out how the keyword list relates to the New General Service List (Brezina and Gablasava, 2015), the New Academic Vocabulary List (Gardner and Davies, 2014) and the Science Word List (Coxhead and Hirsch, 2007), a comparison is performed by means of AntCont Profiler.

The comparative profiling of the above-mentioned wordlists with the Science Textbook Wordlist shows the extent to which there is an overlap between them. The following table shows the percentages covered by each wordlist in the *Science Textbooks Wordlist*, created from the Science Textbooks Corpus, in this study.

Table 20
Coverage values of the wordlists

Wordlist	**Coverage percentage**
New GSL (Brezina and Gablasava, 2015)	32.20 %
New Academic Vocabulary List (Gardner and Davies, 2014)	30.8 %
Science Word List (Coxhead and Hirsch, 2007)	13.30 %

As can be seen from the table, 32.20 percent of the words in the STWL occur in the New GSL; in other words, there is a 32.20-percent overlap between the specialized wordlist and the general service list. When it comes to academic vocabulary, the results show that 30.8 percent of the items in the STWL also appear in the New Academic Vocabulary List, which is slightly lower than the coverage value of the New GSL. In other words, the STWL covers 30.8 percent of the words in the New Academic Vocabulary List. The coverage value for the Science Word List is 13.30 percent, which means 13.30 percent of the words in the *Science Textbooks Word List* also occur in the Science Word List. Given that both wordlists are considered to be specialised wordlists extracted from corpora of scientific domain, the coverage value appears to be relatively low. This might be attributed to the wide range of scientific disciplines in the corpus data on which the Science Word List is based whereas the keyword list developed in this study was created on a more limited number of scientific disciplines, namely physics, chemistry, biology and calculus.

4.4. Findings from the Questionnaire

RQ.1.5. What are the course instructors' perceptions on the usefulness of the items in a key word list based on a corpus of science textbooks?

Using corpus-derived information together with subjective criteria can result in wordlists that is of more value for L2 learning and teaching purposes when compared to depending only on objective data (Dang, 2020). With that in mind, the wordlist derived from the science textbooks corpus was subject to expert opinion. A total of eleven lecturers giving the science courses responded to questionnaires with a different set of individual words. The first questionnaire was composed of the top 223 most frequent items on the list and the other four questionnaires consisted of 210 items each. The rationale behind designing the questionnaire in such a way that each lecturer would work on a different set of data was the desire to get all the items on the list rated by an expert. Sampling would leave out a number of lexical items and a decision on which items to include in the questionnaire would not be made without sacrificing others. Within the purposes of this study, it was intended to construct a pedagogically solid wordlist by employing objective and subjective data; therefore, each item derived from the corpus analysis were used in the qualitative data collection. For the multi-word items, three lecturers provided their opinions on the Likert scale.

Statistical Analysis of Questionnaire 1

The first set of items in Questionnaire 1 (n=223) were analysed statistically and the mean scores were obtained. Table 21 below shows the descriptive statistics for them.

Table 21

Statistical Findings for Questionnaire 1

Descriptive Statistics

	Q1.P1	**Q1.P2.**	**Q1.P.3**	**ARF**	**Ave.Score**
Valid	223	223	213	223	223
Missing	0	0	10	0	0
Mean	4.220	3.395	4.723	327.971	4.085
Std. Error of Mean	0.079	0.108	0.031	16.790	0.067
Std. Deviation	1.186	1.610	0.449	250.734	0.995
Shapiro-Wilk	0.687	0.804	0.560	0.731	0.846
P-value of Shapiro-Wilk	< .001	< .001	< .001	< .001	< .001
Minimum	1.000	1.000	4.000	128.004	1.000
Maximum	5.000	5.000	5.000	1547.456	5.000

The responses of three participants were calculated taking the average of their ratings. As is seen in the table, the average rating for the items in the questionnaire is 4.085 with a standard deviation of value of .995 and standard error value of .067. Given that the items were rated from 1 (not useful at all) to 5 (extremely useful) on a Likert scale, the mean value of 4.085 appears to indicate that the items were mostly found very useful. The mean scores obtained for each participant are 4.220, 3.395, 4.723 respectively.

The average ratings of the participants and the frequency values of the items were checked for correlation using the correlation coefficient *Pearson's r*. Gomes (2013) explains the correlation coefficient as "a kind of measure of the degree of linear relationship between the variables" (p.60). It can take on a value between plus and minus one: $-1 \leq r \leq +1$. A value of $r = +1$ is obtained if high values of one variable are associated with high values of the second variable; in other words, if the value of one variable increases, the value of the other increases too (Gomez, 2013). The statistical analysis of the two variables, namely the average rating scores and the average reduced frequency values of the lexical items in the first questionnaire generated a correlation value of .099 as shown in the table below.

Table 22

Pearson's Correlations for Questionnaire 1

Pearson's Correlations			
Variable		**ARF**	**Ave.Score**
1. ARF	Pearson's r	—	
	p-value	—	
2. Ave.Score	Pearson's r	0.099	—
	p-value	0.139	—

Pearson's r value of .099 shows that there is a weak correlation between these two variables; expressed another way, the average rating score and the average reduced frequency do not correlate significantly. The usefulness ratings of the participants do not increase or decrease in parallel with the ARF values of the items.

Although the mean score for the rating of the items appear to be high, a closer look into the list reveals that some items have received scores lower than 3, which correspond to "not useful" and "not useful at all" in the Likert scale used in the questionnaire. 29 items considered to be not useful are shown in Table 23.

The items are highlighted in the list for further consideration. There exists a variety of options for the items which were rated as not useful. They can either be excluded from the list or ranked with less priority within the list, or kept intact. This will be discussed in more detail in the discussion chapter of the study.

Table 23

Items with scores below 3 in Questionnaire 1

N	**Item**	*CEFR Level*	P1	P2	P3	**Average Rating Score**	**Average Reduced Frequency**
1.	**cell**	*B2*	1	1		**1**	**879.46**
2.	**mass**	*B2*	1	1		**1**	**859.34**
3.	**molecule**	*unlisted*	1	1		**1**	**741.08**
4.	**force**	*B2*	2	1		**1.5**	**859.17**
5.	**ion**	*unlisted*	1	1	4	**2**	**343.75**
6.	**molecular**	*unlisted*	1	1	4	**2**	**162.28**
7.	**organism**	*unlisted*	1	1	4	**2**	**152.17**
8.	**kinetic**	*unlisted*	1	1	4	**2**	**137.01**
9.	**nucleus**	*unlisted*	1	1	4	**2**	**131.48**
10.	**atomic**	*B2*	1	1	4	**2**	**129.10**
11.	**mole**	*unlisted*	1	1	4	**2**	**128.41**
12.	**reaction**	*B2*	3	1		**2**	**662.37**
13.	**species**	*B2*	1	1	5	**2.3**	**150.50**
14.	**particle**	*C2*	2	1	4	**2.3**	**410.94**
15.	**chemical**	*B2*	2	1	4	**2.3**	**378.02**
16.	**bond**	*B2*	2	1	4	**2.3**	**283.49**
17.	**human**	*B1*	2	1	4	**2.3**	**276.18**
18.	**liquid**	*B1*	2	1	4	**2.3**	**230.84**
19.	**datum**	*unlisted*	2	1	4	**2.3**	**219.75**
20.	**internal**	*B2*	2	1	5	**2.6**	**156.34**
21.	**resistance**	*C2*	2	1	5	**2.6**	**142.77**
22.	**earth**	*B1*	3	1	4	**2.6**	**412.14**
23.	**heat**	*B1*	3	1	4	**2.6**	**271.22**
24.	**color**	*unlisted*	3	1	4	**2.6**	**191.64**
25.	**store**	*B1*	3	1	4	**2.6**	**159.28**
26.	**physical**	*B2*	2	2	4	**2.6**	**156.59**
27.	**maintain**	*B2*	3	1	4	**2.6**	**156.49**
28.	**wire**	*B2*	3	1	4	**2.6**	**147.14**
29.	**surround**	*B1*	3	1	4	**2.6**	**145.99**

Statistical Analysis of Questionnaire 2

The statistical analysis of the second set of items (n=220) yielded similar results in terms of average rating score. The mean score was found to be 3.907, which is close to 4, corresponding to the category of "very useful" in the Likert scale. The mean scores of the first and the second participant are 3.800 and 4.014 respectively. This indicates that the items were found mostly very useful by the teachers. The standard deviation value is reported as 0.979 and the standard error of mean as 0.066. The table below demonstrates the statistical findings for Questionnaire 2.

Table 24

Statistical Findings for Questionnaire 2.

Descriptive Statistics				
	Q2.P1	**Q2.P2**	**ARF**	**Average Score**
Valid	220	220	220	220
Missing	0	0	0	0
Mean	3.800	4.014	83.557	3.907
Std. Error of Mean	0.071	0.077	1.339	0.066
Std. Deviation	1.049	1.141	19.859	0.979
Shapiro-Wilk	0.858	0.787	0.924	0.897
P-value of Shapiro-Wilk	< .001	< .001	< .001	< .001
Minimum	1.000	1.000	57.041	1.500
Maximum	5.000	5.000	127.294	5.000

In terms of correlation, Pearson's correlation coefficient was found to be -0.028, as can be seen in the table below. The *r* value of -0.028 indicates a negative correlation which is very weak. That is to say, there is a very weak negative correlation between the expert rating and corpus frequency.

Table 25

Pearson's Correlations for Questionnaire 2

Pearson's Correlations			
Variable		**ARF**	**Average Score**
1. ARF	Pearson's r	—	
	p-value	—	
2. Average Score	Pearson's r	-0.028	—
	p-value	0.677	—

The dataset for the second questionnaire was examined closely, and the items that received an average score below 3 are identified. 27 items that were rated as either "not useful" or "not useful at all" are shown in the table below.

Table 26

Items with scores below 3 in Questionnaire 2

N	Item	CEFR	Q2.P1	Q2 P2	Average Score	ARF
1.	**rod**	*unlisted*	1	2	**1.5**	98.65
2.	**iron**	*B1*	2	1	**1.5**	92.03
3.	**hole**	*B1*	2	1	**1.5**	72.53
4.	**copper**	*B2*	1	2	**1.5**	71.45
5.	**polar**	*unlisted*	2	1	**1.5**	69.69
6.	**loop**	*unlisted*	2	1	**1.5**	60.57
7.	**molar**	*unlisted*	2	1	**1.5**	57.04
8.	**sketch**	*C1*	2	2	**2**	102.48
9.	**slope**	*B2*	2	2	**2**	91.73
10.	**rock**	*B1*	2	2	**2**	79.11
11.	**arrow**	*B2*	2	2	**2**	78.14
12.	**tangent**	*unlisted*	2	2	**2**	77.15
13.	**wavelength**	*C2*	2	2	**2**	69.90
14.	**membrane**	*unlisted*	3	2	**2.5**	111.48
15.	**initially**	*B2*	2	3	**2.5**	105.84
16.	**edge**	*B1*	2	3	**2.5**	103.91
17.	**orbital**	*unlisted*	2	3	**2.5**	88.02
18.	**gravitational**	*unlisted*	2	3	**2.5**	81.61
19.	**mechanical**	*B2*	2	3	**2.5**	81.44
20.	**ocean**	*B1*	3	2	**2.5**	77.68
21.	**beam**	*B2*	3	2	**2.5**	75.75
22.	**vessel**	*unlisted*	3	2	**2.5**	68.10
23.	**spherical**	*unlisted*	2	3	**2.5**	66.57
24.	**voltage**	*unlisted*	2	3	**2.5**	63.26
25.	**ionic**	*unlisted*	3	2	**2.5**	60.73
26.	**mate**	*B1*	3	2	**2.5**	60.71
27.	**vapor**	*unlisted*	3	2	**2.5**	60.43

Statistical Analysis of Questionnaire 3

The third set of items subject to expert opinion included 220 words and was rated by two teachers. The mean score was found to be 3.6 which can be considered to correspond to the "useful" category in the Likert scale. The mean scores of the first and the second participant are 3.5 and 3.6 respectively with a standard deviaton value of .779 and standard error of mean value of .053. The table below demonstrates the descriptive statistics for Questionnaire 3.

Table 27

Statistical Findings for Questionnaire 3.

Descriptive Statistics				
	Q3.P1.	**Q3.P2.**	**ARF**	**Average**
Valid	220	220	220	220
Missing	0	0	0	0
Mean	3.505	3.695	41.998	3.600
Std. Error of Mean	0.071	0.067	0.468	0.053
Std. Deviation	1.058	0.999	6.938	0.779
Shapiro-Wilk	0.870	0.881	0.957	0.937
P-value of Shapiro-Wilk	< .001	< .001	< .001	< .001
Minimum	2.000	1.000	31.065	1.500
Maximum	56.913	5.000	5.000	5.000

The analysis for correlation between the average rating score and average reduced frequency values resulted in a value of .044. Pearson's coefficient of .044 shows that there is almost no correlation between the two variables. In other words, the increase or decrease of the scores do not act together; there appears to be no relationship between the objective frequency data obtained from the corpus and the intuitive data obtained from the teachers. Table 28 shows the correlation values for Questionnaire 3.

Table 28

Pearson's Correlations for Questionnaire 3.

Pearson's Correlations			
Variable		**ARF**	**Average Score**
1. ARF	Pearson's r	—	
	p-value	—	
2. Average	Pearson's r	0.044	—
	p-value	0.515	—

The items on the list that received an average score below 3 are identified and highlighted. These items (n=36), which are found "not useful" by the teacher participants who responded to the questionnaire can be reconsidered for inclusion in the list. For this reason, they are shown with an asterix in the full keyword list which is availabe in Appendix C. The items receving an average score below 3 can be found in table 29 below.

Table 29

Items with scores below 3 in Questionnaire 3

N	Item	CEFR	Q3.P2.	Q3.P1.	Average Rating Score	Average Reduced Frequency
1.	**aqueous**	*Unlisted*	1	2	**1.5**	55.50
2.	**axe**	*Unlisted*	1	2	**1.5**	41.65
3.	**oxide**	*Unlisted*	2	2	**2**	53.13
4.	**oxidation**	*Unlisted*	2	2	**2**	50.34
5.	**chromosome**	*Unlisted*	2	2	**2**	49.05
6.	**acidic**	*Unlisted*	2	2	**2**	47.71
7.	**cation**	*Unlisted*	2	2	**2**	46.10
8.	**tract**	*Unlisted*	2	2	**2**	43.44
9.	**predator**	*C1*	2	2	**2**	41.56
10.	**fiber**	*Unlisted*	2	2	**2**	37.61
11.	**radioactive**	*Unlisted*	2	2	**2**	37.53
12.	**lung**	*B2*	2	2	**2**	35.55
13.	**chamber**	*Unlisted*	2	2	**2**	34.77
14.	**infection**	*B2*	2	2	**2**	32.49
15.	**photosynthesis**	*Unlisted*	2	2	**2**	31.15
16.	**bone**	*B1*	3	2	**2.5**	55.51
17.	**pump**	*B1*	2	3	**2.5**	53.92
18.	**sunlight**	*B2*	3	2	**2.5**	53.78
19.	**seed**	*B2*	3	2	**2.5**	51.11
20.	**biological**	*B2*	3	2	**2.5**	50.31
21.	**steel**	*B2*	2	3	**2.5**	48.86
22.	**tank**	*C1*	2	3	**2.5**	46.59
23.	**cellular**	*Unlisted*	3	2	**2.5**	44.68
24.	**ionization**	*Unlisted*	2	3	**2.5**	42.79
25.	**climate**	*B1*	3	2	**2.5**	41.36
26.	**receptor**	*Unlisted*	3	2	**2.5**	40.21
27.	**pathway**	*Unlisted*	3	2	**2.5**	39.48
28.	**evolutionary**	*Unlisted*	3	2	**2.5**	39.31
29.	**prey**	*C2*	3	2	**2.5**	39.16
30.	**solvent**	*Unlisted*	3	2	**2.5**	35.30
31.	**immune**	*C2*	3	2	**2.5**	35.06
32.	**gland**	*Unlisted*	3	2	**2.5**	33.59
33.	**diffuse**	*Unlisted*	2	3	**2.5**	32.45
34.	**solute**	*Unlisted*	3	2	**2.5**	31.68
35.	**fusion**	*Unlisted*	3	2	**2.5**	31.36
36.	**cluster**	*Unlisted*	3	2	**2.5**	31.25

Statistical Analysis of Questionnaire 4

The statistical analysis of the items in the fourth questionnaire yielded a mean score of 3.732, which would correspond to score 4 (very useful) in the Likert Scale. The mean scores given by the participants are 3.959 and 3.505 respectively. The standard deviation is 0.774 and standard error of mean value is 0.052.

Table 30
Statistical Findings for Questionnaire 4.

Descriptive Statistics				
	Q3.P1.	**Q3.P2.**	**ARF**	**Average Score**
Valid	220	220	220	220
Missing	0	0	0	0
Mean	3.959	3.505	23.239	3.732
Std. Error of Mean	0.064	0.065	0.248	0.052
Std. Deviation	0.948	0.963	3.675	0.774
Shapiro-Wilk	0.851	0.868	0.947	0.931
P-value of Shapiro-Wilk	< .001	< .001	< .001	< .001
Minimum	1.000	1.000	17.891	1.000
Maximum	5.000	5.000	31.021	5.000

Pearson’s correlation value for the ARF and the average participant ratings is 0.034, which indicates that the variables are not significantly correlated.

Table 31
Pearson’s Correlations for Questionnaire 4

Pearson's Correlations			
Variable		**ARF**	**Ave.Score**
1. ARF	Pearson's r	—	
	p-value	—	
2. Ave.Score	Pearson's r	0.034	—
	p-value	0.613	—

Upon the analysis of the items that received a score lower than 3 (useful), which were thought to be not useful, it was found that most of the items- 18 items out of 25- were not categorised according to CEFR level, which might be indicative of the fact that the words are of technical, specialised nature. Some examples to words of this sort are, *tween, torque, pendulum, curvature,* etc. It may be considered necessary to exclude such words from the wordlist considering the teachability criteria.

Table 32
Items with scores below 3 in Questionnaire 4

N	Item	CEFR	Q4.P1	Q4.P2.	Ave.Score	ARF
1.	**tween**	*Unlisted*	1	1	**1**	29.84
2.	**torque**	*Unlisted*	2	1	**1.5**	30.23
3.	**frog**	*B1*	2	1	**1.5**	19.40
4.	**aquatic**	*Unlisted*	2	1	**1.5**	18.99
5.	**tropical**	*B2*	2	2	**2**	18.99
6.	**liver**	*B2*	2	2	**2**	18.85
7.	**astronaut**	*Unlisted*	2	2	**2**	17.98
8.	**cross-sectional**	*Unlisted*	3	2	**2.5**	30.25
9.	**fuse**	*Unlisted*	3	2	**2.5**	29.66
10.	**intestine**	*Unlisted*	2	3	**2.5**	27.78
11.	**intersection**	*Unlisted*	2	3	**2.5**	27.76
12.	**kidney**	*C2*	2	3	**2.5**	27.16
13.	**pendulum**	*Unlisted*	3	2	**2.5**	21.61
14.	**marine**	*Unlisted*	3	2	**2.5**	20.94
15.	**pollen**	*Unlisted*	3	2	**2.5**	20.90
16.	**curvature**	*Unlisted*	4	1	**2.5**	20.47
17.	**equator**	*Unlisted*	3	2	**2.5**	20.41
18.	**terrestrial**	*Unlisted*	3	2	**2.5**	20
19.	**node**	*Unlisted*	3	2	**2.5**	19.94
20.	**inherit**	*C2*	2	3	**2.5**	19.31
21.	**bounce**	*B2*	2	3	**2.5**	18.87
22.	**inward**	*Unlisted*	3	2	**2.5**	18.64
23.	**truck**	*B1*	3	2	**2.5**	18.52
24.	**repel**	*Unlisted*	3	2	**2.5**	18.07
25.	**valve**	*Unlisted*	3	2	**2.5**	17.89

Statistical Analysis of Questionnaire 5

The last set of items in Questionnaire 5 (n=220) were analysed statistically and the mean scores were obtained. The descriptive statistics results show that the average rating for the items in the questionnaire is 3.932 with a standard deviation value of .963 and standard error value of .065. Given that the items were rated from 1 (not useful at all) to 5 (extremely useful) on a Likert scale, the mean value of 3.932 appears to indicate that the items were found mostly useful. The mean scores obtained for each participant are 3.818 and 4.045 respectively. Table 33 shows the descriptive statistics for Questionnaire 5.

Table 33
Statistical Findings for Questionnaire 5.

Descriptive Statistics				
	Q5.P1.	**Q5.P.2**	**ARF**	**Average Rating**
Valid	220	220	220	220
Missing	0	0	0	0
Mean	3.818	4.045	13.729	3.932
Std. Error of Mean	0.107	0.062	0.155	0.065
Std. Deviation	1.589	0.925	2.298	0.963
Shapiro-Wilk	0.705	0.836	0.952	0.892
P-value of Shapiro-Wilk	< .001	< .001	< .001	< .001
Minimum	1.000	1.000	10.003	1.000
Maximum	5.000	5.000	17.875	5.000

The average ratings of the participants and the frequency values of the items were checked for correlation using the correlation coefficient *Pearson's r*. The statistical analysis of the two variables, namely the average rating scores and the average reduced frequency values of the lexical items in the first questionnaire generated a value of r value of .166 as shown in the table below.

Table 34
Pearson's Correlations for Questionnaire 5

Pearson's Correlations			
Variable		**ARF**	**Average Rating**
1. ARF	Pearson's r	—	
	p-value	—	
2. Average Rating	Pearson's r	0.166	—
	p-value	0.014	—

Pearson's r value of .166 indicates a correlation between these two variables; in other words, the mean rating scores of the participants are correlated with the corpus frequency values, though the correlation is not very strong. There appears to be a relation between the increase or decrease of participant ratings and the average reduced frequency values of the items.

The list is analysed in terms of low participant ratings, which might be useful for refining the list. The items the mean scores of which are below 3 are identified. The list below shows the words considered to be “not useful” or “not useful at all”.

Table 35

Items with scores below 3 in Questionnaire 5

N	**Item**	CEFR	Q5.P1.	Q5.P.2	**Average Rating**	ARF
1.	**pea**	*B1*	1	1	**1**	10.08
2.	**microbe**	*unlisted*	1	2	**1.5**	15.16
3.	**athlete**	*B1*	1	2	**1.5**	14.85
4.	**bladder**	*unlisted*	1	2	**1.5**	12.14
5.	**predatory**	*unlisted*	1	2	**1.5**	10
6.	**interstitial**	*unlisted*	1	3	**2**	17.87
7.	**antenna**	*unlisted*	1	3	**2**	14.80
8.	**telescope**	*B2*	1	3	**2**	13.18
9.	**foil**	*unlisted*	1	3	**2**	11.92
10.	**fetus**	*unlisted*	1	3	**2**	10.31
11.	**spacing**	*unlisted*	2	2	**2**	10.04
12.	**semicircle**	*unlisted*	2	2	**2**	10.04
13.	**trajectory**	*unlisted*	1	4	**2.5**	17.75
14.	**feather**	*B2*	2	3	**2.5**	16.82
15.	**nest**	*C2*	1	4	**2.5**	15.30
16.	**inertia**	*unlisted*	1	4	**2.5**	14.67
17.	**pulley**	*unlisted*	1	4	**2.5**	13.96
18.	**muscular**	*unlisted*	1	4	**2.5**	13.80
19.	**bead**	*unlisted*	1	4	**2.5**	13.68
20.	**spider**	*B1*	1	4	**2.5**	13.32
21.	**lizard**	*unlisted*	1	4	**2.5**	13.05
22.	**physiological**	*unlisted*	2	3	**2.5**	13.04
23.	**bee**	*B1*	1	4	**2.5**	12.97
24.	**corn**	*B1*	1	4	**2.5**	12.94
25.	**whale**	*B1*	1	4	**2.5**	12.70
26.	**vein**	*C1*	1	4	**2.5**	11.32
27.	**elongate**	*unlisted*	3	2	**2.5**	11.10
28.	**shark**	*unlisted*	1	4	**2.5**	10.92
29.	**watery**	*unlisted*	1	4	**2.5**	10.83
30.	**rupture**	*unlisted*	2	3	**2.5**	10.82
31.	**nutrition**	*C1*	2	3	**2.5**	10.40
32.	**arctic**	*unlisted*	2	3	**2.5**	10.09

These items can be considered to be excluded from the list; however, expert opinion can be received from a higher number of specialised lecturers for such a decision. Such items are shown with an asterix in the full list for further consideration.

Multi-word terms

It is also possible to reach frequency information regarding word combinations through corpus software. The words commonly found together were investigated through keyness analysis in the Sketch Engine programme for objective data. In order to triangulate the objective quantitative data, subjective intuitive ratings of teachers for the frequency of the corpus-derived multi-word terms were explored. Three teachers rated the usefulness of the 150 items most frequently found in the target corpus. The results are shown in Table 35 below.

Table 36

Statistical Findings for the Questionnaire on Multi-word units

Descriptive Statistics					
	ARF (focus)	**P1**	**P2**	**P3**	**Average Score**
Valid	150	150	150	150	150
Missing	0	0	0	0	0
Mean	24.046	2.727	4.027	3.367	3.373
Std. Error of Mean	1.298	0.053	0.074	0.064	0.040
Std. Deviation	15.900	0.654	0.904	0.789	0.488
Shapiro-Wilk	0.612	0.802	0.766	0.774	0.931
P-value of Shapiro-Wilk	< .001	< .001	< .001	< .001	< .001
Minimum	13.057	1.000	2.000	1.000	1.667
Maximum	121.378	4.000	5.000	5.000	4.333

As is seen from the descriptive statistics, the mean value for the multi-word terms is 3.373, which would correspond to "useful" category in the Likert scale. The standard deviation and standard error of mean are .488 and .040 respectively.

The subjective frequency ratings and the average reduced frequency values based on the corpus data were analysed for correlation. The results show that there is weak positive correlation (*r=.255)* between the variables. That is to say, the increase or decrease in the subjective ratings are independent of the increase or decrease in the objective frequency data. Table 37 below shows the Pearson's correlation statistics.

Table 37

Pearson's Correlations for Questionnaire on Multi-word units

Pearson's Correlations			
Variable		**ARF (focus)**	**Average Score**
1. ARF (focus)	Pearson's r	—	
	p-value	—	
2. Average Score	Pearson's r	0.255	—
	p-value	0.002	—

The multi-word items that received an average score below 3 are identified. These items are found to be not useful by the participants and can require reconsideration on whether to include them in the list or not. Table 32 shows those terms with an average rating score of below 3.

Table 38

Items with scores below 3 in Questionnaire on Multi-word units

	Item	**ARF**	**P1**	**P2**	**P3**	**Ave.Score**
1.	right triangle	14.02	2	3	1	2
2.	accompanying figure	14.16	3	3	1	2.3
3.	internal energy	45.26	2	3	3	2.6
4.	water molecule	40.63	2	3	3	2.6
5.	aqueous solution	39.21	3	2	3	2.6
6.	function of time	30.70	3	3	2	2.6
7.	ideal gas	27.57	2	4	2	2.6
8.	negative value	26.05	2	4	2	2.6
9.	multiple choice	25.54	2	3	3	2.6
10.	digestive tract	23.01	2	3	3	2.6
11.	electromagnetic radiation	13.36	2	3	3	2.6

CHAPTER V

DISCUSSION

5.1. Introduction

This chapter discusses the results obtained through the analysis of data collected for the purposes of this study in connection with the research questions. The findings are interpreted with regard to the research objectives, by elaborating on each research question.

5.2. Evaluation of Research Questions

RQ.1. What are the freshman engineering students' target lexical needs for the science courses?

The whole study is directed towards finding an answer to the first and single research question of "What are the freshman engineering students' target lexical needs for the science courses?, which aims to identify the engineering students' specific vocabulary needs in the science courses they take during their first-year studies. In an attempt to address this main research question, a number of sub-questions are devised, each of which are evaluated below.

RQ.1.1. What are the perceptions of the lecturing staff regarding the target needs of the freshman engineering students?

The first objective of the study, as articulated in the first research question, is to investigate the target needs of the first-year engineering students for the science courses they take as part of the must common courses within the curriculum, through the perceptions of the lecturing staff. To this end, an interview was designed consisting of questions with which it was intended to explore the requirements of the science courses, in other words, what the students would do with the language they learnt, as well as to establish what lacks keep them from attaining the course objectives. Interviews were conducted with seven lecturers, all of whom were giving the courses of Physics, Chemistry, Biology and Calculus. The content was analysed and the results were categorised under necessities, lacks and the lecturers' suggestions for the components of an effective English programme that would meet the specific needs of the freshman students.

Necessities

Regarding the necessities, it was evident from the findings obtained from the interviews that the fundamental requirement of the science courses is comprehension. It was reported to be absolutely necessary to comprehend both written and the spoken content of the courses, mostly based on textbooks and lectures, as well as to understand and answer exam questions. For an acceptable level of comprehension of materials, overall language competence is found to be necessary, as well as a deep knowledge of relevant vocabulary. With regard to vocabulary knowledge, this result is consistent with the results of the studies on the effects of lexical coverage on reading comprehension (Hu & Nation, 2000; Laufer, 1989; Schmitt et al., 2011) suggesting that there is a positive correlation between lexical knowledge and reading comprehension. Schmitt et al. (2011) for instance, found that reading comprehension scores tended to improve as lexical coverage increased above 90%.

Nation's (2006) seminal study of lexical profiling indicated that 6,000 to 7,000 word families were necessary to reach 98% lexical coverage of spoken text, and 8,000 to 9,000 word families were needed to reach 98% lexical coverage of written text. Also, studies that investigated the relationship between lexical coverage and listening comprehension (Bonk, 2000; Van Zeeland & Schmitt, 2013), show that higher lexical coverage ensures better comprehension. Bonk (2000) found that comprehension was best with lexical coverage above 90%, and Van Zeeland & Schmitt (2013) reported that most of the L2 learners had adequate comprehension of a listening passage at a lexical coverage of 90% and 95%. As is evident from these studies, vocabulary knowledge is a significant prerequisite for comprehension, be it written or spoken input. The lecturers interviewed clearly expressed their beliefs that the learners needed to become familiar with the vocabulary they would encounter in the course content in order to get to grips with the core meaning of a text, and this finding is in line with research results in literature.

The lectures given throughout the courses were reported to be mostly based on the textbook information, and thus contained identical content. The interviewees reported the significance of listening skill as a component that needed to be improved by the learners to achieve the course objectives. Given that "listening comprehension is difficult in a second or foreign language" (Lynch and Mendolsohn, 2013, p. 194), and that "remains one of the least understood processes" (Osada 2004, p. 53), it is inevitably worth putting more emphasis on the improvement of this skill. One reason for the challenges faced in listening comprehension is lack of vocabulary knowledge. If the words that the learner knows do not constitute a substantial part of the spoken content, the listening process turns out to be a problematic one with little comprehension. Bloomfield et.al (2010) state that "an obvious factor that can influence comprehension of a spoken passage is the overlap between the listener's vocabulary knowledge and the vocabulary of the passage" (p. 12).

Despite being emphasised by a smaller number of participants (*n=2*), it was deemed necessary that students be able to read equations and theorems properly, which also required a certain level of lexico-grammatical patterns knowledge.

Apart from the requirements of comprehending course content, understanding exam questions and reading equations and theorems, a small number of interviewees (*n=2*) stated that students needed to give presentations as part of the course requirements, which indicated a need for improving speaking skills. Presentation skills entail searching for the appropriate sources of information, comprehension, paraphrasing, summarising and speaking skills, most of which can be related to development of lexical knowledge, as well.

From what is summarised above, one can tentatively suggest that lexical knowledge is an essential construct that lies at the heart of the target needs to be developed by freshman students for attaining the course requirements, namely comprehension of both written and spoken scientific texts and assessment items, responding to exam questions, giving presentations and reading equations and theorems.

Lacks

As regards lacks, which refer to the gap between the target proficiency and the existing proficiency of learners (Hutchinson and Waters, 1987), interviewees' observations regarding the difficulties faced by students in meeting the requirements of the course were delved into. The majority of the lecturers pointed to the fact that students were not able to understand long sentences, exam tasks and high-level vocabulary. The students were also reported to have difficulty in expressing themselves. Two of the lecturers mentioned that they lacked presentation skills, which is somewhat related to the failure to produce speech to communicate ideas or information.

From the teachers' observations, it is possible to conclude that the students lacked overall linguistic competence as well as the necessary lexical knowledge that is needed to comprehend full sentences, as vocabulary constitutes an important component for comprehending written and spoken content. In order to be able to meet the course requirements, students need to improve their linguistic proficiency, deepen their knowledge of vocabulary they come across in the science courses, and improve their speaking and presentation skills.

Suggestions

Asked about their opinion on the possible ways of bridging the gap between the requirements of the courses and the learners' present level of performance, the interviewees pointed to four strands, namely better overall proficiency, familiarity with specific lexis, scientific reading and presentation skills. They expressed their view that students level of proficiency must be higher and they need to have a better knowledge of vocabulary specific to the discipline they are studying. Exposing students to scientific reading texts was suggested as a way of both improving their reading skills in a scientific discipline and familiarising them with specific vocabulary knowledge, which would be of help in attaining the course objectives.

In a nutshell, the findings of the interviews reveal that, from a lexical perspective, vocabulary knowledge is a pivotal element of academic studies in the scientific domain. From the results, it can also be inferred that it would be a fallacy to believe that a one-size-fits-all approach would benefit learners in a specific domain as the content specifications can greatly vary. Rather, it is necessary to adopt an approach that is fit for purpose and to tailor syllabi and materials to learners' specific needs.

RQ.1.2. What specific vocabulary do the science textbooks used by freshman engineering students feature?

The findings of the needs analysis made it clear that the freshman engineering students need to have a higher level of language proficiency and knowledge of vocabulary. In light of these findings, the next step of the study entailed establishing the specific vocabulary found in the science textbooks used by the students in question. To this end, the following research questions were also constructed:

RQ.1.2.1. What are the lexical frequency representations of the science textbooks used by freshman engineering students?

RQ.1.2.2. What keywords and multi-word terms constitute the key vocabulary in the science textbooks of freshman students?

In order to determine the discrepancies between what the students need to learn and what they actually learn, it was necessary to specify the vocabulary featuring the content of the target materials. In pursuit of specifying the lexical content the students are subject to, a corpus was compiled from the textbooks used by freshman engineering students. Gabrielatos (2005) states that textbook corpora allow us to examine language that the learners are exposed to in their studies and can lead to more pedagogically sound materials. The corpus built in this study is comprised of texts in the science textbooks, and thus can be regarded as a specialised corpus, which is believed to be of value in establishing the features of the specific domain. Specialised corpora, as noted by Koester (2010) "provide insights into the particular genres investigated, such as very specific types of scientific (e.g. environmental impact statements) or academic writing (e.g. letters of application)" (p. 68). The target corpus, referred to as the *Science Textbooks Corpus* for the purposes of this study, was compiled using the *Word Sketch Engine s*oftware. The corpus consists of 2,303,096 tokens and 1,898,324 words. The size of the corpus is in line with literature. While large corpora have certain advantages, small corpora can also be useful depending on the purpose it is used for. As noted by Flowerdew (2002), small corpora built for a specific purpose are more likely to provide insights relevant for teaching and learning for specific purposes. Trible (2002) also claims that large corpora do nor cater for the needs of ESP/EAP teachers and learners on the grounds that they provide "either too much data across too large a spectrum, or too little focused data, to be directly helpful to learners with specific learning purposes" (p. 132).

Through the corpus compiled, it was intended to establish the frequent lexical patterns occurring regularly within the textbook content. Frequently encountered items can be learned with ease. Jones and Durrant (2010) state that "the argument for prioritising vocabulary learning on the basis of frequency information is based on the principle that the more frequent a word is, the more important it is to learn" (p. 387). Also, learners can better remember items with a higher frequency of encounter (Trembley et al., 2008). The frequency analysis, which generated a list of 2954 items, was subject to revision where irrelevant items such as non-lexical items, abbreviations, function words, proper names, symbols etc. were removed and rank-ordered according to average reduced frequency value. The final list, the unit of analysis of which is "lemma", comprised of 1688 items with a document frequency of over 2.

Looking at the most frequent items on the list, e.g. *show, find, example, give, point, equation, form, value, see, energy, time, change, result,* one can hold that they are mostly words of general service which can also recurrently appear in a general corpus. Therefore, examining keyness would be rational given that it is the selection of domain-specific vocabulary that is aimed at in the study. To identify the items appearing more frequently in the target corpus than in a general corpus, the two corpora -the target corpus and the benchmark corpus- are compared, and thus a list of key words is generated. Evison (2010) posits that key words are “not necessarily the most frequent words in a corpus, but they are those words which are identified by statistical comparison of a ‘target’ corpus with another, larger corpus, which is referred to as the ‘reference’ or ‘benchmark’ corpus” (p. 127). The target corpus compiled for the purposes of this study, namely the *Science Textbooks Corpus*, was compared with the *BNC* corpus, for keyness. The analysis was based on “lemma” as the unit of analysis. The list, following the manual revision where the irrelevant items, erroneous entries and items with a document frequency value of below two were removed, comprised of 1249 lemmas, ranked according to average reduced frequency value (ARF) highlighting the most frequent as well as evenly dispersed items. Looking at the most frequent words on the list, top ten being *point, equation, form, value, energy, result, call, produce, function*, it can tentatively be suggested that the words appear to be discipline-specific but are not too technical or specialized. As the frequency of occurrence decreases, specificness of the word increases. For instance, the ten least frequent words in the list, *logarithmic, magnification, endangered, prefix, recycle, arctic, pea, spacing, semicircle, predatory*, seem to be more specific to the scientific domain.

The keyword list was then examined for CEFR levels. The items were tagged as *A1, A2, B1, B2, C1* and *C2* according to the Common European Framework levels scale. The items of A1 level were excluded from the list with the rationale that such words can be assumed to have already been learnt at earlier stages of instruction and that they did not bear the quality of being discipline-specific, and thus would not fit the purpose of the wordlist. The final list was reduced to 1195 items. The tagging of the items on the list showed that the majority of the words are B2 level words ($n=269$), followed by B1 ($n=193$), and C1 ($n=119$) level words, which may have an implication for the decision-making processes for course design in terms of exit level targeted for this group of students. Also, the list containing a remarkable number of items of each CEFR level indicates that it is not comprised of purely technical or high-level words that EFL teachers are not likely to be expert in, which again provides a significant baseline for decisions on programme development. Considering the fact that ESP teachers are usually not experts in the target specialised content and may not have background knowledge of the technical area (Sylven, 2013), such a lexical content, free of extreme technicality, would be applicable and practical. As noted by Dudley-Evans and St. John (1998), the ESP practitioner must embody five roles, which are teacher, course designer and materials provider, collaborator, researcher and evaluator. Therefore, it is important that the teaching content be manageable for the teacher. According to Hutchinson and Waters (1987), ESP teachers are "all too often reluctant dwellers in a strange and uncharted land". The reason for such reluctance most probably stems from having to teach in an unfamiliar context. Within the purposes of this study however, the lexical content is not of a very technical nature, which would not bring along an extra demand for the teacher in terms of domain knowledge. More frequent words tend to be of more familiar nature whereas less frequent ones tend to be more specialised and technical. Non-technicality of the most frequent words is similar to what Mudraya (2006) found in her study on lexical frequency. Mudraya (2006), in her study aiming to show how integrating the lexical approach with a corpus-based methodology could improve the way ESP is taught, compiled a corpus from textbooks used in basic engineering disciplines and ran a frequency and keyword analysis on the corpus data. The comparison of the corpus against the BNC Written Sampler showed that the most frequent words in a specialist corpus are sub-technical and non-technical from the academic register.

Trimble (1985) posits that, academic words can have extended meanings in technical contexts, and it is possible that words have totally different meanings in different disciplines. From this perspective, it is also important to note that some words of STWL which look like common words of general service or academic vocabulary might have been used in different senses, and with different patterns in the context of scientific texts.

The following two examples show how some words of high frequency are used in the target corpus (Science Textbooks Corpus) and in the benchmark corpus (British National Corpus).

Table 39

Two examples from the target and the reference corpora

	ARF	Relative freq. in STC	Relative freq. in BNC	KWIC occurrences in STC	KWIC occurences in BNC
base (n)	576	464,59	87,56	•*acid, base and buffer* •*nitrogen-containing base* •*acid-base reaction* •*base ionization* •*strong base* •*weak base*	• *triangular base* • *military base* • *US base* • *logical base* • *information base* • *knowledge base* • *logical base*
reaction (n)	662	1652	66.38	•*nuclear reactions* •*chemical reactions* •*atmospheric reactions* •*chain reaction* •*exergonic reaction* •*reverse reaction*	• *immediate reaction* • *my first reaction* • *sort of reaction* • *public reaction* • *customer reaction* • *excessive reaction* • *skin reaction*

The two words (*base* and *reaction*) selected from within the first 50 most frequently occurring items in the STWL were searched for contextual data. Looking at the first 100 hits produced with the KWIC (Keyword in Context) tool in the Sketch Engine, one can see that the words differ in the way they are used across the two corpora. Taking the example of *base* first, it can be inferred that the word is used in a different technical meaning in the Science Textbooks Corpus (STC), which probably refers to the main part of a substance to which other things are added. In the BNC, on the other hand, it refers to the lowest part of something, as in *triangular base,* or the main place from which an organisation controls their activities, as in *military base.*

The word *reaction* on the other hand is not used in a totally different meaning but it is apparently used in a different sense, which is more technical. The collocates of *reaction* in the Science Textbooks Corpus are mostly of scientific nature (*chemical, nuclear, atmospheric* etc.*)* which adds to the technicality of the word.

Such examples are in line with the point made by Hyland and Tse (2007) in that words can take on different or extended meanings in different disciplines. Specific groups in specific disciplines have a specialised vocabulary which needs to be considered in developing word lists or other materials in an attempt to contribute to L2 learning.

Multi-word terms

One of the central insights to come from corpus linguistics in the last thirty years is the extent to which competent language users draw not only on a lexicon of individual words, but also on a range of lexicalised phrasal units which have come to be known as 'formulaic sequences' (cited in Jones and Durrant, 2010; Wray 2002; Schmitt 2004). Wray (2002) defines formulaic sequence as "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar " (p. 9). Regarding the multi-word units and their examination in terms of keyness, Greaves and Warren (2010) state that "given that multi-word units are so pervasive in language, concgrams can be used to extend the notion of keyness beyond individual words to include the full range of multi-word units" ((p.221). Cheng et al (2009) explain the concept of "concgrams" as "instances of co-occurring words irrespective of whether or not they are contiguous, and irrespective of whether or not they are in the same sequential order". Cortes (2004) thinks that the competency in using multi-word units is an indicator of proficient language use in that specific register or genre. Similarly, Hyland (2008) also thinks that readers and writers participating regularly in a specific discourse are familiar with multi-word units and the absence of discipline-specific multi-word units can indicate a lack of fluency.

The corpus compiled for the purposes of this study is analysed in terms of recurrent multi-part words specifically found in the science textbooks. Formulaic phrases pertaining to a specific discipline can be distinctive from those in a general English context. Phraseology can yield insights about the specialised domain investigated; for instance, Gledhill's study (2000) showed that terminology involving collocations can mirror the recurrent semantics of the specialised domain and that phraseology is part of the defining characteristics of the discourse community (cited in Jones and Durrant, 2010). Nelson's study (2006) shows that words' semantic prosodies differ in business English context and general English context.

The items occurring together with outstanding frequency were identified by means of the software programme Sketch Engine. The analysis involved comparison of the *Science Textbooks Corpus* with the BNC corpus, in a similar fashion to the key word analysis. The items commonly co-occurring extracted from the analysis were revised manually by the researcher for the sake of excluding the irrelevant ones. The final list, comprising of 396 multi-word units, was rank-ordered according to their value of average reduced frequency. From the top ten items of frequency in the list, (*time interval, kinetic energy, electric filed, magnetic field, straight line, potential energy, chemical reaction, hydrogen atom, surface area, internal energy)* one can infer that the multi-word units also reflect the specialised language of the specific domain investigated. These phrases appear to be of more technical nature than the formulaic sequences commonly used in general English contexts. This profile is indicative of the fact that the content of the science textbooks used by engineering students in their first-year features a specialised language, which seems to be different from general English discourse. It is believed that acquiring the repertoire of multi-word units specified in this study can be valuable for learners by contributing to their proficiency.

In a nutshell, the study aimed to find out whether there is a specialised vocabulary in the scientific texts used in engineering discipline, which is different from the vocabulary of general English, and in pursuit of reaching this objective, a corpus was compiled, and subsequently, frequency and keyword analyses were conducted on the corpus data. Finally a key word list and a list of multi-word units were generated. The lists created are of considerable value in that they are corpus-derived, based on objective and quantitative data, and consolidated through subjective and qualitative data collected from teachers delivering the courses. Thus, they can potentially be used in a variety of areas, like curriculum development, syllabus design, material development, and test construction.

RQ.1.3. To what extent does the content of the English preparatory programme meet the lexical needs of freshman engineering students in the science courses?

Following the creation of a keyword list from the corpus compiled from the science textbooks used in the engineering departments, the list was compared with the vocabulary list of the EFL coursebook used by the engineering students during their preparatory education in preparatory school. With the comparison, it was intended to find out the extent to which the materials used in the English preparatory programme cover the lexical needs of engineering students in the science courses they take in their first year tertiary education. Discovering, if any, the discrepancies would triangulate the results of the needs analysis, where there was a remarkable call for the need of vocabulary knowledge for meeting the course requirements.

The likelihood that an EFL coursebook with a general English purpose could contain the lexis required in a specific discipline appears to be low. Also, despite the recent developments in corpus linguistics in terms of materials design, "course books have generally been slow to exploit corpora as a resource" (McCarten, 2010, p. 413). Coursebook developers tend to use content from their own resources, rather than corpus data, which brings along gaps between the "real" language and the language in the coursebook. Biber et. al. (1998) and Cheng and Warren (2007) noted that there are disparities between the language described and modelled in course books and real language use reflected in corpora.

The comparison in this study was made by means of *AntWord Profiler*, a profiling tool developed by Lawrence Anthony. The keyword list, referred to as *Science Textbooks Word List* for the purposes of this study, was compared with the list of the words taught in the EFL coursebook. The results revealed that there was a 12.60 per cent overlap between the two lists. This means that 12.60 percent of the 1195 items in the STWL are covered in the EFL coursebook list. In other words, 151 items in the STWK are taught in the EFL coursebook.

The results are indicative of the fact that the specific academic purposes of the discipline are not reflected in the general English curriculum content used in the preparatory programme, and thus the material fails to mirror the lexis that the students will be subject to during their studies. There exists a gap between the lexis required for the academic studies of engineering students and the lexis taught in the programme, which needs to be closed in order for a fit-for-purpose language education to be achieved. A curriculum that is supposedly designed for the students to meet the requirements of their studies must involve elements that are derived from their needs. Now that a lexical list generated based on key frequency values derived from corpus data is available, the stakeholders can consider tailoring a bespoke curriculum in line with the lexical needs of the students.

RQ.1.4. How does a keyword list based on a corpus of science textbooks relate to the commonly available wordlists, namely the New General Service List, the Academic Vocabulary List and the Science Word List?

It is intended with this research question to investigate how the word list developed in this study relate to three types of wordlists widely used in the field of language teaching, namely the *New General Service List* (Brezina and Gablasova, 2015), the *New Academic Vocabulary List* (Gardner and Davies, 2014) and the *Science Word List* (Coxhead and Hirsch, 2007), with the aim of determining the extent to which they overlap. These three wordlists were chosen for comparison with the rationale that each list represented a different domain; the *New General Service List* is, as its name suggests, has a focus of English for general purposes; the *New Academic Vocabulary List* represents the academic field, and the *Science Word List* is derived from a corpus in the scientific domain.

The comparison was performed through *Antword Profiler*. The coverage value for the New GSL was found to be 32.20 per cent, which means that 32.20 percent of the words in the STWL also appeared in the New GSL. A similar result was obtained for the New Academic Vocabulary List, which is 30.8 per cent. These two values are indicative of the fact that a certain proportion of the words in the STWL are words of general service and words used frequently in the academic domain. Yet still, the discipline-specific word list informed by the science textbooks corpus can be considered as differing widely from a general service list or an academic wordlist. This result is in line with what Martinez and his colleagues (Martinex et al., 2009) found in their study on identifying field-specific vocabulary through specialised corpus. Their study revealed that the specialised corpus they compiled in the field of agriculture, the AgroCorpus comprising of 826,416 words, contained only ninety-two word families from Coxhead's AWL.

When it comes to comparing the list with a specialised wordlist, namely the Science Word List, the analysis yields a coverage value of 13.30 percent, which is surprisingly low given that the two lists are from almost the same domain. The reason for this low coverage figure might be that the science wordlist represented a wider number of disciplines (*n=14*) whereas the STWL represented a smaller number of disciplines (*n=4*).

This comparison also consolidated the conclusion that there is a need for a discipline-specific, specialised word list peculiar to the domain investigated. The science courses taken in the engineering departments in the first year appear to feature a specific lexis, according to which English instruction should be shaped. Acknowledging the contribution of the wordlists like GSL, and AWL, Hyland and Tse (2007) believe that the problematic aspect of such lists is the assumption that a single inventory can be representative of the vocabulary of every academic discourse and thus be useful to all students regardless of their field of study. The findings of their study show that the coverage of AWL is not evenly distributed across the whole corpus, meaning that some items occur more frequently in certain disciplines, and that words can take on different meanings in different disciplines. Similarly, the results of this study also confirm the need for a wordlist featuring the specialised language of the specific domain investigated which the specific group of learners can benefit from.

RQ.1.5. What are the course instructors' perceptions on the usefulness of the items in a key word list based on a corpus of science textbooks?

<u>*Individual Words*</u>

In order to establish a useful and pedagogically solid wordlist that is based not only on purely quantitative, low-inference corpus data, but also on subjective, expert opinion, it was necessary to explore teachers' perceptions. Jones and Durrant (2010, p. 387) point to the importance of "substantial human guidance" as corpus software is not capable of building a word list that is pedagogically useful. Also, Dang et.al, (2022) state that "teacher perceptions of word usefulness can provide useful insight into the value of the items that make up a word list" (p. 622).

With the objective of triangulating the corpus-derived data, 11 teachers were asked about their opinion on the usefulness of the items on the wordlist. The participants responded to different sets of words. The following table lays out the statistical results for each questionnaire, namely the mean rating score, average reduced frequency score and Pearson's correlation value. Also shown in the table is the number of items that are scored below 3 by the respondents; expressed another way, the items that were found to be not useful.

Table 40

Summary of Questionnaire Findings

	Mean Rating Score	***Pearson's r***	***Average Reduced Score Mean***	***Number of items rated below 3***
Questionnaire 1 (223 items)	4.085	*0.099*	*327.971*	*29*
Questionnaire 2 (220 items)	3.907	*-0.028*	*83.5*	*27*
Questionnaire 3 (220 items)	3.600	*0.044*	*41.9*	*36*
Questionnaire 4 (220 items)	3.732	*0.034*	*23.239*	*25*
Questionnaire 5 (220 items)	3.932	*0.166*	*13.729*	*32*

* Total number of items: 1103

The findings indicate no significant correlation between teachers' ratings and the corpus frequency figures. The correlation values of *.099, -.028, .044, .034, .166* point to weak or no statistically significant correlation overall. This may be attributed to the relatively low number of participants whose intuitions are explored. However, this result does not contradict with recent studies which show low correlations between corpus figures and intuition figures. For instance, Alderson (2007) found a correlation value of *.67*, and Schmitt and Dunham (1999) found a value of *.53–.65*. Also notable is the high variability between raters in both studies (cited in Schmitt, 2010, p. 68). Brzoza (2018) compared objective frequency data of Polish and English words with the frequency judgments of L1 users and found a weak correlation between the two variables; thus, he suggested combining objective and subjective lexical frequency values. However, there are also studies which showed more significant correlations between objective and subjective lexical frequency measures. For example, Okamoto (2015) who explored the relationship between lexical frequency based on corpora and native speakers' ratings, concluded that the word frequency in corpora seemed to be closely related to native speakers' intuitions regarding word frequency. She recommended making use of both objective and subjective approaches to identifying target vocabulary and constructing word lists for EFL textbooks. According to McGee (2008), the divergence of corpus data and intuitive data is not surprising because different corpora can also differ on word frequencies, and he advises considering both corpus-based and intuitive-based data as useful. He and Godfroid (2019) found a moderate correlation between the frequency of academic words in the COCA and COCA Academic corpus and teacher perceptions of the usefulness of these words. The study conducted by Dang et al. (2022) investigated the usefulness of four well-known wordlists using teacher perceptions of word usefulness and learner vocabulary knowledge as the criteria. They found strong correlations between teacher groups and teachers perceived BNC/COCA2000 to have more useful words. The mixed results of the studies mentioned indicate that further research is necessary regarding the relationship between subjective intuitive frequency data and objective corpus-based frequency data.

Of the 1103 items in the STWL, where A2 level words were also excluded for collecting subjective data, 149 items received a score below 3, which means that 149 items were not useful according to the intuitions of the teacher participants. Such items can either be excluded from the list or can be rank-ordered in terms of priority accordingly. What to do with these items can be decided according to the purpose of using the wordlist by the user. Therefore, it was found appropriate to keep the items of this sort in the list and suggest that further data be collected for subjective frequency ratings.

Multi-word terms

Tremblay, et.al (2008) state that each time a lexical item is experienced it leaves a memory trace, and that this effect applies not only to individual words but also lexical items. In other words, as the frequency of lexical bundles increases, they are remembered better. Given that multi-word units are pervasive in language and that corpus data is limited in capturing the formulaic language uses, it can make sense to employ "the other main way of determining frequency – user intuitions" (Schmitt, 2010, p. 67).

In order to collect intuitive data on the multi-word units derived from corpus data, expert opinion was asked for the most frequent 150 items on the list. Three lecturers responded to the questionnaire. The descriptive statistics for these items indicate a mean score of *3.373*. The correlation between the objective frequency data and intuitive ratings of the teachers is weak (*r= 0.255*).

Studies focusing on intuitive frequency are limited in number, have mixed results and mostly focus on the differences between native and non-native participant ratings with regard to frequency of occurrence of lexical items. Hoffman and Lehmann (2000) explored native and non-native speaker intuitions about 55 word pairings with the aim of testing the sensitivity of non-native speakers to the frequency of words occurring together. They found that native speakers' predictions were 70% correct while non-native speakers' were only 30%. Siyanova and Schmitt (2008) focused on the intuitive ratings of the participants in terms of the frequency of collocations that had high, mid and low frequency profiles. They found that native speakers made more accurate predictions that are closer to the frequencies in the BNC, in comparison to non-native speakers and that the correlation between the native speakers intuitions and the corpus frequency values was high. In McGee's study (2009), where he compared the corpus frequency values with native English teachers' intuitions regarding the most frequent collocates of certain adjectives, the results showed that there was a great difference between teacher's intuitions and corpus data. In other words, there was no correlation between subjective and objective frequency measures. Siyanova-Chanturia and Spina

(2015) explored collocational frequency intuitions of the native and non-native Italians and corpus frequency data. The results showed that there was correlation between L1 group intuitions and corpus data in terms of low-frequency collocations, but a very strong correlation between both L1 and L1 speakers' intuitions and corpus data in terms of high frequency collocations. Also, the correlation between the variables in question was weak for medium and low frequency collocations whereas it was strong for the very low frequency collocations. A recent study by Cangır (2021) investigated the extent to which the association measures indicating collocational strength correlate with EFL instructors' intuitions regarding collocational frequency. The results show that there is a strong correlation between collocational frequency intuitions and objective collocational frequency measures extracted from corpus.

One can tentatively conclude from these studies that the results regarding the relationship between intuitive frequency data and objective corpus-based data have mixed results and the results obtained in this study, showing hardly any correlation between the subjective and objective frequency values, overall, are therefore not surprising. It is evident that the subjective frequency measures based on intuition and objective frequency measures based on corpus data require further investigation.

CHAPTER VI

CONCLUSION

6.1. Introduction

The last chapter of the study is dedicated to the summary of the research and the conclusions drawn from the whole research process, as well as the pedagogical implications of the study and suggestions for further research.

6.2. Summary of the Research

One of the most remarkable developments in vocabulary studies lately has been the exploration of corpus data as a basis for vocabulary research. With the advent of computerized analysis, corpus-based studies gained momentum and a number of large-scale corpora have been developed (e.g. BNC, COCA, CANCODE, MICASE, BASE, etc.). Frequency, one of the most important characteristics of vocabulary that affects many aspects of lexical acquisition (Schmitt, 2010), can ideally be explored in corpus data, and wordlists can be developed based on corpus frequency measures. Schmitt (2010) states that "language learners typically acquire higher frequency vocabulary before lower frequency vocabulary" (p. 14).

The fields of ESP (English for Specific Purposes) and EAP (English for Academic Purposes) have also benefited from the developments in corpus linguistics. According to Boulton et al. (2012), corpus-based ESP analysis is "evolving in promising directions and being gradually enriched by new methods and applications, and, true to its origins, by empirical investigations that have robust theoretical foundations" (p. 3). According to Gavioli (2005), "corpus tools or corpus-based approaches are part of growing amalgamation of technology and language learning for specific purposes" (p.17). Nesi (2012) also advocates the use of corpora in ESP studies since "they help make stronger and statistically supported claims" (p.420). Indeed, "one advantage of corpus studies is being able to gather and analyze a large amount of text including texts which actual learners and teachers use in their classrooms and courses" (Coxhead & Demecheleer, 2018, p. 87).

With the belief that it is students' specific target context that can provide most reliable data for making sound decisions as regards course curriculum, syllabus, materials and assessment components, this study is based on data collected from within the target context of engineering students. It is prompted by the assumption that engineering students need to acquire a specific lexis to perform their studies and that a bespoke wordlist representing the vocabulary prevalent in the texts used in the basic science courses they take would be of value for them. Therefore, the ultimate goal of the study was to identify the lexical needs of freshman engineering students for the must science courses they take, namely Physics, Chemistry, Biology and Calculus, and develop a needs-driven wordlist. To this end, a set of research questions were devised to reach the main goal of the study. These questions addressed the following issues:

- perceptions of the lecturing staff as regards the target lexical needs of the engineering students,
- the specific vocabulary featured in the science textbooks used by the engineering students**,**
- the extent to which the target lexical needs are covered in the English preparatory programme,
- the overlap between the specific lexis derived from the corpus and the new GSL, the new Academic Vocabulary List and the Science Word List
- the perceptions of lecturing staff in terms of the high-frequency lexical items needed by the engineering students

Needs analysis

The first stage of the study is comprised of a needs analysis with the purpose of identifying what the science courses' requirements are, what the students needed for meeting those requirements and what they lacked-that is the skills or knowledge they needed but did not possess. With the purpose of collecting data on the target needs, subject teachers were interviewed through questions devised according to the needs analysis framework of Hutchinson and Waters (1987). The data was analysed through content analysis and outstanding themes and patterns were identified. The results showed that, in addition to a good command of English proficiency, students needed to be familiar with the lexis they were to encounter in the science courses for better comprehension of written and spoken content.

Corpus compilation

In the second stage of the study, a corpus aimed at first-year engineering students was compiled with a view to creating a wordlist on which to base teaching materials, assessment constructs and other relevant decisions. The corpus was compiled of the Physics, Chemistry, Calculus and Biology textbooks used in the first-year studies, by means of the Sketch Engine software. For the compilation, all the material was converted into .txt format and the irrelevant data such as preface, table of contents, tables, figures, headings, and appendices were removed. Then, the files were uploaded into the Sketch Engine software and a corpus of 2,303,096 tokens and 1,898,324 words was created. The corpus, referred to as the *Science Textbooks Corpus* for the purposes of this study, consisted of 4 sub-corpora, each belonging to a different subject, namely physics, chemistry, calculus and biology.

Frequency and Keyness Analyses

Following the compilation of the corpus, a frequency analysis with a threshold frequency value of fifty was conducted in order to investigate which vocabulary is especially frequent in the collection of the textbooks. The frequency profiles were computed and the resulting list of items with a frequency value of over fifty consisted of 2954 items, which also included function words, such as *the, a, and* etc. The list was revised and irrelevant items were removed, which resulted in a final list of 1688 lexical items.

Following the frequency analysis, the next step was investigation of keyness. A keyword analysis was performed, the purpose of which is to identify the key words occurring with a remarkably higher frequency in the target corpus but occurring with a low frequency in a reference corpus. The target corpus –*Science Textbooks Corpus*- was compared with a reference corpus- the *British National Corpus.* The unit of analysis was *lemma* and the minimum document frequency was set as two on the grounds that the total number of sub-corpora is four and the items that were frequent were required to occur in at least half of the sub-corpora. The list generated by the computerized analysis was then revised manually for irrelevant items and erroneous entries. The list, consisting of 1249 lemmas, was rank-ordered according to average reduced frequency values (ARF). The items on the list were then categorized according to the CEFR levels. A1 level items were removed from the list given that items of this level cannot be specific to a certain discipline, but rather are words of general English, presupposed to have been learnt at earlier stages of the learning process. The final list, referred to as *Science Textbooks Word List* was reduced to 1195 lemmas.

The keyness analysis was also conducted for multi-word units appearing with a high frequency in the target corpus in comparison to the reference corpus. The list generated was checked for inconsistencies, irrelevant items and erroneous entries. Following the removal of the items, the final list consisted of 379 items.

Wordlist comparison

Having constructed a corpus-informed word list based on the criteria of frequency, dispersion and keyness, the Science Textbooks Word List was compared to the list of words taught in the preparatory programme, in order to investigate the extent to which the English programme content covers the target lexical items needed by the freshman engineering students. The analysis conducted by means of *AntWord Profiler* yielded a coverage value of 12.6 per cent; in other words, 12.6 per cent of the words in the STWL occurred in the list of vocabulary taught in the preparatory programme. Based on the coverage value obtained, one can conclude that the material used in the preparatory programme fails to provide a good coverage of the freshman engineering students' target lexical needs in the science courses they take.

With the objective of finding out how the *Science Textbooks Word List* relates to the new GSL, the Academic Vocabulary List and the Science Word List, the lists were compared and coverage results were obtained. The analyses yielded the coverage values of 32.20 %, 30.8 %, 13.30 % respectively. Expressed another way, only 32.20 percent of the words in the *Science Textbooks Word List* appear in the new GSL; 30.8 percent of the words appear in the Academic Vocabulary List and 13.30 percent of the words appear in the Science Word List. These coverage values lead to the conclusion that none of the word lists, be it general service, academic or science word list, can fully mirror the specific lexis derived from corpus data specific to the science courses taken by the engineering students. It is also apparent that an empirically derived vocabulary list specific to learners' needs will pay substantial dividends, when effective mechanisms are in place to exploit this source for the benefit of the learners.

Teachers' intuitions

In order to combine objective, quantitative corpus data with subjective, qualitative data, and thus develop a pedagogically solid wordlist, teachers' opinions were explored with regard to the usefulness of the items on the list. The items were rated by teachers on a 5-point likert scale, with 5 being the "extremely useful" and 1 "not useful" at all. The mean scores showed that the items were mostly found useful. The items that received a score below 3, which corresponds to "not useful" or "not useful at all" were identified and marked in the wordlist. In terms of the relationship between the teachers' beliefs or perceptions on the usefulness of the items and the corpus frequency data, no significant correlation was found between the variables, which was not surprising considering the mixed results of the previous studies in the literature.

6.3. Pedagogical Implications

A number of pedagogical implications can arise from the study, which adopted a quantitative, low-inference data collection approach consolidated with subjective, expert opinion. First of all, corpus-informed and pedagogically convenient wordlists are valuable assets for both teachers and learners. The lexical choices that material developers and textbook writers can be arbitrary; therefore, development of a wordlist based on authentic usage and integrating it into the curriculum or syllabus is remarkable. Also, every school or group of learners have their own goals and objectives; their purposes of learning English inevitably vary, and thus a bespoke curriculum that is fit-for-purpose would serve learners' needs better. In the Turkish context at tertiary level, students, whose general proficiency level is not sufficient to perform academic studies, receive one-year general English education in the preparatory schools before starting their majors. It is common practice that these learners are taught general English, yet endeavor to perform their studies through academic English, in fact academic English specific to the discipline they are studying in. The understanding that learners in different disciplines have varying needs have also been confirmed by the lecturing staff delivering the science courses in the context of this study, lending further support to the usefulness of a tailored lexical content specific to their needs. Assuming that these students' target needs are covered through the current programme content would be a complacent attitude toward the problem. In this respect, Hyland and Tse (2007) state that "Within each discipline or course, students need to acquire the specialized discourse competencies that will allow them to succeed in their studies and participate as group members" (pp. 248-9). In that respect, having established the freshman engineering students' target lexical needs and developing a corpus-derived, pedagogically convenient wordlist, this study is believed to provide the basis on which a curriculum or a course syllabus can be built.

Frequency, being a psycholinguistic reality, is a pivotal element of any lexical content to be used in language teaching. Regardless of the context the language is taught, it is prudent to consult to corpus data, despite its limitations. Employing a corpus-informed, frequency-based approach, this study provides a valuable tool for the learners and the teachers. From the teachers' perspective, a curriculum or syllabus can be developed around the corpus-derived wordlist, assessment constructs can be devised based on frequency information, and materials can be designed using authentic contexts in which the frequent lexical items occur. For the learners, learning and retrieving items they are supposed to come across frequently can be more plausible. Being exposed to vocabulary in contexts reflecting their further studies can also have a priming effect.

The wordlist constructed in this study has benefitted from human guidance, which contributes to its teachable nature. It is of significant value to consolidate the empirically built wordlist through intuitive expert data. Some items on the wordlist were considered as “not useful” by the teachers, which were marked in the list for further consideration. The corpus-informed wordlist based on quantitative objective data can be fine-grained with the help of expert opinion. As such, the list can serve as a core element which is subject to change and improve throughout the process of programme development and evaluation.

Another point worth considering is that, the corpus compiled for the purposes of this study, can be exploited for DDL (data driven learning). Learners can directly use the corpus for their own learning processes, by using concordances and KWIC (keyword in context) tools. They can explore the specialized corpora and make inferences as to lexical patterns and collocations frequently used in the target context. Teachers can also prepare materials using the contextual elements where the vocabulary is used. It was highlighted by the lecturing staff that it would be useful for students to familiarize with the lexical content through reading passages related to the scientific domain, which would help them get to grips with the material they would need to cover during their studies. As noted by Boulton (2016) “…corpora can be useful in preparing all kinds of pedagogical materials and resources, from general to specialized dictionaries to grammar books and usage manuals, from syllabus design to testing, from wordlists to course books” (p.3).

Finally, although the target wordlist is primarily aimed at the students of engineering faculty, students of science departments - departments of physics, chemistry, biology and mathematics - can also benefit from this list of specialized lexis. The corpus being a compilation of science textbooks, those students who are studying in any scientific discipline or teachers teaching students of a scientific discipline can also make use of the list. Similarly, the corpus data can also be used for authentic ESP material development in science disciplines.

6.4. Limitations of the Study and Suggestions for Further Research

The current study aiming to establish the lexical target needs of engineering students and construct a word list intended as a primary inventory based on corpus frequency measures, has several limitations which should be acknowledged.

The greatest limitation to the study is the low number of participants from whom qualitative data is collected. The study would have sounder results if the scope for subjective data were wider and a higher number of teachers had been asked about their opinion regarding the word list items. The reason for this limitation was the intention of collecting data solely from the subject teachers giving the science courses; academic members of other faculties could as well participate in the questionnaires.

Another point which could be perceived as a limitation is the fact that the corpus data is mainly based on written texts. In order to have a more balanced corpus, spoken content could have been included, which is a laborious and time-consuming task. A sample of lectures from each subject area could have been recorded, transcribed and uploaded for corpus compilation. Despite the fact that the lectures are reported to be based mostly on textbook information, they could have provided insight into the characteristics of the spoken discourse used in the lectures.

Exam tasks were reported to be a source of challenge for the students. Examining the tasks in the courses could have provided strong data regarding the lexical profiles of the assessment content; however, data of this sort was impossible to reach for confidentiality reasons.

It is also possible to extend the scope of the study by including EFL teachers to collect data on the teachability of the wordlist items, assuming that a syllabus would be developed around the word list. Data of this sort would guide the sequencing of the lexical components of a syllabus or curriculum.

Another further extension could be suggesting a sample lesson plan for a course based on corpus data. The wordlist items can be taught using the texts in the corpus, with some modification, in the form of cloze texts and reading passages, as well as writing practice tasks.

Lastly, the *Science Textbooks Word List* could be compared with the target wordlists of other EFL coursebooks widely used in the preparatory programmes at tertiary level, in order to determine the degree of overlap between them.

REFERENCES

Alderson, J. C. (2007). Judging the frequency of English words. *Applied Linguistics*, *28*(3), 383-409. https://doi.org/10.1093/applin/amm024

Almut, K. (2010). Building small specialised corpora. In A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 66-79). Routledge.

Anthony, L. (2019). Tools and strategies for Data-Driven Learning (DDL) in the EAP writing classroom. In , K. Hyland, & L. L Wong (Eds.), *Specialised English: New directions in ESP and EAP research and practice* (pp. 179-194). Routledge.

Baker, P. (2006). *Using corpora in discourse analysis*. A&C Black.

Barber, C. L. (1962). Some measurable characteristics of modern scientific prose. In C. L. Barber (Ed.), *Contributions to English syntax and philology* (pp. 21-43). Acta Universitatis Gothoburgensis.

Basturkmen, H. (2010). *Developing courses in English for specific purposes.* Springer.

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

Becker, H. S. (1970). *Sociological Work: Method and Substance*. Routledge.

Berwick, R. (1989). Needs assessment in language programming: From theory to practice. In R. K. Johnson (Ed.), *The second language curriculum* (pp. 48-62). Cambridge University Press.

Biber, D. (1993). Representativeness in corpus design. *Literary and linguistic computing*, *8*(4), 243-257. https://doi.org/10.1093/llc/8.4.243

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins.

Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching?. *Studies in Second Language Acquisition*, 24(2), 199-208. https://doi.org/10.1017/S0272263102002048

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied linguistics*, *25*(3), 371-405. https://doi.org/10.1093/applin/25.3.371

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

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *The Longman Grammar of Spoken and Written English.* Pearson Education.

Bloomfield, A., Wayland, S. C., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2010). *What makes listening difficult? Factors affecting second language listening comprehension*. University of Maryland.

Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension. *International journal of listening*, 14(1), 14-31. https://doi.org/10.1080/10904018.2000.10499033

Boulton, A. (2012). What data for data-driven learning?. *EuroCALL Review*, *20*(1), 23-27.

Bowker, L., & Pearson, J. (2002). *Working with specialized language: a practical guide to using corpora*. Routledge.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics*, 36(1), 1-22. https://doi.org/10.1093/applin/amt018

Browne, C. (2014). A new general service list: The better mousetrap we've been looking for. *Vocabulary learning and Instruction*, *3*(2), 1-10. doi: http://dx.doi.org/10.7820/vli.v03.2.browne

Brzoza, B. (2018). Word frequency counts: Linking corpus data to user's perception in linguistic research. *Lingvisticæ Investigationes*, 41(2), 224–239. https://doi.org/10.1075/li.00021.brz

Burton, G. (2012). Corpora and coursebooks: Destined to be strangers forever? *Corpora*, 7, 1, 91-108.

Cangır, H. (2021). Objective and subjective collocational frequency: Association strength measures and EFL teacher intuitions. *Pedagogical Linguistics*, *2*(1), 64-91. https://doi.org/10.1075/pl.20014.can

Carter, R. (1998). Orders of reality: CANCODE, communication and culture. *ELT Journal*, 52, 43-56. https://doi.org/10.1093/elt/52.1.43

Carter, R. (2004). *Language and Creativity: The art of common talk*. Routledge.

Carter, R. A. & McCarthy, M. J. (2006) *Cambridge Grammar of English*. Cambridge University Press.

Carver, D. (1983). Some propositions about ESP. *The ESP journal*, *2*(2), 131-137.

Chen, Q., & Ge, G. C. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes*, *26*(4), 502-514. https://doi.org/10.1016/j.esp.2007.04.003

Cheng, W. (2012). *Exploring Corpus Linguistics: Language in action*. Routledge.

Cheng, W., Greaves, C., Sinclair, J. M., & Warren, M. (2009). Uncovering the extent of the phraseological tendency: Towards a systematic analysis of concgrams. *Applied Linguistics*, 30(2), 236-252. https://doi.org/10.1093/applin/amn039

Cheng, W., & Warren, M. (2007). Checking understandings: Comparing textbooks and a corpus of spoken English in Hong Kong. *Language Awareness*, 16(3), 190-207. https://doi.org/10.2167/la455.0

Chen, X., & Dong, Y. (2019). Evaluating objective and subjective frequency measures in L2 lexical processing. *Lingua*, *230*, 102738. https://doi.org/10.1016/j.lingua.2019.102738

Chung, T. M. (2003). *Identifying technical vocabulary*. [Unpublished Ph.D. thesis]. Victoria University of Wellington.

Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System*, *32*(2), 251-263. https://doi.org/10.1016/j.system.2003.11.008

Cobb, T. and Horst, M. (2015). Learner corpora and lexis, in S. Granger, G. Gilquin and F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Linguistics* (pp. 185–206). Cambridge University Press.

Cortes, V. (2002) 'Lexical Bundles in Freshman Composition', in D. Biber, S. Fitzmaurice and R. Reppen (Eds.), *Using Corpora to Explore Linguistic Variation* (pp. 131–45). John Benjamins.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*(4), 397–423. https://doi.org/10.1016/j.esp.2003.12.001

Coxhead, A. (1998). *An academic word list*. School of Linguistics and Applied Language Studies, Victoria University of Wellington.

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

Coxhead, A. (2012). Vocabulary and ESP. In B. Paltridge and S. Starfield (Eds), *The Handbook of English for Specific Purposes* (pp. 115-132).Wiley Blackwell. https://doi.org/10.1002/9781118339855.ch6

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., & Hirsch, D. (2007). A pilot science-specific word list. *Revue Française De Linguistique Appliquée*, *XII*(2), 65-78. https://doi.org/10.3917/rfla.122.0065

Creswell, J. W. (1999). Mixed-method research: Introduction and application. In G. J. Cizek (Ed.), *Handbook of educational policy* (pp. 455-472). Academic press.

Dang, T. N. Y. (2020). Corpus-based word lists in second language vocabulary research, learning, and teaching. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp.288-304). Routledge.

Dang, T. N. Y & Webb, S (2016). Making an essential word list for beginners. In I.S.OP. Nation (Ed.), *Making and Using Word Lists for Language Learning and Testing* (pp. 153-167). John Benjamins. https://doi.org/10.1075/z.208.15ch15

Dang, T. N. Y., Webb, S., & Coxhead, A. (2022). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research*, 26(4), 617-641. https://doi.org/10.1177/1362168820911189

Davies, M., & Gardner, D. (2013). *A frequency dictionary of contemporary American English: Word sketches, collocates and thematic lists*. Routledge.

Dudley-Evans, T. (1998). An Overview of ESP in the 1990s. US Department of Education. https://files.eric.ed.gov/fulltext/ED424775.pdf

Dudley-Evans, T., & St John, M. J. (1998). *Developments in English for specific Purposes: A Multi-Disciplinary Approach.* Cambridge University Press.

Durrant, P. (2014). Discipline and level specificity in university students' written vocabulary. *Applied Linguistics*, 35(3), 328-356. https://doi.org/10.1093/applin/amt016

Durrant, P. (2015) Lexical bundless and disciplinary variation in university students' writing: mapping the theories. *Applied Linguistics*, 38(2), 165-193. https://doi.org/10.1093/applin/amv011

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

Ellis, N. C., Simpson-Vlach, R. I. T. A., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL quarterly*, *42*(3), 375-396. https://doi.org/10.1002/j.1545-7249.2008.tb00137.x

Evans, D. (2007*). Corpus building and investigation for the Humanities: An on-line information pack about corpus investigation techniques for the Humanities*. University of Nottingham.

Evison, J. (2010). What are the basics of analysing a corpus?. In A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 122-135). Routledge.

Fielding N., & Fielding J. (1986). *Linking data. The articulation of qualitative and quantitative methods in social research*. Sage.

Firth, J.R. (1957). *Papers in Linguistics 1934-1951*. Oxford University Press.

Flowerdew, J. (1993). Concordancing as a tool in course design. *System*, *21*(2), 231-244. https://doi.org/10.1016/0346-251X(93)90044-H

Flowerdew, J. (2002). Genre in the classroom: A linguistic approach. In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives*, (pp. 91-102). Lawrence Erlbaum Associates Publishers.

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International journal of corpus linguistics*, 14(3), 393-417. https://doi.org/10.1075/ijcl.14.3.05flo

Flowerdew, L. (2012). Corpus-based discourse analysis, in J. P. Gee and M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 174-187). Routledge.

Francis, N. and Kucera, H. (1982). *Frequency Analysis of English Usage: Lexicon and grammar*. Houghton Mifflin.

Gabrielatos, C. (2005). Corpora and Language Teaching: Just a Fling or Wedding Bells?. *Tesl-Ej*, 8(4). https://files.eric.ed.gov/fulltext/EJ1068106.pdf

Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In C. Taylor & A. Marchi, *Corpus Approaches to Discourse* (pp. 225-258). Routledge.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction*. Longman Publishing.

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

Gatehouse, K. (2001). Key issues in English for specific purposes (ESP) curriculum development. *The internet TESL journal*, *7*(10), 1-10.

Gavioli, L. (2005). *Exploring corpora for ESP learning*. John Benjamins.

Gilmore, A. (2015). Research into practice: The influence of discourse studies on language descriptions and task design in published ELT materials. *Language Teaching*, *48*(4), 506-530. https://doi.org/10.1017/S0261444815000269

Gilner, L. (2011). A primer on the general service list. *Reading in a Foreign Language*, 23(1), 65-83. http://hdl.handle.net/10125/66658

Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6(4), 319-335. https://doi.org/10.1016/j.jeap.2007.09.007

Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes*, *19*(2), 115-135. https://doi.org/10.1016/S0889-4906(98)00015-5

Gómez, P. C. (2013). *Statistical methods in language and linguistic research*. Equinox.

Gouverneur, C. (2008). The phraseological patterns of high-frequency verbs in advanced English for general purposes: A corpus driven approach to EFL textbook analysis. in F. Meunier and S. Granger (Eds.), *Phraseology in Foreign Language Learning and Teaching* (p. 223-246). John Benjamins.

Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, S. Petch-Tyson, & J. Hung (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching*. John Benjamins Publishing Company. http://digital.casalini.it/9789027296238

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, *1*(1), 7-24. https://doi.org/10.1075/ijlcr.1.1.01gra

Greaves, C., & Warren, M. (2010). What can a corpus tell us about multi-word units. In A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 212-226). Routledge.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255–274. https://doi.org/10.3102/01623737011003255

Guetterman, T. C., & Fetters, M. D. (2018). Two methodological approaches to the integration of mixed methods and case study designs: A systematic review. *American Behavioral Scientist*, *62*(7), 900-918. https://doi.org/10.1177/0002764218772641

Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Bazell, J.C. Catford, M. A. K. Halliday and R. H. Robins (Eds.), *In Memory of J. R. Firth* (pp. 148-162). Longman.

Handford, M. (2010). What can a corpus tell us about specialist genres. In A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 255-269). Routledge.

He, X., & Godfroid, A. (2019). Choosing words to teach: A novel method for vocabulary selection and its practical application. *Tesol Quarterly*, 53(2), 348-371. https://doi.org/10.1002/tesq.483

Higgins, J. J. (1966). Hard facts: Notes on teaching English to science students. *ELT Journal*, *21*(1), 55-60. https://doi.org/10.1093/elt/XXI.1.55

Hirsh, D. (2004). *A Functional Representation of Academic Vocabulary*. [Unpublished PhD thesis]. Victoria University of Wellington.

Hoffmann, S., & Lehmann, H. M. (2000). Collocational Evidence from the British National Corpus. In J. M. Kirk (Ed.), *Corpora Galore: Analyses and Techniques in Describing English* (pp. 17-32). Rodopi. https://doi.org/10.1163/9789004485211_005

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

Hu, M. & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403-430.

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

Hutchinson, T., & Waters, A. (1987). *English for specific purposes*. Cambridge University Press.

Hyland, K. (2007). English for specific purposes: Some influences and impacts. In: Cummins, J., Davison, C. (Eds.), *International Handbook of English Language Teaching* (pp. 391-402). Springer.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*, *27*(1), 4-21. https://doi.org/10.1016/j.esp.2007.06.001

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

Inman M. (1978). Lexical analysis of scientific and technical prose. In Trimble, L., Todd-Trimble, M. and Drobnic, K (Eds.), *ESP: Science and Technology* (pp. 242-256). English Language Institute, Oregon State University.

It-ngam, T., & Phoocharoensil, S. (2015). The development of science academic word list. *Indonesian Journal of Applied Linguistics*, *8*(3), 657-667. doi: 10.17509/ijal.v8i3.15269

Iwai, T., Kondo, K., Limm, S. J. D., Ray, E. G., Shimizu, H., & Brown, J. D. (1999). *Japanese language needs analysis.* http://www.nflrc.hawaii.edu/Networks/NW13/NW13.pdf

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In Odlin (Ed.), *Perspectives on Pedagogical Grammar* (pp. 293–313). Cambridge University Press.

Jones M. & Durrant P. (2010). What can a corpus tell us about vocabulary teaching materials? In A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 341-357). Routledge.

Jordan, R. R. (1997). *English for academic purposes: A guide and resource book for teachers*. Cambridge University Press.

Kennedy, G. (1992). Preferred ways of putting things with implications for language teaching. In J. Svartvik (Ed.), *Directions in corpus linguistics* (pp. 335-378). Walter de Gruyter.

Kennedy, G. (2004). *The Contribution of corpus linguistics to language teaching: Three decades of promise*. 25th Icame Conference. Verona.

Koester, A. (2006). *Investigating workplace discourse*. Routledge.

Koester, A. (2010). *Workplace discourse*. A & C Black.

Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research*, *4*(5), 900. doi:10.4304/jltr.4.5.900-913

Konstantakis, N. (2007). Creating a business word list for teaching business English. *Estudios de Lingüística Inglesa Aplicada,* 7, 79-102. http://hdl.handle.net/11441/34157

Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT journal*, 59(4), 322-332. https://doi.org/10.1093/elt/cci061

Krippendorff, K. (2003). *Content analysis: An introduction to its methodology*. Sage.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension. In C. Lauren, M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316-323). Multilingual Matters Ltd.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22 (1), 15-30. http://hdl.handle.net/10125/66648

Lee, D. Y. (2010). What corpora are available?. In A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 107-121). Routledge.

Leech, G. (1991). The state of art in corpus linguistics. In K. Aijmer and B. Altenberg (Eds.), *English Corpus Linguistics*, (pp. 8-29). Longman.

Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, 22, 42-53. doi: 10.1016/j.jeap.2016.01.008

Li, S. L., & Pemberton, R. (1994). An investigation of students' knowledge of academic and subtechnical vocabulary. *Proceedings of the Joint Seminar on Corpus Linguistics and Lexicology* (pp. 183–196). Hong Kong University of Science and Technology. https://hdl.handle.net/1783.1/1089

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

Lynch, T., & Mendelsohn, D. (2013). Listening. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 190-206). Routledge.

Mackey, W. F. (1965). *Language teaching analysis*. Longmans.

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

McCarten, J. (2010). Corpus-informed course book design. In A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 413-427). Routledge.

McCrostie, J. (2007). Investigating the accuracy of teachers' word frequency intuitions. *RELC Journal*, 38(1), 53–66. https://doi.org/10.1177/0033688206076158

McDonough, J. (1984). *ESP in perspective: A practical guide*. Taylor & Francis.

McEnery, T. and Wilson, A. (1996) *Corpus Linguistics*. Edinburgh University Press.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Taylor & Francis.

McGee, I. (2009). Adjective-noun collocations in elicited and corpus data: Similarities, differences and the whys and wherefores. *Corpus Linguistics and Linguistic Theory*, (5), 79–103. https://doi.org/10.1515/CLLT.2009.004

Meara, P. (1992). Network structures and vocabulary acquisition in a foreign language. In P. J, Arnaud &, H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 62-70). Macmillan.

Meunier, F. and Reppen, R. (2015). Corpus versus non-corpus-informed pedagogical materials: Grammar as the focus. In D. Biber and R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics* (pp. 498-514). Cambridge University Press.

Merriam, S. B. (1998). *Qualitative research and case study applications in education* (2nd ed.). Jossey-Bass Publishers.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Multilingual Matters.

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in English as a foreign language. *Insights into non-native vocabulary teaching and learning*, *52*, 83-98.

Moon, R. (1997). Vocabulary connections: Multi-word stems in English. In N. Schmitt and M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 40-63). Cambridge University Press.

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

Munby, J. (1978). *Communicative syllabus design*. Cambridge University Press.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening?. *Canadian modern language review*, *63*(1), 59-82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P. (2011). Research into practice: Vocabulary. *Language Teaching*, *44*(4), 529-539.

Nation, I. S. P. (2013). My ideal vocabulary teaching course. In J., Macalister & I. S. P., Nation (Eds.), *Case studies in language curriculum design* (pp. 61-74). Routledge.

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

Nation, I. S. P., & Hwang, K. (1995). Where would general service vocabulary stop and special purposes vocabulary begin?. *System*, *23*(1), 35-41. https://doi.org/10.1016/0346-251X(94)00050-G

Nation, I. S. P. & Macalister, John (2010). *Language Curriculum Design*. Routledge.

Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge University Press.

Nelson, M. (2006). Semantic associations in Business English: A corpus-based analysis. *English for Specific Purposes*, *25*(2), 217-234.

https://doi.org/10.1016/j.esp.2005.02.008

Nelson, M. (2010). Building a written corpus. In A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 53-65). Routledge.

Nesi, H. (2012). ESP and corpus studies. In B. Paltridge & S. Starfield (Eds.), *The Handbook of English for Specific Purposes* (pp. 407-426). John Wiley & Sons.

Nunan, D. (2004). *Task-based language teaching*. Cambridge University Press.

Nurweni, A., & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes*, 18(2), 161-175. https://doi.org/10.1016/S0889-4906(98)00005-2

O'Keeffe, A., McCarthy, M. and Carter, R. (2007). *From Corpus to Classroom*. Cambridge University Press.

Okamoto, M. (2015). Is corpus word frequency a good yardstick for selecting words to teach? Threshold levels for vocabulary selection. *System*, 51, 1–10. https://doi.org/10.1016/j.system.2015.03.004

Oostdijk, N. (1991). *Corpus linguistics and the automatic analysis of English*. Rodopi.

Osada, N. (2004). Listening comprehension research: A brief review of the past thirty years. *Dialogue*, *3*(1), 53-66.

Paltridge, B. (2002). Thesis and dissertation writing: An examination of published advice and actual practice. *English for Specific Purposes*, *21*(2), 125-143. https://doi.org/10.1016/S0889-4906(00)00025-9

Paltridge, B. (2012). Genre and English for specific purposes. In B. Paltridge & S. Starfield (Eds.), *The Handbook of English for Specific Purposes* (pp. 347-366). John Wiley & Sons.

Partington, A. (1998). *Patterns and Meanings*. John Benjamins.

Rahman, M. (2015). English for Specific Purposes (ESP): A Holistic Review. *Universal Journal of Educational Research*, *3*(1), 24-31.

Rayson, Paul and Roger Garside. (2000). Comparing Corpora Using Frequency Profiling. *NAACL-ANLP 2000 Workshop: Syntactic and Semantic Complexity in Natural Language Processing Systems*. http://acl.ldc.upenn.edu/W/W00/W00-0901.pdf

Richterich R. & Chancerel, J.C. (1980). *Identifying the needs of adults learning a foreign language*. Pergamon Press.

Robinson, P. C. (1991). *ESP today: A practitioner's guide*. Prentice Hall.

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205-225. https://doi.org/10.1017/S0267190511000055

Schmitt, N. (2000) *Vocabulary in Language Teaching*. Cambridge University Press.

Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing, and use*. John Benjamins Publishing.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Springer.

Schmitt, N., & Dunham, B. (1999). Exploring native and non-native intuitions of word frequency. *Second Language Research*, *15*(4), 389-411.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26-43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, *47*(4), 484-503. https://doi.org/10.1017/S0261444812000018

Scott, M. (1997). PC analysis of key words—and key key words. *System*, *25*(2), 233-245. https://doi.org/10.1016/S0346-251X(97)00011-0

Scott, M. and Tribble, C. (2006) *Textual Patterns: Key Words and Corpus Analysis in Language Education.* John Benjamins.

Scrivener, J. (2005). *Learning teaching*. Macmillan.

Seedhouse, P. (1995). Needs analysis and the general English classroom. *English Language Teaching Journal*, 49 (1) 59-65. https://doi.org/10.1093/elt/49.1.59

Sinclair, J. (1987). Collocation: A Progress Report, in R. Steele and T. Threadgold (Eds), *Language Topics: Essays in Honour of Michael Halliday* (pp. 319–31). John Benjamins.

Sinclair, J. (1987). *Lookingup: An account of the COBUILD project in lexical computing*. Collins.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching*, (pp. 140-160). Routledge.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, *64*(3), 429-458. https://doi.org/10.3138/cmlr.64.3.429

Siyanova-Chanturia, A., & Spina, S. (2015). Investigation of Native Speaker and Second Language Learner Intuition of Collocation Frequency. *Language Learning*, 65(3), 533–562. https://doi.org/10.1111/lang.12125

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139-152. https://doi.org/10.1080/09571730802389975

Stein, G. (2017). Some Thoughts on the issue of core vocabularies: A response to Vaclav Brezina and Dana Gablasova:'Is there a core general vocabulary?' Introducing the New General Service List. *Applied Linguistics*, *38*(5), 759-763. https://doi.org/10.1093/applin/amw027

Strevens, P. (1988). The learner and teacher of ESP. In D. Chamberlain & R. J. Baumgardner (Eds.) *ESP in the classroom: Practice and evaluation* (pp. 91-119). Modern English Publications and the British Council.

Stubbs, M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture*. Blackwell publishers.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell publishers.

Stubbs, M. (2007). On texts, corpora and models of language. In M. Hoey, M. Mahlberg, M. Stubbs & W. Teubert (Eds.), *Text, discourse and corpora: Theory and analysis* (pp. 127-161). Continuum.

Sukman, K., Triwatwaranon, W., Munkongdee, T., & Chumnumnawin, N. (2022). A Corpus-Based Study of Lexical Collocations of Keywords Found in Online Business News Articles. *European Journal of English Language Teaching*, *7*(3). http://dx.doi.org/10.46827/ejel.v7i3.4275

Swales, J. M., & Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Sylvén, L. K. (2013). CLIL in Sweden–why does it not work? A metaperspective on CLIL across contexts in Europe. *International Journal of Bilingual Education and Bilingualism*, 16(3), 301-320. https://doi.org/10.1080/13670050.2013.777387

Szudarski, P. (2017). *Corpus linguistics for vocabulary: A guide for research*. Routledge.

Thorndike E. L. (1921). Measurement in education. *Teachers College Record*, 22(5), 371–379.

Tognini-Bonelli, E. (2010). Theoretical overview of the evolution of corpus linguistics, in A. O'Keevffe and M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 14-27). Routledge.

Tremblay, A., Baayen, H., Derwing, B., & Libben, G. (2008, June 19-20). *Lexical bundles and working memory: An ERP study*. [Conference presentation]. Formulaic Language Research Network Conference, University of Nottingham.

Tribble, C. (2002). Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew (Ed.), *Academic Discourse* (pp. 131–149). Longman

Trimble, L. (1985). *English for science and technology: A discourse approach*. Cambridge University Press

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. doi: 10.1016/j.jeap.2013.07.001

Van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension?. *Applied linguistics*, 34(4), 457-479. https://doi.org/10.1093/applin/ams074

Wang, J., Liang, S. L., & Ge, G. C. (2008). Establishment of a medical academic word list. *English for Specific Purposes, 27*(4), 442-458. doi:10.1016/j.esp.2008.05.003

Ward, J. (2007). Collocation and technicality in EAP engineering. *Journal of English for Academic Purposes*, *6*(1), 18-35. https://doi.org/10.1016/j.jeap.2006.10.001

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

Webb, S., & Nation, P. (2013). Computer-assisted vocabulary load analysis. In C. Chappelle (Ed.) *Encyclopedia of Applied Linguistics* (pp. 844–853). Wiley-Blackwell.

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

West, M. (1953). *A General Service List of English Words*. Longman.

Widdowson, H. G. (1998). Context, community, and authentic language. *TESOL quarterly*, *32*(4), 705-716. https://doi.org/10.2307/3588001

Willis, D. (1990). *The lexical syllabus*. Collins.

Willis, D. and Willis, J. (1988). *Collins COBUILD English Course*. Collins COBUILD

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge University Press.

Xodabande, I., & Xodabande, N. (2020). Academic Vocabulary in Psychology Research Articles: A Corpus-Based Study. *MEXTESOL Journal*, *44*(3), 1-21.

Xue, G., & Nation, I. S. (1984). A university word list. *Language learning and communication*, *3*(2), 215-229. https://doi.org/10.2307/3587951

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

Yin, Robert K. (2014). *Case study research: Design and methods.* Sage.

Zorzi, D. (2001). The pedagogic use of spoken corpora: Learning discourse markers in Italian. In G. Aston (Ed.), *Learning with corpora* (pp. 85-107). Athelstan.

APPENDICES

A. INTERVIEW QUESTIONS

1. What are the requirements of the course? E.g. full comprehension of the written materials, participating in discussions, understanding lectures, delivering presentations, writing academic articles etc.

2. What linguistic skills and sub-skills do students need to fulfil the requirements of the course? E.g. Reading, Writing, Listening, Speaking, Grammar, Vocabulary

2.1. How do you rate the significance of these skills from 1 to 4 for the fulfilment of the requirements? (1-Very significant, 2-Significant, 3-Somewhat significant, 4-Not significant?

3. What do you think students mostly need for sufficient coverage of a) written materials? b) lectures?

4. What do the students mostly have difficulty in during their studies?

5. To what extent do the students that have completed the preparatory programme possess the necessary skills to be successful in the course?

5.1. In which areas do you think they need improvement?

6. Do you think the materials covered in the science course feature a specific lexis?

7. Do you think engineering students' lexical needs differ from those of students from other disciplines?

8. Do you think engineering students would benefit more from a language programme specifically designed for them?

8.1. If yes, what features can characterise such a programme?
8.2 If not, why do you think such a programme is not necessary?

B. FREQUENCY LIST

	Item	Freq.	Relative Frequency	Average Reduced Frequency	Rel. DOCF
1.	show	4082	1772.40	2149.40	100
2.	find	4459	1936.09	2048.18	100
3.	example	3650	1584.82	1904.98	100
4.	give	3358	1458.04	1751.37	100
5.	point	4730	2053.76	1647.81	100
6.	equation	4524	1964.31	1547.46	100
7.	form	3177	1379.45	1485.96	100
8.	value	3925	1704.23	1455.36	100
9.	see	2498	1084.63	1401.08	100
10.	energy	5630	2444.54	1385.48	100
11.	time	3395	1474.10	1382.68	100
12.	change	3293	1429.81	1252.28	100
13.	result	2170	942.21	1202.46	100
14.	call	2409	1045.98	1200.78	100
15.	make	2082	904.00	1161.09	100
16.	water	4074	1768.92	1155.21	100
17.	produce	2639	1145.85	1140.44	100
18.	small	2262	982.16	1130.26	100
19.	number	2896	1257.44	1057.49	100
20.	function	3888	1688.16	1056.59	100
21.	move	2666	1157.57	1027.44	100
22.	many	2070	898.79	1023.05	100
23.	increase	2469	1072.04	1017.93	100
24.	follow	1957	849.73	979.89	100
25.	would	2058	893.58	972.75	100
26.	constant	2514	1091.57	955.76	100
27.	large	1951	847.12	946.57	100
28.	most	2010	872.74	919.17	100
29.	system	3056	1326.91	907.43	100
30.	cell	5311	2306.03	879.47	100
31.	determine	1876	814.56	874.28	100
32.	describe	1646	714.69	864.25	100
33.	mass	3384	1469.33	859.35	100
34.	force	4023	1746.78	859.18	100

35.	occur	1834	796.32	852.65	100
36.	solution	3048	1323.44	830.47	100
37.	high	1682	730.32	765.29	100
38.	different	1608	698.19	753.25	100
39.	contain	1583	687.34	752.13	100
40.	line	2327	1010.38	749.05	100
41.	molecule	3143	1364.68	741.08	100
42.	unit	1682	730.32	728.98	100
43.	take	1370	594.85	728.56	100
44.	surface	2469	1072.04	726.46	100
45.	section	1381	599.63	723.97	100
46.	part	1434	622.64	709.27	100
47.	consider	1315	570.97	709.20	100
48.	know	1326	575.75	700.28	100
49.	cause	1696	736.40	683.12	100
50.	atom	3648	1583.95	674.09	100
51.	equal	1462	634.80	673.78	100
52.	reaction	3807	1652.99	662.37	100
53.	case	1211	525.81	660.46	100
54.	way	1162	504.54	656.32	100
55.	speed	2748	1193.18	635.21	100
56.	low	1307	567.50	606.04	100
57.	require	1132	491.51	583.11	100
58.	assume	1279	555.34	580.79	100
59.	after	1206	523.64	579.98	100
60.	base	1631	708.18	576.56	100
61.	direction	1884	818.03	575.33	100
62.	obtain	1227	532.76	573.32	100
63.	process	1586	688.64	568.71	100
64.	calculate	1349	585.73	566.23	100
65.	work	1614	700.80	560.57	100
66.	great	1067	463.29	545.07	100
67.	include	1224	531.46	537.23	100
68.	type	1238	537.54	536.21	100
69.	temperature	2442	1060.31	530.37	100
70.	object	2401	1042.51	515.09	100
71.	length	1639	711.65	513.61	100
72.	represent	1153	500.63	512.27	100
73.	long	1066	462.86	511.02	100
74.	distance	1448	628.72	504.56	100
75.	explain	1177	511.05	503.82	100
76.	become	1038	450.70	502.30	100
77.	right	1101	478.05	501.73	100

78.	apply	1082	469.80	500.16	100
79.	table	1278	554.91	496.42	100
80.	place	966	419.44	496.03	100
81.	important	942	409.01	495.06	100
82.	rate	1772	769.40	492.17	100
83.	charge	3007	1305.63	488.94	100
84.	gas	2322	1008.21	486.20	100
85.	light	2190	950.89	486.00	100
86.	term	1059	459.82	485.84	100
87.	area	1530	664.32	485.27	100
88.	state	1321	573.58	479.53	100
89.	remain	907	393.82	478.61	100
90.	provide	952	413.36	475.34	100
91.	end	1205	523.21	474.92	100
92.	total	1194	518.43	472.82	100
93.	electron	2878	1249.62	467.03	100
94.	measure	1117	485.00	466.96	100
95.	zero	1291	560.55	465.86	100
96.	chapter	1040	451.57	455.59	100
97.	need	827	359.08	454.25	100
98.	problem	1086	471.54	444.79	100
99.	positive	1196	519.30	442.45	100
100.	answer	929	403.37	441.77	100
101.	leave	875	379.92	441.69	100
102.	body	1690	733.79	441.64	100
103.	structure	1713	743.78	437.96	100
104.	difference	1053	457.21	434.64	100
105.	let	874	379.49	431.48	100
106.	above	949	412.05	431.15	100
107.	field	2907	1262.21	431.10	100
108.	depend	794	344.75	424.10	100
109.	less	874	379.49	423.69	100
110.	just	742	322.18	421.48	100
111.	amount	1030	447.22	412.67	100
112.	earth	1462	634.80	412.15	100
113.	side	1001	434.63	411.25	100
114.	particle	1939	841.91	410.94	100
115.	air	1345	584.00	405.04	100
116.	position	1245	540.58	404.75	100
117.	note	854	370.81	404.27	100
118.	single	806	349.96	403.68	100
119.	negative	1060	460.25	401.48	100
120.	new	916	397.73	400.39	100

121.	mean	759	329.56	395.90	100
122.	carry	970	421.17	387.91	100
123.	define	947	411.19	386.62	100
124.	condition	879	381.66	381.86	100
125.	reach	816	354.31	381.20	100
126.	allow	814	353.44	378.85	100
127.	consist	839	364.29	378.65	100
128.	chemical	1183	513.66	378.03	100
129.	curve	1725	748.99	375.21	100
130.	decrease	888	385.57	374.81	100
131.	region	1309	568.37	374.71	100
132.	pass	809	351.27	373.43	100
133.	write	960	416.83	371.69	100
134.	group	1684	731.19	370.62	100
135.	like	708	307.41	370.29	100
136.	although	680	295.25	369.53	100
137.	involve	787	341.71	367.46	100
138.	simple	819	355.61	364.28	100
139.	volume	1547	671.70	364.05	100
140.	expression	943	409.45	363.75	100
141.	therefore	733	318.27	362.69	100
142.	possible	696	302.20	362.51	100
143.	set	717	311.32	362.32	100
144.	center	1182	513.22	362.13	100
145.	graph	1721	747.26	361.98	100
146.	acid	2049	889.67	359.70	100
147.	magnitude	1228	533.20	359.50	100
148.	motion	1353	587.47	358.61	100
149.	could	676	293.52	355.73	100
150.	product	1111	482.39	351.79	100
151.	similar	653	283.53	347.96	100
152.	law	1246	541.01	346.56	100
153.	reduce	837	363.42	345.24	100
154.	well	610	264.86	344.51	100
155.	ion	1867	810.65	343.75	100
156.	below	739	320.87	343.72	100
157.	effect	756	328.25	338.12	100
158.	suppose	703	305.24	337.39	100
159.	method	916	397.73	335.43	100
160.	together	660	286.57	335.43	100
161.	here	641	278.32	334.58	100
162.	pressure	1853	804.57	334.30	100
163.	several	612	265.73	332.43	100

164.	start	717	311.32	330.61	100
165.	add	872	378.62	329.82	100
166.	order	828	359.52	324.87	100
167.	maximum	1027	445.92	323.53	100
168.	compare	606	263.12	322.21	100
169.	plant	1842	799.79	321.98	100
170.	radius	1270	551.43	319.29	100
171.	begin	598	259.65	317.32	100
172.	close	721	313.06	316.67	100
173.	present	688	298.73	314.99	100
174.	lead	699	303.50	313.68	100
175.	release	967	419.87	313.64	100
176.	might	594	257.91	312.06	100
177.	potential	1676	727.72	311.83	100
178.	interval	1209	524.95	309.37	100
179.	act	866	376.02	308.59	100
180.	quantity	769	333.90	306.72	100
181.	angle	1275	553.60	305.51	100
182.	general	590	256.18	304.58	100
183.	help	685	297.43	303.90	100
184.	level	902	391.65	303.89	100
185.	always	550	238.81	302.69	100
186.	while	566	245.76	301.67	100
187.	initial	941	408.58	300.33	100
188.	limit	1072	465.46	298.60	100
189.	plane	1219	529.29	298.20	100
190.	source	839	364.29	297.54	100
191.	average	928	402.94	295.20	100
192.	common	609	264.43	293.55	100
193.	go	532	230.99	293.55	100
194.	current	1951	847.12	293.53	100
195.	study	572	248.36	292.82	100
196.	year	1014	440.28	289.64	100
197.	illustrate	535	232.30	287.83	100
198.	element	1412	613.09	287.10	100
199.	density	1164	505.41	285.11	100
200.	approach	737	320.00	284.47	100
201.	bond	2112	917.03	283.49	75
202.	far	541	234.90	283.34	100
203.	solid	1105	479.79	280.78	100
204.	factor	721	313.06	280.75	100
205.	turn	622	270.07	280.15	100
206.	space	751	326.08	278.72	100

207.	travel	862	374.28	277.79	100
208.	electric	1720	746.82	277.01	100
209.	velocity	1469	637.84	277.00	75
210.	animal	1245	540.58	276.82	100
211.	say	607	263.56	276.79	100
212.	human	1058	459.38	276.19	100
213.	good	543	235.77	275.59	100
214.	material	780	338.67	275.05	100
215.	component	897	389.48	274.49	100
216.	certain	514	223.18	274.46	100
217.	property	747	324.35	271.23	100
218.	heat	1478	641.74	271.23	100
219.	indicate	564	244.89	270.98	100
220.	appear	558	242.28	269.94	100
221.	next	483	209.72	269.81	100
222.	come	494	214.49	267.08	100
223.	discuss	478	207.55	262.22	100
224.	express	613	266.16	260.91	100
225.	fact	472	204.94	260.20	100
226.	relate	555	240.98	258.32	100
227.	power	929	403.37	258.07	100
228.	fall	635	275.72	256.03	100
229.	rest	740	321.31	251.31	100
230.	axis	1011	438.97	251.07	100
231.	every	505	219.27	250.90	100
232.	keep	439	190.61	250.58	100
233.	substance	814	353.44	249.16	100
234.	develop	611	265.30	248.93	100
235.	page	563	244.45	248.35	100
236.	shape	585	254.01	247.99	100
237.	expect	568	246.62	247.74	100
238.	particular	460	199.73	247.32	100
239.	situation	556	241.41	244.94	100
240.	pair	814	353.44	244.24	100
241.	metal	1150	499.33	243.77	100
242.	hydrogen	987	428.55	241.13	75
243.	range	542	235.34	238.05	100
244.	hold	467	202.77	237.53	100
245.	separate	531	230.56	236.93	100
246.	size	569	247.06	236.75	100
247.	life	892	387.30	236.24	100
248.	near	503	218.40	232.45	100
249.	carbon	1131	491.08	231.59	100

250.	still	392	170.21	231.12	100
251.	liquid	939	407.71	230.85	100
252.	inside	567	246.19	230.64	100
253.	step	712	309.15	229.98	100
254.	relative	534	231.86	228.52	100
255.	enter	575	249.66	228.30	100
256.	sum	777	337.37	228.25	100
257.	solve	556	241.41	227.66	100
258.	across	701	304.37	227.44	100
259.	natural	697	302.64	226.94	100
260.	relationship	477	207.11	226.57	100
261.	formula	995	432.03	226.56	100
262.	equilibrium	1182	513.22	225.48	100
263.	compound	1157	502.37	224.08	100
264.	lie	570	247.49	223.83	100
265.	think	388	168.47	223.15	100
266.	vary	462	200.60	221.51	100
267.	sometimes	389	168.90	221.40	100
268.	left	498	216.23	220.68	100
269.	datum	626	271.81	219.76	100
270.	horizontal	764	331.73	218.88	100
271.	diagram	612	265.73	217.14	100
272.	oxygen	781	339.11	216.99	100
273.	instead	367	159.35	216.91	100
274.	blood	1345	584.00	216.36	100
275.	estimate	617	267.90	215.78	100
276.	exist	511	221.88	215.20	100
277.	normal	689	299.16	215.18	100
278.	model	578	250.97	214.68	100
279.	direct	521	226.22	214.39	100
280.	convert	547	237.51	213.72	100
281.	draw	485	210.59	211.79	100
282.	individual	561	243.59	210.57	100
283.	half	419	181.93	208.36	100
284.	suggest	412	178.89	208.26	100
285.	vector	1538	667.80	208.18	100
286.	exercise	731	317.40	208.02	100
287.	differ	407	176.72	206.84	100
288.	vertical	613	266.16	205.22	100
289.	specific	606	263.12	204.78	100
290.	top	535	232.30	203.52	100
291.	wave	1915	831.49	203.10	100
292.	whether	426	184.97	202.42	100

293.	people	687	298.29	201.54	100
294.	directly	371	161.09	201.01	100
295.	path	786	341.28	200.89	100
296.	enough	390	169.34	200.81	100
297.	identify	450	195.39	200.80	100
298.	divide	428	185.84	200.27	100
299.	net	651	282.66	199.79	100
300.	car	822	356.91	199.51	100
301.	locate	489	212.32	199.15	100
302.	among	448	194.52	198.62	100
303.	replace	400	173.68	198.54	100
304.	look	375	162.82	198.00	100
305.	open	567	246.19	197.89	100
306.	evaluate	483	209.72	196.37	100
307.	reason	404	175.42	196.04	100
308.	connect	607	263.56	194.82	100
309.	short	373	161.96	194.53	100
310.	happen	375	162.82	193.80	100
311.	transfer	680	295.25	193.18	100
312.	wall	635	275.72	192.52	100
313.	food	776	336.94	192.40	100
314.	various	335	145.46	192.01	100
315.	break	492	213.63	191.78	100
316.	color	687	298.29	191.64	100
317.	complete	381	165.43	190.82	100
318.	information	449	194.95	190.48	100
319.	refer	388	168.47	190.25	100
320.	acceleration	1183	513.66	189.71	100
321.	final	550	238.81	189.63	100
322.	series	894	388.17	189.37	100
323.	behavior	516	224.05	189.02	100
324.	strong	604	262.26	188.56	100
325.	again	347	150.67	187.36	100
326.	combine	383	166.30	186.76	100
327.	observe	424	184.10	186.64	100
328.	calculation	517	224.48	186.57	100
329.	concept	428	185.84	186.56	100
330.	name	579	251.40	186.24	100
331.	protein	1184	514.09	185.50	75
332.	remove	384	166.73	183.74	100
333.	origin	577	250.53	182.49	100
334.	rule	663	287.87	182.01	100
335.	free	418	181.49	181.96	100

336.	original	381	165.43	181.25	100
337.	flow	615	267.03	180.87	100
338.	double	587	254.87	179.81	100
339.	accord	352	152.84	178.75	100
340.	coordinate	702	304.81	178.62	100
341.	useful	333	144.59	178.43	100
342.	parallel	610	264.86	178.25	100
343.	experiment	491	213.19	177.65	100
344.	height	581	252.27	177.40	100
345.	able	325	141.11	174.87	100
346.	attach	474	205.81	174.70	100
347.	question	323	140.25	174.65	100
348.	period	537	233.16	173.98	100
349.	addition	383	166.30	172.84	100
350.	derive	339	147.19	172.52	100
351.	opposite	447	194.09	172.48	100
352.	concentration	809	351.27	170.69	100
353.	rise	375	162.82	170.52	100
354.	blue	403	174.98	170.21	100
355.	sample	672	291.78	169.91	100
356.	associate	349	151.54	168.79	100
357.	sign	472	204.94	168.74	100
358.	drop	407	176.72	167.96	100
359.	complex	593	257.48	167.68	100
360.	principle	363	157.61	166.83	100
361.	third	317	137.64	166.63	100
362.	notice	319	138.51	166.17	100
363.	exert	684	296.99	165.91	100
364.	square	427	185.40	165.75	100
365.	sphere	797	346.06	164.88	100
366.	get	343	148.93	164.84	100
367.	matter	400	173.68	164.65	100
368.	ratio	447	194.09	163.71	100
369.	outside	371	161.09	162.51	100
370.	molecular	700	303.94	162.28	100
371.	straight	398	172.81	162.12	100
372.	ground	494	214.49	161.74	100
373.	combination	388	168.47	161.31	100
374.	variable	590	256.18	160.91	100
375.	throughout	323	140.25	160.11	100
376.	run	357	155.01	159.71	100
377.	store	550	238.81	159.29	100
378.	true	321	139.38	159.21	100

379.	support	460	199.73	158.88	100
380.	day	415	180.19	158.80	100
381.	focus	419	181.93	158.59	100
382.	generate	423	183.67	158.37	100
383.	likely	358	155.44	158.34	100
384.	continue	303	131.56	157.01	100
385.	physical	322	139.81	156.60	100
386.	maintain	407	176.72	156.49	100
387.	internal	568	246.62	156.35	100
388.	major	418	181.49	155.63	100
389.	idea	348	151.10	155.47	100
390.	lose	349	151.54	155.24	100
391.	active	393	170.64	155.00	100
392.	choose	330	143.29	154.91	100
393.	location	335	145.46	154.76	100
394.	view	340	147.63	154.69	100
395.	control	499	216.66	153.58	100
396.	block	801	347.79	153.40	100
397.	appropriate	300	130.26	152.80	100
398.	additional	274	118.97	152.69	100
399.	bottom	370	160.65	152.53	100
400.	organism	821	356.48	152.17	100
401.	theory	568	246.62	151.45	100
402.	quiz	458	198.86	151.24	50
403.	further	265	115.06	151.01	100
404.	almost	330	143.29	150.92	100
405.	thin	448	194.52	150.59	100
406.	definition	425	184.53	150.58	100
407.	species	652	283.10	150.50	75
408.	face	348	151.10	149.72	100
409.	perpendicular	555	240.98	149.02	100
410.	population	1151	499.76	148.51	100
411.	establish	355	154.14	147.84	100
412.	since	358	155.44	147.83	100
413.	circle	559	242.72	147.49	100
414.	introduce	281	122.01	147.44	100
415.	wire	943	409.45	147.15	100
416.	generally	290	125.92	146.44	100
417.	list	345	149.80	146.28	100
418.	surround	391	169.77	145.99	100
419.	predict	327	141.98	145.89	100
420.	rapidly	318	138.08	145.68	100
421.	perform	306	132.86	144.85	100

422.	create	316	137.21	144.71	100
423.	portion	310	134.60	144.18	100
424.	frequency	1046	454.17	143.91	100
425.	identical	311	135.04	143.82	100
426.	live	501	217.53	143.64	100
427.	grow	533	231.43	143.25	100
428.	learn	296	128.52	142.88	100
429.	multiple	348	151.10	142.85	100
430.	cycle	672	291.78	142.80	100
431.	resistance	723	313.93	142.77	100
432.	word	279	121.14	142.36	100
433.	central	457	198.43	141.06	100
434.	activity	474	205.81	139.64	100
435.	fill	376	163.26	139.45	100
436.	feature	301	130.69	139.42	100
437.	understand	255	110.72	139.17	100
438.	typical	266	115.50	139.08	100
439.	against	330	143.29	139.06	100
440.	formation	455	197.56	138.73	100
441.	least	296	128.52	138.67	100
442.	test	406	176.28	138.61	100
443.	nearly	261	113.33	138.49	100
444.	respect	384	166.73	138.09	100
445.	ball	792	343.88	137.12	100
446.	kinetic	737	320.00	137.01	100
447.	variety	276	119.84	136.82	100
448.	hence	315	136.77	136.31	100
449.	weight	473	205.38	135.60	100
450.	relatively	271	117.67	135.35	100
451.	quick	453	196.69	135.34	100
452.	application	291	126.35	134.43	100
453.	analyze	284	123.31	134.39	100
454.	undergo	314	136.34	134.17	100
455.	basic	326	141.55	133.64	100
456.	finally	232	100.73	133.21	100
457.	hand	304	132.00	132.72	100
458.	circular	447	194.09	132.65	100
459.	continuous	559	242.72	131.53	100
460.	later	250	108.55	131.50	100
461.	nucleus	715	310.45	131.48	100
462.	substitute	300	130.26	130.60	100
463.	approximately	260	112.89	130.24	100
464.	easily	233	101.17	130.00	100

465.	arise	290	125.92	129.78	100
466.	early	368	159.78	129.52	100
467.	atomic	687	298.29	129.10	75
468.	person	356	154.57	128.87	100
469.	plot	340	147.63	128.76	100
470.	uniform	522	226.65	128.40	100
471.	stop	324	140.68	128.29	100
472.	slightly	271	117.67	128.16	100
473.	integral	1004	435.93	128.00	100
474.	dna	1200	521.04	127.95	50
475.	simply	236	102.47	127.50	100
476.	environment	428	185.84	127.29	100
477.	nature	235	102.04	127.13	100
478.	spring	700	303.94	126.18	100
479.	mol	537	233.16	126.06	75
480.	contact	320	138.94	125.31	100
481.	treat	269	116.80	125.24	100
482.	key	246	106.81	124.92	100
483.	minimum	387	168.03	124.87	100
484.	meter	325	141.11	124.69	100
485.	completely	248	107.68	124.69	100
486.	slow	301	130.69	124.57	100
487.	though	229	99.43	124.46	100
488.	cover	301	130.69	123.92	100
489.	experience	295	128.09	123.91	100
490.	gene	1105	479.79	123.87	50
491.	derivative	864	375.15	123.83	100
492.	fluid	690	299.60	123.64	100
493.	itself	213	92.48	123.58	100
494.	special	230	99.87	123.52	100
495.	affect	260	112.89	123.32	100
496.	correct	285	123.75	123.06	100
497.	available	267	115.93	122.59	100
498.	magnetic	1422	617.43	122.26	100
499.	correspond	276	119.84	121.83	100
500.	cylinder	530	230.13	121.78	100
501.	million	386	167.60	121.56	100
502.	entire	227	98.56	121.32	100
503.	imagine	245	106.38	121.26	100
504.	site	318	138.08	121.17	100
505.	extend	251	108.98	120.83	100
506.	tell	230	99.87	120.60	100
507.	upper	295	128.09	120.23	100

508.	presence	255	110.72	119.98	100
509.	phase	572	248.36	119.67	100
510.	discussion	218	94.66	119.55	100
511.	muscle	800	347.36	118.77	75
512.	mixture	506	219.70	118.70	100
513.	significant	303	131.56	118.61	100
514.	proportional	325	141.11	118.18	100
515.	practice	202	87.71	117.61	100
516.	necessary	212	92.05	117.10	100
517.	action	378	164.13	117.04	100
518.	typically	262	113.76	116.87	100
519.	supply	287	124.61	116.50	100
520.	tend	240	104.21	116.48	100
521.	own	246	106.81	116.25	100
522.	return	265	115.06	116.08	100
523.	choice	253	109.85	115.72	100
524.	real	373	161.96	115.60	100
525.	equivalent	284	123.31	115.56	100
526.	absorb	407	176.72	114.47	100
527.	characteristic	239	103.77	114.31	100
528.	plate	748	324.78	114.26	100
529.	play	267	115.93	113.85	100
530.	layer	405	175.85	113.72	100
531.	role	258	112.02	113.41	100
532.	reverse	273	118.54	112.94	100
533.	cross	338	146.76	112.86	100
534.	want	215	93.35	112.59	100
535.	conclude	222	96.39	112.58	100
536.	prevent	290	125.92	112.31	100
537.	linear	372	161.52	112.24	100
538.	corresponding	234	101.60	111.87	100
539.	try	206	89.44	111.74	100
540.	membrane	780	338.67	111.48	100
541.	piece	261	113.33	110.95	100
542.	production	332	144.15	109.80	100
543.	prove	231	100.30	109.69	100
544.	analysis	245	106.38	109.40	100
545.	balance	331	143.72	108.81	100
546.	statement	237	102.90	108.73	100
547.	contribute	250	108.55	108.61	100
548.	standard	467	202.77	108.19	100
549.	loss	265	115.06	107.84	100
550.	late	200	86.84	107.74	100

551.	diameter	366	158.92	107.65	100
552.	growth	556	241.41	107.58	100
553.	external	352	152.84	107.58	100
554.	ability	258	112.02	107.28	100
555.	pattern	437	189.74	107.17	100
556.	construct	217	94.22	106.99	100
557.	tube	422	183.23	106.98	100
558.	apart	227	98.56	106.66	100
559.	actually	178	77.29	106.14	100
560.	seem	219	95.09	106.00	100
561.	main	192	83.37	105.90	100
562.	initially	237	102.90	105.84	100
563.	effective	230	99.87	104.60	100
564.	edge	275	119.40	103.91	100
565.	drive	240	104.21	103.86	100
566.	bacterium	656	284.83	103.86	75
567.	review	210	91.18	103.84	100
568.	dissolve	414	179.76	103.80	100
569.	observation	230	99.87	103.79	100
570.	last	194	84.23	103.44	100
571.	shell	478	207.55	103.42	100
572.	green	309	134.17	103.37	100
573.	upward	314	136.34	102.99	100
574.	electrical	301	130.69	102.76	100
575.	sketch	313	135.90	102.49	100
576.	cut	250	108.55	102.47	100
577.	link	240	104.21	102.33	100
578.	eat	323	140.25	102.22	100
579.	reflect	453	196.69	102.07	100
580.	distribution	362	157.18	101.75	100
581.	scale	319	138.51	101.48	100
582.	outer	279	121.14	101.41	100
583.	she	302	131.13	101.38	100
584.	fast	243	105.51	101.28	100
585.	little	215	93.35	101.20	100
586.	device	271	117.67	100.98	100
587.	raise	237	102.90	100.31	100
588.	ring	412	178.89	99.84	100
589.	world	235	102.04	99.65	100
590.	root	590	256.18	99.45	100
591.	bring	202	87.71	99.35	100
592.	rotate	462	200.60	99.14	100
593.	chain	443	192.35	99.08	100

594.	rod	535	232.30	98.65	100
595.	enzyme	589	255.74	98.56	75
596.	symbol	248	107.68	98.29	100
597.	yield	270	117.23	98.21	100
598.	independent	250	108.55	98.07	100
599.	explore	160	69.47	97.83	100
600.	join	228	99.00	97.66	100
601.	discover	228	99.00	97.37	100
602.	stand	207	89.88	97.02	100
603.	exactly	208	90.31	96.76	100
604.	multiply	214	92.92	96.75	100
605.	development	373	161.96	96.59	100
606.	angular	699	303.50	96.57	75
607.	moment	534	231.86	96.56	100
608.	wide	172	74.68	96.49	100
609.	fraction	283	122.88	96.48	100
610.	highly	210	91.18	96.36	100
611.	theorem	631	273.98	96.16	75
612.	twice	184	79.89	96.09	100
613.	approximate	259	112.46	95.75	100
614.	white	298	129.39	95.19	100
615.	sure	185	80.33	95.09	100
616.	age	266	115.50	94.97	100
617.	encounter	181	78.59	94.95	100
618.	overall	280	121.58	94.72	100
619.	pre	196	85.10	94.69	100
620.	ture	205	89.01	94.47	100
621.	except	167	72.51	94.22	100
622.	respectively	188	81.63	94.20	100
623.	land	317	137.64	93.53	100
624.	image	685	297.43	93.26	100
625.	circuit	901	391.21	93.26	100
626.	recall	178	77.29	93.22	100
627.	account	198	85.97	92.52	100
628.	iron	398	172.81	92.04	100
629.	slope	461	200.17	91.73	100
630.	response	398	172.81	91.63	100
631.	sound	683	296.56	91.52	100
632.	atmosphere	304	132.00	91.44	100
633.	probably	232	100.73	91.33	100
634.	easy	158	68.60	91.04	100
635.	ideal	431	187.14	90.92	100
636.	technique	178	77.29	90.90	100

637.	six	212	92.05	90.83	100
638.	egg	567	246.19	90.82	100
639.	text	191	82.93	89.41	100
640.	mechanism	348	151.10	89.38	75
641.	difficult	163	70.77	89.30	100
642.	degree	222	96.39	89.23	100
643.	design	186	80.76	89.22	100
644.	slowly	182	79.02	88.95	100
645.	tissue	531	230.56	88.54	100
646.	picture	177	76.85	88.54	100
647.	tree	456	197.99	88.41	100
648.	sodium	368	159.78	88.37	75
649.	orbital	1123	487.60	88.02	100
650.	nitrogen	368	159.78	88.02	100
651.	segment	342	148.50	88.00	100
652.	modern	234	101.60	87.83	100
653.	measurement	218	94.66	87.29	100
654.	fish	458	198.86	87.21	100
655.	sequence	430	186.71	87.18	100
656.	displacement	419	181.93	86.95	75
657.	interaction	268	116.37	86.94	100
658.	event	267	115.93	86.67	100
659.	simplify	180	78.16	86.55	100
660.	procedure	173	75.12	86.41	100
661.	eventually	183	79.46	86.18	100
662.	scientist	245	106.38	86.15	100
663.	weak	399	173.25	85.89	100
664.	consistent	174	75.55	85.84	100
665.	hot	255	110.72	85.78	100
666.	clear	169	73.38	85.69	100
667.	evolve	340	147.63	85.52	75
668.	phenomenon	192	83.37	85.44	100
669.	meet	160	69.47	85.41	100
670.	bind	369	160.22	85.31	100
671.	disease	443	192.35	84.96	75
672.	organic	370	160.65	84.87	75
673.	ice	370	160.65	84.66	100
674.	expand	222	96.39	84.49	100
675.	extremely	192	83.37	84.41	100
676.	nutrient	475	206.24	84.36	75
677.	book	203	88.14	84.22	100
678.	basis	172	74.68	83.81	100
679.	commonly	159	69.04	83.74	100

680.	sun	339	147.19	83.68	100
681.	demonstrate	154	66.87	83.67	100
682.	ray	619	268.77	83.64	100
683.	label	201	87.27	83.45	100
684.	serve	168	72.95	83.43	100
685.	repeat	171	74.25	83.40	100
686.	salt	331	143.72	83.35	100
687.	especially	199	86.41	83.34	100
688.	foot	238	103.34	83.15	100
689.	neither	157	68.17	83.08	100
690.	examine	138	59.92	82.41	100
691.	front	272	118.10	82.09	100
692.	string	689	299.16	81.98	100
693.	advantage	181	78.59	81.93	100
694.	gravitational	440	191.05	81.62	100
695.	influence	196	85.10	81.57	100
696.	mechanical	261	113.33	81.45	100
697.	strike	236	102.47	81.40	100
698.	actual	156	67.73	81.30	100
699.	bear	238	103.34	81.29	100
700.	already	148	64.26	81.21	100
701.	newton	364	158.05	81.09	75
702.	head	265	115.06	81.09	100
703.	bar	357	155.01	80.68	100
704.	stable	272	118.10	80.62	100
705.	perhaps	156	67.73	80.47	100
706.	movement	282	122.44	80.29	100
707.	versus	213	92.48	80.16	100
708.	glass	330	143.29	79.98	100
709.	fuel	333	144.59	79.97	100
710.	member	293	127.22	79.91	100
711.	remember	145	62.96	79.81	100
712.	build	148	64.26	79.79	100
713.	sea	311	135.04	79.57	100
714.	signal	338	146.76	79.50	100
715.	pull	243	105.51	79.48	100
716.	recognize	171	74.25	79.42	100
717.	cool	269	116.80	79.40	100
718.	united	225	97.69	79.39	100
719.	transport	346	150.23	79.22	100
720.	downward	230	99.87	79.17	100
721.	rock	306	132.86	79.11	100
722.	evidence	216	93.79	79.11	100

723.	radiation	426	184.97	78.91	100
724.	principal	233	101.17	78.74	100
725.	detect	233	101.17	78.56	100
726.	fundamental	214	92.92	78.39	100
727.	arrow	229	99.43	78.15	100
728.	room	209	90.75	78.13	100
729.	variation	186	80.76	78.04	100
730.	receive	233	101.17	77.85	100
731.	pure	300	130.26	77.79	75
732.	ocean	309	134.17	77.68	100
733.	approximation	295	128.09	77.64	75
734.	summarize	153	66.43	77.63	100
735.	excess	225	97.69	77.50	100
736.	instant	259	112.46	77.46	75
737.	states	215	93.35	77.39	100
738.	instance	142	61.66	77.32	100
739.	tangent	453	196.69	77.15	75
740.	dimension	220	95.52	76.96	100
741.	unknown	202	87.71	76.75	100
742.	smooth	285	123.75	76.75	75
743.	sense	171	74.25	76.60	100
744.	assumption	189	82.06	76.43	100
745.	experimental	165	71.64	76.28	100
746.	closed	280	121.58	76.12	100
747.	similarly	136	59.05	76.07	100
748.	beam	450	195.39	75.76	100
749.	flat	212	92.05	75.57	100
750.	longer	142	61.66	75.19	100
751.	visible	183	79.46	75.17	100
752.	spread	186	80.76	75.14	100
753.	agent	304	132.00	75.10	75
754.	eye	321	139.38	75.08	100
755.	fire	215	93.35	74.92	100
756.	display	179	77.72	74.79	100
757.	heart	391	169.77	74.61	100
758.	quite	143	62.09	74.41	100
759.	course	142	61.66	74.23	100
760.	organ	313	135.90	74.14	100
761.	push	197	85.54	74.05	100
762.	primary	316	137.21	73.88	100
763.	enclose	242	105.08	73.57	100
764.	configuration	343	148.93	73.43	75
765.	reference	247	107.25	73.42	100

766.	proton	415	180.19	73.39	75
767.	friction	471	204.51	73.24	100
768.	old	203	88.14	73.07	100
769.	satisfy	209	90.75	72.84	100
770.	conduct	257	111.59	72.58	100
771.	hole	247	107.25	72.54	100
772.	mine	125	54.27	72.47	100
773.	dry	220	95.52	72.42	100
774.	black	198	85.97	72.35	75
775.	distribute	163	70.77	71.62	100
776.	familiar	122	52.97	71.52	100
777.	brain	420	182.36	71.51	100
778.	ask	131	56.88	71.49	100
779.	copper	334	145.02	71.45	100
780.	essential	163	70.77	71.37	100
781.	put	124	53.84	71.29	100
782.	deal	129	56.01	71.28	100
783.	gram	179	77.72	71.27	100
784.	contrast	142	61.66	71.00	100
785.	fly	247	107.25	70.96	100
786.	exceed	133	57.75	70.75	100
787.	simultaneously	136	59.05	70.74	100
788.	watch	189	82.06	70.69	100
789.	gain	160	69.47	70.67	100
790.	propose	174	75.55	70.59	100
791.	check	141	61.22	70.36	100
792.	hour	182	79.02	70.21	100
793.	skin	315	136.77	70.21	100
794.	extreme	215	93.35	70.19	100
795.	synthesize	295	128.09	70.02	50
796.	eliminate	139	60.35	69.95	100
797.	wavelength	581	252.27	69.91	75
798.	polar	400	173.68	69.69	100
799.	consequence	124	53.84	69.65	100
800.	avoid	144	62.52	69.58	100
801.	cold	224	97.26	69.35	100
802.	past	161	69.91	69.12	100
803.	kind	133	57.75	69.04	100
804.	child	267	115.93	69.02	100
805.	displace	174	75.55	68.94	100
806.	valid	146	63.39	68.85	100
807.	hint	174	75.55	68.83	100
808.	strength	195	84.67	68.80	100

809.	floor	221	95.96	68.77	100
810.	branch	266	115.50	68.76	100
811.	deliver	271	117.67	68.64	100
812.	assign	185	80.33	68.63	100
813.	specify	155	67.30	68.62	100
814.	shift	193	83.80	68.60	100
815.	inner	215	93.35	68.46	100
816.	vessel	268	116.37	68.10	100
817.	absolute	240	104.21	67.98	100
818.	mind	151	65.56	67.93	100
819.	interact	144	62.52	67.91	100
820.	acquire	163	70.77	67.85	100
821.	accelerate	265	115.06	67.83	100
822.	wind	212	92.05	67.76	100
823.	stage	262	113.76	67.60	100
824.	partial	302	131.13	67.58	100
825.	attract	196	85.10	67.57	100
826.	engine	377	163.69	67.48	100
827.	deep	145	62.96	67.48	100
828.	distinguish	146	63.39	67.44	100
829.	react	330	143.29	67.39	75
830.	medium	299	129.83	67.35	100
831.	dioxide	242	105.08	67.20	75
832.	nuclear	344	149.36	66.93	100
833.	following	145	62.96	66.88	100
834.	sugar	347	150.67	66.85	100
835.	deter	111	48.20	66.59	100
836.	spherical	271	117.67	66.58	100
837.	synthesis	253	109.85	66.43	100
838.	reactant	445	193.22	66.23	50
839.	fail	168	72.95	66.18	100
840.	send	147	63.83	66.06	100
841.	warm	223	96.83	65.95	100
842.	separation	196	85.10	65.91	75
843.	student	176	76.42	65.88	100
844.	description	131	56.88	65.70	100
845.	previous	113	49.06	65.70	100
846.	genetic	416	180.63	65.65	100
847.	respond	180	78.16	65.63	100
848.	achieve	111	48.20	65.31	100
849.	coefficient	285	123.75	65.24	75
850.	subject	137	59.49	65.23	100
851.	reveal	150	65.13	65.17	100

852.	arrangement	151	65.56	64.85	100
853.	detail	114	49.50	64.66	100
854.	leaf	407	176.72	64.53	100
855.	isolate	154	66.87	64.53	75
856.	thing	117	50.80	64.52	100
857.	thick	188	81.63	64.20	100
858.	oil	241	104.64	63.90	100
859.	burn	196	85.10	63.74	100
860.	stimulate	359	155.88	63.58	75
861.	investigate	118	51.24	63.48	100
862.	male	494	214.49	63.43	75
863.	voltage	589	255.74	63.26	100
864.	lack	213	92.48	63.19	100
865.	research	133	57.75	63.13	100
866.	full	129	56.01	62.93	100
867.	verify	124	53.84	62.92	100
868.	north	228	99.00	62.92	100
869.	yellow	173	75.12	62.89	100
870.	damage	190	82.50	62.79	100
871.	closely	141	61.22	62.68	100
872.	consequently	117	50.80	62.67	75
873.	heavy	149	64.70	62.50	100
874.	arrive	152	66.00	62.41	100
875.	mix	176	76.42	62.39	100
876.	laboratory	133	57.75	62.34	100
877.	immediately	118	51.24	62.26	100
878.	ensure	126	54.71	62.24	100
879.	computer	135	58.62	61.92	100
880.	readily	123	53.41	61.90	100
881.	depth	176	76.42	61.87	100
882.	pound	163	70.77	61.86	100
883.	notation	161	69.91	61.80	75
884.	compose	190	82.50	61.75	100
885.	bound	361	156.75	61.69	100
886.	particularly	129	56.01	61.57	100
887.	conversion	240	104.21	61.48	75
888.	occupy	165	71.64	61.21	100
889.	otherwise	92	39.95	61.20	100
890.	rotation	260	112.89	61.09	100
891.	numerical	140	60.79	61.07	100
892.	battery	553	240.11	61.02	100
893.	ionic	358	155.44	60.73	50
894.	mate	206	89.44	60.72	100

895.	female	438	190.18	60.71	75
896.	match	125	54.27	60.67	100
897.	percentage	135	58.62	60.66	100
898.	exhibit	149	64.70	60.61	100
899.	hard	134	58.18	60.58	100
900.	loop	549	238.37	60.58	75
901.	aluminum	249	108.12	60.51	100
902.	vapor	387	168.03	60.43	75
903.	conservation	199	86.41	60.33	100
904.	division	369	160.22	60.25	100
905.	mathematical	123	53.41	60.22	100
906.	attempt	112	48.63	59.93	100
907.	boundary	227	98.56	59.86	100
908.	gravity	176	76.42	59.67	100
909.	triangle	211	91.62	59.61	100
910.	record	142	61.66	59.56	100
911.	planet	243	105.51	59.38	100
912.	paper	193	83.80	59.36	100
913.	die	211	91.62	59.31	100
914.	error	232	100.73	59.22	100
915.	intermediate	160	69.47	59.16	100
916.	essentially	130	56.45	59.11	100
917.	capacity	258	112.02	59.08	100
918.	stretch	185	80.33	59.05	100
919.	appendix	166	72.08	59.01	100
920.	dark	218	94.66	58.92	75
921.	normally	125	54.27	58.85	100
922.	ence	109	47.33	58.57	100
923.	composition	226	98.13	58.49	100
924.	evolution	353	153.27	58.28	75
925.	interior	179	77.72	58.23	100
926.	arrange	129	56.01	58.10	100
927.	domain	422	183.23	58.07	100
928.	lower	139	60.35	57.71	100
929.	interpret	115	49.93	57.59	100
930.	requirement	110	47.76	57.59	100
931.	precisely	113	49.06	57.49	100
932.	emit	200	86.84	57.48	100
933.	tional	117	50.80	57.47	100
934.	read	120	52.10	57.39	100
935.	middle	123	53.41	57.15	100
936.	molar	336	145.89	57.04	75
937.	width	198	85.97	57.04	100

938.	proceed	136	59.05	56.91	100
939.	hair	254	110.29	56.83	75
940.	alternative	118	51.24	56.75	100
941.	manner	105	45.59	56.69	100
942.	protect	195	84.67	56.66	75
943.	young	206	89.44	56.62	100
944.	concern	99	42.99	56.52	100
945.	themselves	125	54.27	56.47	100
946.	bird	289	125.48	56.47	75
947.	consume	183	79.46	56.45	100
948.	slide	244	105.94	56.41	75
949.	cube	206	89.44	56.35	100
950.	hit	158	68.60	56.15	100
951.	round	148	64.26	56.04	100
952.	earlier	92	39.95	55.63	100
953.	bone	361	156.75	55.51	100
954.	aqueous	372	161.52	55.51	75
955.	chemistry	209	90.75	55.33	75
956.	share	172	74.68	55.25	75
957.	ignore	128	55.58	55.07	100
958.	capture	208	90.31	54.98	100
959.	precise	106	46.03	54.88	100
960.	cost	203	88.14	54.87	100
961.	imply	106	46.03	54.84	100
962.	equally	110	47.76	54.72	100
963.	whole	106	46.03	54.66	100
964.	mi	170	73.81	54.52	100
965.	reasonable	96	41.68	54.48	100
966.	disk	280	121.58	54.38	100
967.	treatment	152	66.00	54.33	100
968.	ml	472	204.94	54.31	75
969.	forward	144	62.52	54.30	100
970.	possibility	117	50.80	54.15	100
971.	minute	131	56.88	54.10	100
972.	operate	182	79.02	54.00	100
973.	box	212	92.05	53.94	100
974.	pump	227	98.56	53.92	100
975.	conductor	432	187.57	53.90	50
976.	enable	96	41.68	53.84	100
977.	sunlight	218	94.66	53.78	100
978.	mostly	134	58.18	53.77	100
979.	modify	133	57.75	53.75	100
980.	destroy	166	72.08	53.72	100

981.	extent	152	66.00	53.70	100
982.	sheet	181	78.59	53.62	100
983.	representation	126	54.71	53.61	100
984.	exchange	215	93.35	53.36	75
985.	container	253	109.85	53.28	100
986.	oxide	340	147.63	53.13	75
987.	track	150	65.13	53.08	100
988.	conclusion	118	51.24	53.07	100
989.	plasma	336	145.89	53.04	75
990.	collision	430	186.71	52.92	100
991.	man	160	69.47	52.86	100
992.	agree	96	41.68	52.71	100
993.	prepare	169	73.38	52.70	100
994.	comparison	98	42.55	52.65	100
995.	proper	118	51.24	52.55	100
996.	survive	201	87.27	52.50	100
997.	stay	104	45.16	52.44	100
998.	fully	108	46.89	52.43	100
999.	trace	128	55.58	52.30	100
1000.	favor	170	73.81	52.29	100
1001.	transform	124	53.84	52.24	100
1002.	home	134	58.18	52.23	100
1003.	percent	276	119.84	52.23	100
1004.	clearly	100	43.42	52.20	100
1005.	terminal	258	112.02	52.19	100
1006.	somewhat	107	46.46	52.04	100
1007.	tiny	137	59.49	51.97	75
1008.	kill	202	87.71	51.94	75
1009.	motor	231	100.30	51.86	100
1010.	atmospheric	162	70.34	51.65	100
1011.	escape	132	57.31	51.61	100
1012.	decay	278	120.71	51.58	100
1013.	wonder	96	41.68	51.56	100
1014.	chloride	217	94.22	51.55	75
1015.	local	275	119.40	51.48	100
1016.	exam	100	43.42	51.45	100
1017.	collect	135	58.62	51.27	100
1018.	resemble	129	56.01	51.21	100
1019.	storage	126	54.71	51.13	100
1020.	manufacture	137	59.49	51.13	100
1021.	seed	387	168.03	51.12	100
1022.	decide	108	46.89	51.11	100
1023.	impossible	108	46.89	51.03	100

1024.	empty	128	55.58	51.00	100
1025.	meaning	94	40.81	50.97	100
1026.	mark	96	41.68	50.94	100
1027.	interest	100	43.42	50.89	100
1028.	compute	129	56.01	50.85	75
1029.	quickly	105	45.59	50.80	100
1030.	proportion	131	56.88	50.79	100
1031.	forest	384	166.73	50.77	75
1032.	explanation	115	49.93	50.75	100
1033.	stem	319	138.51	50.72	75
1034.	possess	119	51.67	50.37	75
1035.	oxidation	466	202.34	50.34	50
1036.	orbit	286	124.18	50.32	100
1037.	biological	177	76.85	50.31	75
1038.	suspend	132	57.31	50.24	100
1039.	insect	270	117.23	50.22	75
1040.	finite	177	76.85	50.19	100
1041.	fit	94	40.81	50.06	100
1042.	momentum	510	221.44	50.02	100
1043.	x-axis	371	161.09	50.00	50
1044.	switch	267	115.93	49.97	100
1045.	despite	105	45.59	49.73	100
1046.	integrate	190	82.50	49.70	100
1047.	cylindrical	175	75.98	49.69	100
1048.	thousand	128	55.58	49.68	100
1049.	periodic	246	106.81	49.68	100
1050.	differentiate	189	82.06	49.65	100
1051.	scientific	166	72.08	49.62	100
1052.	freely	99	42.99	49.58	100
1053.	coil	413	179.32	49.46	100
1054.	unlike	105	45.59	49.41	100
1055.	tie	94	40.81	49.15	100
1056.	health	206	89.44	49.15	75
1057.	reproduce	250	108.55	49.14	100
1058.	geometry	164	71.21	49.09	75
1059.	chromosome	657	285.27	49.06	50
1060.	death	161	69.91	49.05	100
1061.	automobile	113	49.06	49.05	100
1062.	steel	167	72.51	48.86	100
1063.	behave	110	47.76	48.78	100
1064.	hypothesis	189	82.06	48.63	100
1065.	denote	115	49.93	48.58	100
1066.	upon	95	41.25	48.37	100
1067.	alone	82	35.60	48.32	100

1068.	strongly	110	47.76	48.29	75
1069.	regardless	98	42.55	48.21	100
1070.	interesting	93	40.38	48.18	100
1071.	fruit	263	114.19	48.17	100
1072.	contribution	109	47.33	48.11	100
1073.	column	179	77.72	48.09	100
1074.	ten	103	44.72	48.04	100
1075.	throw	169	73.38	48.01	100
1076.	continuously	98	42.55	48.00	100
1077.	select	114	49.50	47.92	100
1078.	content	118	51.24	47.88	100
1079.	emerge	140	60.79	47.84	100
1080.	electromagnetic	243	105.51	47.83	100
1081.	furthermore	115	49.93	47.82	100
1082.	acidic	239	103.77	47.72	50
1083.	environmental	163	70.77	47.69	75
1084.	transmit	166	72.08	47.68	100
1085.	reduction	324	140.68	47.66	100
1086.	waste	190	82.50	47.65	100
1087.	science	102	44.29	47.55	100
1088.	alter	111	48.20	47.52	100
1089.	photograph	112	48.63	47.46	100
1090.	fat	264	114.63	47.12	75
1091.	summary	96	41.68	47.10	100
1092.	critical	188	81.63	47.08	100
1093.	project	106	46.03	47.07	100
1094.	carefully	73	31.70	47.03	100
1095.	drug	194	84.23	46.99	75
1096.	transition	280	121.58	46.95	75
1097.	substitution	214	92.92	46.90	100
1098.	touch	137	59.49	46.89	100
1099.	feed	176	76.42	46.88	100
1100.	discovery	124	53.84	46.87	100
1101.	selection	281	122.01	46.86	75
1102.	eight	113	49.06	46.79	100
1103.	century	105	45.59	46.77	100
1104.	soil	298	129.39	46.69	75
1105.	none	100	43.42	46.61	100
1106.	tank	221	95.96	46.60	100
1107.	coat	148	64.26	46.59	100
1108.	primarily	114	49.50	46.53	100
1109.	mercury	262	113.76	46.53	100
1110.	tance	110	47.76	46.47	100
1111.	calcium	172	74.68	46.42	75

1112.	hundred	125	54.27	46.34	100
1113.	intensity	341	148.06	46.30	100
1114.	understanding	92	39.95	46.21	100
1115.	mile	135	58.62	46.16	100
1116.	fourth	106	46.03	46.14	100
1117.	cation	226	98.13	46.11	100
1118.	prediction	92	39.95	46.09	100
1119.	screen	246	106.81	46.08	100
1120.	lake	170	73.81	46.07	100
1121.	entirely	86	37.34	46.04	100
1122.	corner	126	54.71	45.95	100
1123.	cord	276	119.84	45.91	75
1124.	attack	214	92.92	45.88	75
1125.	convenient	92	39.95	45.84	100
1126.	stant	105	45.59	45.73	100
1127.	greek	87	37.78	45.73	100
1128.	pole	260	112.89	45.67	100
1129.	stick	135	58.62	45.62	100
1130.	roughly	119	51.67	45.40	100
1131.	adjust	94	40.81	45.19	100
1132.	usual	81	35.17	45.19	100
1133.	arbitrary	105	45.59	44.97	75
1134.	cancel	110	47.76	44.88	75
1135.	parent	199	86.41	44.85	75
1136.	farther	100	43.42	44.83	100
1137.	rapid	105	45.59	44.71	100
1138.	importance	75	32.56	44.70	100
1139.	cellular	208	90.31	44.68	75
1140.	tension	285	123.75	44.56	100
1141.	mineral	182	79.02	44.46	100
1142.	opening	128	55.58	44.45	75
1143.	hang	144	62.52	44.45	100
1144.	report	93	40.38	44.26	100
1145.	unique	115	49.93	44.08	100
1146.	soon	91	39.51	43.95	100
1147.	induce	337	146.32	43.89	100
1148.	combustion	221	95.96	43.71	75
1149.	rectangle	247	107.25	43.70	100
1150.	tail	156	67.73	43.67	100
1151.	abundant	145	62.96	43.66	75
1152.	previously	93	40.38	43.65	100
1153.	sufficient	82	35.60	43.58	100
1154.	absence	82	35.60	43.48	100

1155.	bright	183	79.46	43.47	75
1156.	star	160	69.47	43.47	100
1157.	adjacent	145	62.96	43.37	100
1158.	structural	150	65.13	43.30	75
1159.	weigh	155	67.30	43.15	100
1160.	extract	119	51.67	43.11	75
1161.	ordinary	90	39.08	43.10	100
1162.	isolated	145	62.96	42.98	75
1163.	rectangular	158	68.60	42.98	75
1164.	regulate	220	95.52	42.91	75
1165.	flower	386	167.60	42.90	100
1166.	thickness	149	64.70	42.88	100
1167.	ionization	389	168.90	42.79	50
1168.	indeed	63	27.35	42.78	100
1169.	sufficiently	84	36.47	42.78	100
1170.	deposit	124	53.84	42.68	100
1171.	bend	135	58.62	42.67	100
1172.	operation	86	37.34	42.61	100
1173.	rare	118	51.24	42.60	100
1174.	ammonia	147	63.83	42.58	75
1175.	briefly	82	35.60	42.58	100
1176.	resultant	197	85.54	42.57	75
1177.	pose	82	35.60	42.56	100
1178.	vertically	122	52.97	42.54	100
1179.	adult	206	89.44	42.48	75
1180.	partially	103	44.72	42.46	100
1181.	ally	87	37.78	42.43	100
1182.	right-hand	134	58.18	42.38	100
1183.	reproduction	249	108.12	42.37	50
1184.	exact	89	38.64	42.35	100
1185.	purpose	81	35.17	42.35	100
1186.	geometric	154	66.87	42.25	75
1187.	accurate	82	35.60	42.15	100
1188.	building	119	51.67	42.07	100
1189.	efficient	97	42.12	42.02	100
1190.	something	70	30.39	41.99	100
1191.	expose	121	52.54	41.90	100
1192.	electricity	124	53.84	41.90	100
1193.	narrow	114	49.50	41.89	100
1194.	researcher	193	83.80	41.85	75
1195.	extra	83	36.04	41.80	100
1196.	tendency	121	52.54	41.79	100
1197.	tool	91	39.51	41.78	100
1198.	count	96	41.68	41.75	100

1199.	axe	180	78.16	41.66	100
1200.	impact	129	56.01	41.63	100
1201.	negligible	106	46.03	41.59	100
1202.	predator	251	108.98	41.57	50
1203.	functional	170	73.81	41.56	100
1204.	minimize	98	42.55	41.53	100
1205.	infinite	144	62.52	41.46	75
1206.	melt	165	71.64	41.37	100
1207.	climate	211	91.62	41.37	100
1208.	letter	94	40.81	41.35	100
1209.	moon	173	75.12	41.34	100
1210.	physic	119	51.67	41.34	100
1211.	amino	282	122.44	41.32	50
1212.	wheel	252	109.42	41.28	75
1213.	window	140	60.79	41.25	100
1214.	sulfur	237	102.90	41.24	100
1215.	frame	302	131.13	41.23	100
1216.	walk	104	45.16	41.17	100
1217.	improve	87	37.78	41.14	100
1218.	crystal	212	92.05	41.09	75
1219.	orientation	108	46.89	41.09	100
1220.	plus	86	37.34	40.91	100
1221.	arm	142	61.66	40.83	100
1222.	radial	206	89.44	40.82	100
1223.	thereby	89	38.64	40.78	100
1224.	mention	70	30.39	40.77	100
1225.	practical	85	36.91	40.72	100
1226.	future	107	46.46	40.67	100
1227.	limited	90	39.08	40.53	100
1228.	concentrate	81	35.17	40.43	100
1229.	observer	376	163.26	40.38	100
1230.	leg	155	67.30	40.37	75
1231.	differential	192	83.37	40.33	100
1232.	sit	85	36.91	40.27	100
1233.	receptor	368	159.78	40.21	50
1234.	diversity	256	111.15	40.20	50
1235.	characterize	86	37.34	40.15	100
1236.	silver	201	87.27	40.14	100
1237.	confine	82	35.60	40.05	100
1238.	amplitude	319	138.51	40.05	75
1239.	stationary	129	56.01	39.88	75
1240.	week	122	52.97	39.81	100
1241.	existence	79	34.30	39.76	100

1242.	photo	88	38.21	39.65	100
1243.	wish	81	35.17	39.63	100
1244.	fairly	82	35.60	39.56	100
1245.	secondary	240	104.21	39.53	75
1246.	pathway	162	70.34	39.49	75
1247.	ship	113	49.06	39.47	100
1248.	desire	97	42.12	39.43	100
1249.	symmetry	159	69.04	39.43	100
1250.	faster	80	34.74	39.41	100
1251.	uniformly	132	57.31	39.41	100
1252.	american	81	35.17	39.40	100
1253.	connection	96	41.68	39.38	100
1254.	evolutionary	225	97.69	39.32	50
1255.	rain	189	82.06	39.28	100
1256.	stress	131	56.88	39.27	75
1257.	accompany	78	33.87	39.21	100
1258.	prey	230	99.87	39.16	75
1259.	rocket	160	69.47	39.13	75
1260.	inverse	191	82.93	39.12	75
1261.	apparatus	104	45.16	39.08	75
1262.	insert	121	52.54	39.00	100
1263.	significantly	78	33.87	38.95	100
1264.	careful	68	29.53	38.92	100
1265.	split	102	44.29	38.91	100
1266.	classify	87	37.78	38.84	100
1267.	dash	96	41.68	38.81	100
1268.	film	218	94.66	38.70	75
1269.	surroundings	173	75.12	38.67	75
1270.	beginning	74	32.13	38.66	100
1271.	conserve	140	60.79	38.65	100
1272.	restrict	75	32.56	38.62	100
1273.	alternate	128	55.58	38.61	100
1274.	powerful	73	31.70	38.60	100
1275.	contract	166	72.08	38.56	100
1276.	wood	102	44.29	38.55	100
1277.	integration	245	106.38	38.47	100
1278.	nerve	226	98.13	38.42	75
1279.	tip	140	60.79	38.40	100
1280.	neglect	100	43.42	38.34	100
1281.	alcohol	223	96.83	38.34	75
1282.	naturally	109	47.33	38.29	100
1283.	originally	71	30.83	38.27	100
1284.	claim	74	32.13	38.27	100
1285.	attraction	135	58.62	38.27	100

1286.	feel	91	39.51	38.21	100
1287.	believe	77	33.43	38.13	100
1288.	host	212	92.05	38.10	75
1289.	month	102	44.29	37.99	100
1290.	retain	102	44.29	37.91	100
1291.	aspect	78	33.87	37.90	100
1292.	accept	78	33.87	37.80	100
1293.	input	146	63.39	37.70	100
1294.	fiber	225	97.69	37.62	75
1295.	output	138	59.92	37.61	100
1296.	radioactive	292	126.79	37.54	100
1297.	really	67	29.09	37.51	100
1298.	solar	131	56.88	37.49	100
1299.	thermal	218	94.66	37.47	100
1300.	confirm	79	34.30	37.46	100
1301.	couple	97	42.12	37.37	100
1302.	massive	93	40.38	37.33	100
1303.	plastic	106	46.03	37.33	75
1304.	spend	85	36.91	37.30	100
1305.	beneath	91	39.51	37.14	100
1306.	south	133	57.75	37.11	100
1307.	glucose	292	126.79	37.06	50
1308.	direct	94	40.81	37.04	100
1309.	peak	88	38.21	37.03	100
1310.	map	138	59.92	37.00	100
1311.	calculus	132	57.31	36.93	75
1312.	rely	87	37.78	36.92	100
1313.	boil	176	76.42	36.85	100
1314.	depict	78	33.87	36.80	100
1315.	difficulty	67	29.09	36.79	100
1316.	virus	356	154.57	36.72	75
1317.	stream	103	44.72	36.56	100
1318.	revolution	129	56.01	36.49	100
1319.	target	145	62.96	36.36	100
1320.	appearance	90	39.08	36.33	100
1321.	chance	115	49.93	36.32	100
1322.	integer	154	66.87	36.31	75
1323.	expansion	214	92.92	36.26	100
1324.	categorize	89	38.64	36.21	75
1325.	horizontally	112	48.63	36.18	100
1326.	altitude	120	52.10	36.11	100
1327.	electrode	335	145.46	36.09	75
1328.	widely	86	37.34	35.99	100
1329.	restore	99	42.99	35.97	100

1330.	pulse	295	128.09	35.88	75
1331.	woman	180	78.16	35.78	75
1332.	transmission	104	45.16	35.75	100
1333.	ecosystem	297	128.96	35.73	50
1334.	strand	267	115.93	35.69	100
1335.	extension	74	32.13	35.68	100
1336.	reasoning	68	29.53	35.60	100
1337.	lung	206	89.44	35.56	100
1338.	analogous	81	35.17	35.53	100
1339.	complicated	64	27.79	35.53	100
1340.	microscopic	111	48.20	35.51	75
1341.	vacuum	101	43.85	35.40	100
1342.	accomplish	66	28.66	35.34	100
1343.	tall	86	37.34	35.33	100
1344.	relation	78	33.87	35.32	100
1345.	solvent	249	108.12	35.30	50
1346.	transformation	125	54.27	35.28	100
1347.	cancer	180	78.16	35.14	100
1348.	gaseous	157	68.17	35.14	50
1349.	preceding	71	30.83	35.13	100
1350.	winter	113	49.06	35.11	100
1351.	array	88	38.21	35.07	100
1352.	immune	288	125.05	35.07	50
1353.	version	76	33.00	34.92	100
1354.	generation	179	77.72	34.92	100
1355.	cubic	135	58.62	34.90	100
1356.	specialized	127	55.14	34.86	75
1357.	construction	76	33.00	34.78	100
1358.	chamber	112	48.63	34.78	100
1359.	rigid	148	64.26	34.74	100
1360.	chemist	117	50.80	34.73	75
1361.	forth	75	32.56	34.64	100
1362.	algebraic	112	48.63	34.61	75
1363.	instantaneous	182	79.02	34.59	100
1364.	lens	591	256.61	34.59	75
1365.	pool	113	49.06	34.53	100
1366.	east	121	52.54	34.51	100
1367.	nonzero	90	39.08	34.47	75
1368.	press	79	34.30	34.45	100
1369.	row	88	38.21	34.30	100
1370.	dot	101	43.85	34.26	100
1371.	seek	73	31.70	34.25	100
1372.	counterclockwise	125	54.27	34.21	100

1373.	unchanged	69	29.96	34.19	100
1374.	responsible	70	30.39	34.16	75
1375.	contraction	185	80.33	34.16	100
1376.	dense	99	42.99	34.09	100
1377.	cable	175	75.98	34.09	100
1378.	similarity	98	42.55	34.08	100
1379.	closer	69	29.96	34.00	100
1380.	trap	89	38.64	33.96	100
1381.	promote	112	48.63	33.94	75
1382.	random	173	75.12	33.86	100
1383.	city	75	32.56	33.73	100
1384.	capacitor	712	309.15	33.70	75
1385.	gland	247	107.25	33.59	75
1386.	compress	151	65.56	33.59	100
1387.	helium	141	61.22	33.57	100
1388.	potassium	111	48.20	33.54	75
1389.	covalent	271	117.67	33.48	50
1390.	care	66	28.66	33.46	100
1391.	ultimately	71	30.83	33.45	100
1392.	rubber	86	37.34	33.35	100
1393.	core	122	52.97	33.31	100
1394.	filter	112	48.63	33.30	100
1395.	copy	163	70.77	33.24	100
1396.	neutral	120	52.10	33.23	75
1397.	soluble	148	64.26	33.22	75
1398.	parameter	117	50.80	33.20	100
1399.	family	104	45.16	33.19	100
1400.	distant	89	38.64	33.18	100
1401.	nervous	198	85.97	33.06	50
1402.	cone	154	66.87	33.04	75
1403.	diverse	137	59.49	33.00	50
1404.	brown	102	44.29	32.92	75
1405.	greatly	65	28.22	32.89	75
1406.	gasoline	115	49.93	32.87	100
1407.	odd	107	46.46	32.83	100
1408.	outline	71	30.83	32.76	100
1409.	lift	121	52.54	32.75	100
1410.	national	78	33.87	32.68	100
1411.	band	105	45.59	32.65	100
1412.	identity	100	43.42	32.62	100
1413.	infection	150	65.13	32.50	75
1414.	everywhere	77	33.43	32.47	100
1415.	diffuse	158	68.60	32.46	75
1416.	engineer	63	27.35	32.46	100
1417.	originate	67	29.09	32.44	75

1418.	sweep	75	32.56	32.42	100
1419.	insulate	115	49.93	32.38	100
1420.	aid	64	27.79	32.35	100
1421.	enormous	80	34.74	32.19	75
1422.	country	124	53.84	32.17	75
1423.	steady	93	40.38	32.17	75
1424.	medical	88	38.21	32.15	100
1425.	knowledge	70	30.39	32.13	100
1426.	roll	108	46.89	32.08	100
1427.	category	74	32.13	32.03	100
1428.	successful	64	27.79	31.92	100
1429.	disorder	169	73.38	31.90	75
1430.	efficiency	147	63.83	31.89	100
1431.	unfortunately	63	27.35	31.89	100
1432.	mental	59	25.62	31.80	100
1433.	patient	135	58.62	31.74	100
1434.	solute	282	122.44	31.68	75
1435.	bulb	179	77.72	31.62	100
1436.	apparent	74	32.13	31.62	100
1437.	machine	77	33.43	31.56	100
1438.	orange	74	32.13	31.54	75
1439.	centimeter	81	35.17	31.47	100
1440.	accumulate	84	36.47	31.38	100
1441.	fusion	113	49.06	31.37	75
1442.	river	106	46.03	31.34	100
1443.	properly	57	24.75	31.31	100
1444.	projection	105	45.59	31.27	100
1445.	cluster	106	46.03	31.25	75
1446.	nothing	52	22.58	31.22	100
1447.	photosynthesis	194	84.23	31.16	75
1448.	mirror	442	191.92	31.15	100
1449.	oscillate	156	67.73	31.07	75
1450.	advanced	63	27.35	31.03	100
1451.	overlap	115	49.93	31.02	100
1452.	electronic	97	42.12	31.02	100
1453.	jump	84	36.47	31.00	100
1454.	disappear	73	31.70	30.98	75
1455.	balloon	158	68.60	30.95	100
1456.	bubble	91	39.51	30.89	100
1457.	crucial	88	38.21	30.88	75
1458.	train	116	50.37	30.84	100
1459.	cavity	173	75.12	30.81	75
1460.	him	77	33.43	30.77	75
1461.	repre	55	23.88	30.76	100

1462.	exponential	216	93.79	30.74	100
1463.	flight	94	40.81	30.73	100
1464.	stomach	148	64.26	30.63	50
1465.	fore	62	26.92	30.48	100
1466.	necessarily	52	22.58	30.44	100
1467.	metallic	125	54.27	30.44	100
1468.	attention	51	22.14	30.39	100
1469.	incorporate	62	26.92	30.37	75
1470.	handle	70	30.39	30.33	100
1471.	proof	103	44.72	30.31	100
1472.	cross-sectional	113	49.06	30.26	75
1473.	torque	308	133.73	30.23	50
1474.	attractive	69	29.96	30.22	75
1475.	ear	136	59.05	30.19	75
1476.	offer	56	24.32	30.17	100
1477.	obey	63	27.35	30.16	100
1478.	industrial	72	31.26	30.09	100
1479.	helpful	53	23.01	30.07	100
1480.	search	74	32.13	30.04	100
1481.	pack	85	36.91	30.00	75
1482.	exit	83	36.04	29.98	100
1483.	mountain	86	37.34	29.98	100
1484.	house	71	30.83	29.98	100
1485.	activate	133	57.75	29.93	100
1486.	correctly	51	22.14	29.92	100
1487.	habitat	198	85.97	29.86	50
1488.	tween	74	32.13	29.85	50
1489.	distinct	75	32.56	29.81	100
1490.	triple	135	58.62	29.79	100
1491.	clockwise	98	42.55	29.75	100
1492.	stone	87	37.78	29.74	100
1493.	big	56	24.32	29.73	100
1494.	fuse	114	49.50	29.67	75
1495.	recent	83	36.04	29.52	75
1496.	toxic	98	42.55	29.50	75
1497.	likewise	58	25.18	29.44	100
1498.	parabola	171	74.25	29.36	50
1499.	safe	51	22.14	29.34	100
1500.	milk	116	50.37	29.27	100
1501.	definite	137	59.49	29.26	100
1502.	argument	61	26.49	29.20	100
1503.	x-ray	109	47.33	29.19	100
1504.	strip	120	52.10	29.18	100
1505.	dead	92	39.95	29.08	100

1506.	exception	66	28.66	29.07	75
1507.	gold	106	46.03	29.03	75
1508.	gradually	59	25.62	28.96	100
1509.	channel	142	61.66	28.93	100
1510.	chlorine	137	59.49	28.90	75
1511.	dog	104	45.16	28.90	75
1512.	orient	70	30.39	28.90	100
1513.	network	90	39.08	28.90	75
1514.	pipe	157	68.17	28.88	100
1515.	emission	130	56.45	28.85	75
1516.	mother	156	67.73	28.82	75
1517.	shoot	170	73.81	28.80	75
1518.	convention	65	28.22	28.78	100
1519.	radio	123	53.41	28.74	100
1520.	three-dimensional	96	41.68	28.69	100
1521.	road	105	45.59	28.69	100
1522.	related	62	26.92	28.64	100
1523.	dipole	305	132.43	28.64	50
1524.	fossil	189	82.06	28.63	100
1525.	regular	54	23.45	28.52	100
1526.	sharp	71	30.83	28.50	100
1527.	neutron	199	86.41	28.47	100
1528.	frictionless	172	74.68	28.42	50
1529.	collide	106	46.03	28.41	100
1530.	coast	83	36.04	28.40	100
1531.	reversible	161	69.91	28.37	100
1532.	class	62	26.92	28.35	100
1533.	sensitive	82	35.60	28.31	100
1534.	sand	81	35.17	28.30	100
1535.	pond	72	31.26	28.29	100
1536.	capable	63	27.35	28.26	75
1537.	collection	57	24.75	28.24	100
1538.	rough	84	36.47	28.20	100
1539.	physics	72	31.26	28.12	100
1540.	overcome	58	25.18	28.08	100
1541.	regard	45	19.54	28.05	100
1542.	ionize	133	57.75	28.03	75
1543.	oscillation	177	76.85	28.01	100
1544.	suitable	72	31.26	28.01	100
1545.	attain	52	22.58	28.00	100
1546.	daughter	202	87.71	27.96	100
1547.	night	72	31.26	27.94	75
1548.	span	73	31.70	27.89	100
1549.	catch	70	30.39	27.89	100

1550.	benefit	114	49.50	27.88	75
1551.	intestine	158	68.60	27.79	50
1552.	intersection	87	37.78	27.77	75
1553.	leak	66	28.66	27.76	100
1554.	history	134	58.18	27.75	75
1555.	fine	51	22.14	27.73	100
1556.	flux	390	169.34	27.72	75
1557.	spontaneously	89	38.64	27.66	100
1558.	effectively	44	19.10	27.62	100
1559.	spectrum	169	73.38	27.58	75
1560.	diffusion	157	68.17	27.55	100
1561.	hypothetical	72	31.26	27.55	100
1562.	phosphate	157	68.17	27.48	50
1563.	hollow	86	37.34	27.44	75
1564.	bridge	117	50.80	27.44	100
1565.	ancient	76	33.00	27.43	100
1566.	hear	129	56.01	27.37	75
1567.	gap	77	33.43	27.35	100
1568.	standing	104	45.16	27.34	100
1569.	accompanying	118	51.24	27.29	100
1570.	decline	110	47.76	27.25	75
1571.	sex	163	70.77	27.25	50
1572.	solu	93	40.38	27.24	100
1573.	kidney	191	82.93	27.17	75
1574.	dependent	69	29.96	27.16	100
1575.	melting	133	57.75	27.11	100
1576.	consideration	60	26.05	27.09	100
1577.	effort	70	30.39	27.05	100
1578.	universe	76	33.00	26.96	100
1579.	argue	59	25.62	26.95	75
1580.	liter	88	38.21	26.95	100
1581.	speak	50	21.71	26.90	100
1582.	positively	87	37.78	26.87	75
1583.	mechanic	103	44.72	26.86	100
1584.	successive	53	23.01	26.84	100
1585.	intersect	89	38.64	26.83	75
1586.	remainder	59	25.62	26.80	100
1587.	increasingly	64	27.79	26.78	100
1588.	dangerous	64	27.79	26.75	100
1589.	island	149	64.70	26.74	100
1590.	huge	70	30.39	26.71	75
1591.	cap	76	33.00	26.71	100
1592.	airplane	75	32.56	26.67	100
1593.	disperse	101	43.85	26.64	100

1594.	plan	57	24.75	26.62	100
1595.	pick	53	23.01	26.59	100
1596.	fashion	52	22.58	26.57	100
1597.	preserve	90	39.08	26.55	100
1598.	suffer	77	33.43	26.52	75
1599.	wear	54	23.45	26.51	100
1600.	rank	64	27.79	26.47	100
1601.	combined	48	20.84	26.38	100
1602.	permit	50	21.71	26.34	100
1603.	pigment	97	42.12	26.33	75
1604.	blow	56	24.32	26.32	100
1605.	means	45	19.54	26.25	100
1606.	commercial	62	26.92	26.21	100
1607.	extensive	61	26.49	26.19	100
1608.	soft	59	25.62	26.11	100
1609.	west	70	30.39	26.05	75
1610.	tant	53	23.01	26.04	75
1611.	denominator	92	39.95	26.04	75
1612.	spot	85	36.91	26.02	100
1613.	physicist	51	22.14	25.97	100
1614.	frequently	57	24.75	25.95	100
1615.	metric	52	22.58	25.88	100
1616.	cloud	65	28.22	25.71	75
1617.	newly	59	25.62	25.71	100
1618.	representative	50	21.71	25.68	100
1619.	broad	55	23.88	25.57	100
1620.	win	56	24.32	25.18	100
1621.	unable	52	22.58	25.17	100
1622.	task	52	22.58	25.14	100
1623.	incorrect	53	23.01	25.00	100
1624.	university	48	20.84	24.99	100
1625.	perfectly	62	26.92	24.98	100
1626.	permanent	61	26.49	24.92	100
1627.	rearrange	47	20.41	24.90	100
1628.	anything	61	26.49	24.89	100
1629.	underlie	54	23.45	24.87	100
1630.	distinction	52	22.58	24.79	100
1631.	realize	44	19.10	24.78	100
1632.	subsequent	45	19.54	24.78	100
1633.	possibly	47	20.41	24.73	100
1634.	progress	57	24.75	24.39	100
1635.	slight	41	17.80	24.24	75
1636.	weather	59	25.62	24.21	100

1637.	strategy	59	25.62	24.07	100
1638.	wet	53	23.01	23.97	75
1639.	numerous	47	20.41	23.93	100
1640.	serious	55	23.88	23.91	100
1641.	designate	55	23.88	23.72	75
1642.	inject	72	31.26	23.69	75
1643.	currently	52	22.58	23.68	100
1644.	everyday	50	21.71	23.62	100
1645.	employ	49	21.28	23.60	100
1646.	unstable	62	26.92	23.59	100
1647.	suddenly	55	23.88	23.56	100
1648.	technology	56	24.32	23.42	100
1649.	entry	41	17.80	23.36	100
1650.	circumstance	42	18.24	23.25	100
1651.	insight	42	18.24	23.22	100
1652.	devise	47	20.41	23.10	75
1653.	wrong	40	17.37	23.06	100
1654.	advance	46	19.97	23.04	100
1655.	significance	43	18.67	22.99	100
1656.	notion	45	19.54	22.95	100
1657.	subtract	58	25.18	22.95	100
1658.	deduce	51	22.14	22.90	100
1659.	poor	54	23.45	22.88	75
1660.	detector	57	24.75	22.88	100
1661.	proportionality	52	22.58	22.83	75
1662.	minus	44	19.10	22.83	100
1663.	wait	56	24.32	22.78	100
1664.	brief	43	18.67	22.76	100
1665.	resist	57	24.75	22.76	100
1666.	elevation	66	28.66	22.75	100
1667.	constitute	50	21.71	22.72	100
1668.	fortunately	55	23.88	22.70	100
1669.	persist	53	23.01	22.61	75
1670.	trial	50	21.71	22.55	100
1671.	halfway	44	19.10	22.41	100
1672.	accuracy	47	20.41	22.17	100
1673.	guide	43	18.67	22.16	100
1674.	recover	48	20.84	22.16	100
1675.	accurately	40	17.37	22.14	100
1676.	minor	43	18.67	22.00	100
1677.	stance	43	18.67	21.98	100
1678.	belong	46	19.97	21.86	100
1679.	independently	51	22.14	21.82	100

1680.	organize	54	23.45	21.79	75
1681.	finding	42	18.24	21.79	100
1682.	sell	44	19.10	21.71	100
1683.	goal	40	17.37	21.57	100
1684.	straight-line	44	19.10	21.49	75
1685.	mathematically	47	20.41	21.01	75
1686.	visualize	41	17.80	20.96	100
1687.	sight	40	17.37	20.75	100
1688.	invent	41	17.80	20.57	100

C. KEYWORD LIST

	Item	*PoS*	*CEFR level*	ARF	Freq.	Rel. Freq.	Rel. DOCF
1.	point	*n*	*A2*	**1647.81**	4730	2053.8	100
2.	equation	*n*	*C1*	**1547.46**	4524	1964.3	100
3.	form	*v/n*	*A2*	**1485.96**	3177	1379.4	100
4.	value	*v/n*	*B1*	**1455.36**	3925	1704.2	100
5.	energy	*n*	*B1*	**1385.48**	5630	2444.5	100
6.	result	*v/n*	*B1*	**1202.46**	2170	942.21	100
7.	call	*v/n*	*A2*	**1200.78**	2409	1046	100
8.	produce	*v*	*B1*	**1140.44**	2639	1145.8	100
9.	function	*n*	*B2*	**1056.59**	3888	1688.2	100
10.	move	*v*	*A2*	**1027.44**	2666	1157.6	100
11.	increase	*v/n*	*B1*	**1017.93**	2469	1072	100
12.	follow	*v*	*A2*	**979.89**	1957	849.73	100
13.	constant	*adj*	*B2*	**955.76**	2514	1091.6	100
14.	large	*adj*	*A2*	**946.57**	1951	847.12	100
15.	system	*n*	*B1*	**907.43**	3056	1326.9	100
16.	cell*	*n*	*B2*	**879.47**	5311	2306	100
17.	determine	*v*	*C1*	**874.28**	1876	814.56	100
18.	describe	*v*	*A2*	**864.25**	1646	714.69	100
19.	mass*	*n*	*B2*	**859.35**	3384	1469.3	100
20.	force*	*v/n*	*B2*	**859.18**	4023	1746.8	100
21.	occur	*v*	*B2*	**852.65**	1834	796.32	100
22.	solution	*n*	*B1*	**830.47**	3048	1323.4	100
23.	high	*adj*	*A2*	**765.29**	1682	730.32	100
24.	contain	*v*	*B1*	**752.13**	1583	687.34	100
25.	line	*n*	*A2*	**749.05**	2327	1010.4	100
26.	molecule*	*n*		**741.08**	3143	1364.7	100
27.	unit	*n*	*B1*	**728.98**	1682	730.32	100
28.	surface	*n*	*B2*	**726.46**	2469	1072	100
29.	section	*n*	*B1*	**723.97**	1381	599.63	100
30.	consider	*v*	*B1*	**709.20**	1315	570.97	100
31.	cause	*v/n*	*B2*	**683.12**	1696	736.4	100
32.	equal	*adj*	*B1*	**673.78**	1462	634.8	100
33.	reaction*	*n*	*B2*	**662.37**	3807	1653	100
34.	speed	*n*	*B1*	**635.21**	2748	1193.2	100
35.	low	*adj*	*A2*	**606.04**	1307	567.5	100
36.	require	*v*	*B1*	**583.11**	1132	491.51	100
37.	assume	*v*	*B2*	**580.79**	1279	555.34	100

38.	base	*v/n*	*B1*	**576.56**	1631	708.18	100
39.	direction	*n*	*B1*	**575.33**	1884	818.03	100
40.	obtain	*v*	*B2*	**573.32**	1227	532.76	100
41.	process	*n*	*B2*	**568.71**	1586	688.64	100
42.	calculate	*v*	*B2*	**566.23**	1349	585.73	100
43.	type	*n*	*A2*	**536.21**	1238	537.54	100
44.	temperature	*n*	*A2*	**530.37**	2442	1060.3	100
45.	object	*n*	*B1*	**515.09**	2401	1042.5	100
46.	length	*n*	*B1*	**513.61**	1639	711.65	100
47.	represent	*v*	*B2*	**512.27**	1153	500.63	100
48.	distance	*n*	*B1*	**504.56**	1448	628.72	100
49.	explain	*v*	*A2*	**503.82**	1177	511.05	100
50.	apply	*v*	*B1*	**500.16**	1082	469.8	100
51.	rate	*v/n*	*B2*	**492.17**	1772	769.4	100
52.	charge	*v/n*	*B1*	**488.94**	3007	1305.6	100
53.	term	*n*	*A2*	**485.84**	1059	459.82	100
54.	area	*n*	*A2*	**485.27**	1530	664.32	100
55.	state	*adj/n*	*B2*	**479.53**	1321	573.58	100
56.	remain	*v*	*B1*	**478.61**	907	393.82	100
57.	measure	*v/n*	*B2*	**466.96**	1117	485	100
58.	positive	*adj*	*B1*	**442.45**	1196	519.3	100
59.	structure	*n*	*B2*	**437.96**	1713	743.78	100
60.	difference	*n*	*A2*	**434.64**	1053	457.21	100
61.	let	*v*	*A2*	**431.48**	874	379.49	100
62.	field	*n*	*A2*	**431.10**	2907	1262.2	100
63.	depend	*v*		**424.10**	794	344.75	100
64.	amount	*n*	*B1*	**412.67**	1030	447.22	100
65.	earth*	*n*	*B1*	**412.15**	1462	634.8	100
66.	side	*n*	*A2*	**411.25**	1001	434.63	100
67.	particle*	*n*	*C2*	**410.94**	1939	841.91	100
68.	air	*n*	*A2*	**405.04**	1345	584	100
69.	position	*n*	*B1*	**404.75**	1245	540.58	100
70.	single	*adj*	*A2*	**403.68**	806	349.96	100
71.	define	*v*	*B2*	**386.62**	947	411.19	100
72.	condition	*n*	*B1*	**381.86**	879	381.66	100
73.	reach	*v*	*B1*	**381.20**	816	354.31	100
74.	allow	*v*	*B1*	**378.85**	814	353.44	100
75.	consist	*v*		**378.65**	839	364.29	100
76.	chemical*	*adj/n*	*B2*	**378.03**	1183	513.66	100
77.	curve	*v/n*	*B2*	**375.21**	1725	748.99	100
78.	decrease	*v/n*	*B1*	**374.81**	888	385.57	100
79.	region	*n*	*B1*	**374.71**	1309	568.37	100
80.	pass	*v*	*A2*	**373.43**	809	351.27	100
81.	involve	*v*	*B1*	**367.46**	787	341.71	100

82.	simple	*adj*	*A2*	**364.28**	819	355.61	100
83.	volume	*n*	*B1*	**364.05**	1547	671.7	100
84.	expression	*n*	*B2*	**363.75**	943	409.45	100
85.	center	*n*		**362.13**	1182	513.22	100
86.	graph	*n*	*B2*	**361.98**	1721	747.25	100
87.	magnitude	*n*		**359.50**	1228	533.2	100
88.	motion	*n*	*C2*	**358.61**	1353	587.47	100
89.	product	*n*	*B1*	**351.79**	1111	482.39	100
90.	similar	*adj*	*B1*	**347.96**	653	283.53	100
91.	law	*n*	*B1*	**346.56**	1246	541.01	100
92.	reduce	*v*	*B1*	**345.24**	837	363.42	100
93.	ion*	*n*		**343.75**	1867	810.65	100
94.	suppose	*v*	*B1*	**337.39**	703	305.24	100
95.	method	*n*	*B1*	**335.43**	916	397.73	100
96.	pressure	*n*	*B2*	**334.30**	1853	804.57	100
97.	several	*adj*	*A2*	**332.43**	612	265.73	100
98.	add	*v*	*A2*	**329.82**	872	378.62	100
99.	compare	*v*	*B1*	**322.21**	606	263.12	100
100.	radius	*n*		**319.29**	1270	551.43	100
101.	release	*v*	*B2*	**313.64**	967	419.87	100
102.	potential	*adj*	*B2*	**311.83**	1676	727.72	100
103.	interval	*n*	*B1*	**309.37**	1209	524.95	100
104.	act	*v/n*	*B1*	**308.59**	866	376.02	100
105.	quantity	*n*	*B1*	**306.72**	769	333.9	100
106.	angle	*n*	*C1*	**305.51**	1275	553.6	100
107.	level	*n*	*A2*	**303.89**	902	391.65	100
108.	initial	*adj*	*B1*	**300.33**	941	408.58	100
109.	source	*n*	*B2*	**297.54**	839	364.29	100
110.	average	*n*	*B1*	**295.20**	928	402.94	100
111.	common	*adj*	*B1*	**293.55**	609	264.43	100
112.	current	*adj*	*B2*	**293.53**	1951	847.12	100
113.	illustrate	*v*	*B2*	**287.83**	535	232.3	100
114.	density	*n*	*C1*	**285.11**	1164	505.41	100
115.	approach	*v/n*	*B1*	**284.47**	737	320	100
116.	bond*	*v/n*	*B2*	**283.49**	2112	917.03	75
117.	solid	*adj*	*B2*	**280.78**	1105	479.79	100
118.	factor	*n*	*B2*	**280.75**	721	313.06	100
119.	space	*n*	*A2*	**278.72**	751	326.08	100
120.	velocity	*n*		**277.00**	1469	637.84	75
121.	human*	*n*	*B1*	**276.19**	1058	459.38	100
122.	material	*n*	*B1*	**275.05**	780	338.67	100
123.	component	*n*	*C1*	**274.49**	897	389.48	100
124.	certain	*adj*	*B1*	**274.46**	514	223.18	100
125.	property	*n*	*B1*	**271.23**	747	324.35	100

126.	heat*	*v/n*	*B1*	**271.23**	1478	641.74	100
127.	indicate	*v*	*B2*	**270.98**	564	244.89	100
128.	discuss	*v*	*A2*	**262.22**	478	207.55	100
129.	express	*v*	*B2*	**260.91**	613	266.16	100
130.	relate	*v*	*C2*	**258.32**	555	240.98	100
131.	power	*n*	*B1*	**258.07**	929	403.37	100
132.	rest	*v/n*	*A2*	**251.31**	740	321.31	100
133.	axis	*n*		**251.07**	1011	438.97	100
134.	substance	*n*	*B2*	**249.16**	814	353.44	100
135.	shape	*n*	*B1*	**247.99**	585	254.01	100
136.	situation	*n*	*B1*	**244.94**	556	241.41	100
137.	separate	*v/adj*	*B1*	**236.93**	531	230.56	100
138.	size	*n*	*A2*	**236.75**	569	247.06	100
139.	liquid*	*n*	*B1*	**230.85**	939	407.71	100
140.	step	*n*	*B1*	**229.98**	712	309.15	100
141.	relative	*adj*	*B1*	**228.52**	534	231.86	100
142.	enter	*v*	*A2*	**228.30**	575	249.66	100
143.	sum	*n*	*B1*	**228.25**	777	337.37	100
144.	solve	*v*	*B1*	**227.66**	556	241.41	100
145.	natural	*adj*	*B1*	**226.94**	697	302.64	100
146.	relationship	*n*	*B1*	**226.57**	477	207.11	100
147.	formula	*n*	*C1*	**226.56**	995	432.03	100
148.	equilibrium	*n*		**225.48**	1182	513.22	100
149.	compound	*n*		**224.08**	1157	502.37	100
150.	lie	*v*	*A2*	**223.83**	570	247.49	100
151.	vary	*v*	*B2*	**221.51**	462	200.6	100
152.	datum*	*n*		**219.76**	626	271.81	100
153.	horizontal	*adj*	*C1*	**218.88**	764	331.73	100
154.	diagram	*n*	*B1*	**217.14**	612	265.73	100
155.	blood	*n*	*A2*	**216.36**	1345	584	100
156.	estimate	*v/n*	*B2*	**215.78**	617	267.9	100
157.	exist	*v*	*B1*	**215.20**	511	221.88	100
158.	model	*n*	*A2*	**214.68**	578	250.97	100
159.	direct	*v/adj*	*B1*	**214.39**	521	226.22	100
160.	convert	*v*	*B2*	**213.72**	547	237.51	100
161.	vector	*n*		**208.18**	1538	667.8	100
162.	differ	*v*	*B2*	**206.84**	407	176.72	100
163.	vertical	*adj*	*C1*	**205.22**	613	266.16	100
164.	specific	*adj*	*B2*	**204.78**	606	263.12	100
165.	due	*adj*	*B1*	**204.12**	539	234.03	100
166.	wave	*n*	*B1*	**203.10**	1915	831.49	100
167.	directly	*adv*	*B1*	**201.01**	371	161.09	100
168.	path	*n*	*A2*	**200.89**	786	341.28	100
169.	identify	*adv*	*B2*	**200.80**	450	195.39	100
170.	divide	*v*	*B1*	**200.27**	428	185.84	100

171.	locate	*v*	*B1*	**199.15**	489	212.32	100
172.	replace	*v*	*B1*	**198.54**	400	173.68	100
173.	evaluate	*v*	*C1*	**196.37**	483	209.72	100
174.	connect	*v*	*B1*	**194.82**	607	263.56	100
175.	color*	*n*		**191.64**	687	298.29	100
176.	refer	*v*	*C2*	**190.25**	388	168.47	100
177.	acceleration	*n*		**189.71**	1183	513.66	100
178.	final	*adj*	*A2*	**189.63**	550	238.81	100
179.	series	*n*	*B1*	**189.37**	894	388.17	100
180.	behavior	*n*		**189.02**	516	224.05	100
181.	strong	*adj*	*A2*	**188.56**	604	262.26	100
182.	combine	*v*	*B2*	**186.76**	383	166.3	100
183.	observe	*v*	*B2*	**186.64**	424	184.1	100
184.	calculation	*n*	*B2*	**186.57**	517	224.48	100
185.	concept	*n*	*B2*	**186.56**	428	185.84	100
186.	remove	*v*	*B1*	**183.74**	384	166.73	100
187.	origin	*n*	*B2*	**182.49**	577	250.53	100
188.	rule	*n*	*B1*	**182.01**	663	287.87	100
189.	original	*adj*	*B1*	**181.25**	381	165.43	100
190.	flow	*v/n*	*B1*	**180.87**	615	267.03	100
191.	double	*adj*	*A2*	**179.81**	587	254.87	100
192.	coordinate	*v*		**178.62**	702	304.81	100
193.	useful	*adj*	*A2*	**178.43**	333	144.59	100
194.	parallel	*adj*	*C2*	**178.25**	610	264.86	100
195.	experiment	*v/n*	*B1*	**177.65**	491	213.19	100
196.	height	*n*	*B1*	**177.40**	581	252.27	100
197.	fix	*v*	*B1*	**175.51**	409	177.59	100
198.	attach	*v*	*B1*	**174.70**	474	205.81	100
199.	addition	*n*	*B1*	**172.84**	383	166.3	100
200.	derive	*v*	*C1*	**172.52**	339	147.19	100
201.	concentration	*n*	*B2*	**170.69**	809	351.27	100
202.	sample	*n*	*B2*	**169.91**	672	291.78	100
203.	associate	*v*	*C1*	**168.79**	349	151.54	100
204.	sign	*n*	*A2*	**168.74**	472	204.94	100
205.	drop	*v/n*	*B1*	**167.96**	407	176.72	100
206.	complex	*adj*	*B2*	**167.68**	593	257.48	100
207.	principle	*n*	*C1*	**166.83**	363	157.61	100
208.	exert	*v*		**165.91**	684	296.99	100
209.	square	*n*	*A2*	**165.75**	427	185.4	100
210.	sphere	*n*	*C1*	**164.88**	797	346.06	100
211.	ratio	*n*	*C1*	**163.71**	447	194.09	100
212.	molecular*	*adj*		**162.28**	700	303.94	100
213.	straight	*adj*	*A2*	**162.12**	398	172.81	100
214.	combination	*n*	*B2*	**161.31**	388	168.47	100
215.	variable	*adj/n*	*C1*	**160.91**	590	256.18	100

216.	store*	*n*	*B1*	**159.29**	550	238.81	100
217.	focus	*v/n*	*B2*	**158.59**	419	181.93	100
218.	generate	*v*	*B2*	**158.37**	423	183.67	100
219.	physical*	*adj*	*B2*	**156.60**	322	139.81	100
220.	maintain*	*v*	*B2*	**156.49**	407	176.72	100
221.	internal*	*adj*	*B2*	**156.35**	568	246.62	100
222.	active	*adj*	*B1*	**155.00**	393	170.64	100
223.	location	*n*	*B1*	**154.76**	335	145.46	100
224.	block	*v/n*	*B1*	**153.40**	801	347.79	100
225.	appropriate	*adj*	*B2*	**152.80**	300	130.26	100
226.	additional	*adj*	*B2*	**152.69**	274	118.97	100
227.	organism*	*n*		**152.17**	821	356.48	100
228.	theory	*n*	*B2*	**151.45**	568	246.62	100
229.	thin	*adj*	*A2*	**150.59**	448	194.52	100
230.	definition	*n*	*B2*	**150.58**	425	184.53	100
231.	species*	*n*	*B2*	**150.50**	652	283.1	75
232.	perpendicular	*adj*		**149.02**	555	240.98	100
233.	population	*n*	*B1*	**148.51**	1151	499.76	100
234.	circle	*v/n*	*A2*	**147.49**	559	242.72	100
235.	wire*	*n*	*B2*	**147.15**	943	409.45	100
236.	generally	*adv*	*B1*	**146.44**	290	125.92	100
237.	surround*	*v*	*B1*	**145.99**	391	169.77	100
238.	predict	*v*	*B1*	**145.89**	327	141.98	100
239.	rapidly	*adv*	*B2*	**145.68**	318	138.08	100
240.	perform	*v*	*B1*	**144.85**	306	132.86	100
241.	portion	*n*	*B2*	**144.18**	310	134.6	100
242.	frequency	*n*		**143.91**	1046	454.17	100
243.	identical	*adj*	*B2*	**143.82**	311	135.04	100
244.	multiple	*adj*	*C1*	**142.85**	348	151.1	100
245.	cycle	*v/n*	*B1*	**142.80**	672	291.78	100
246.	resistance*	*n*	*C2*	**142.77**	723	313.93	100
247.	central	*adj*	*B1*	**141.06**	457	198.43	100
248.	fill	*v*	*A2*	**139.45**	376	163.26	100
249.	typical	*adj*	*B1*	**139.08**	266	115.5	100
250.	formation	*n*	*C2*	**138.73**	455	197.56	100
251.	respect	*v/n*	*B1*	**138.09**	384	166.73	100
252.	kinetic*	*adj*		**137.01**	737	320	100
253.	variety	*n*	*A2*	**136.82**	276	119.84	100
254.	weight	*n*	*B1*	**135.60**	473	205.38	100
255.	relatively	*adv*	*B2*	**135.35**	271	117.67	100
256.	analyze	*v*		**134.39**	284	123.31	100
257.	undergo	*v*	*C1*	**134.17**	314	136.34	100
258.	basic	*adj*	*B1*	**133.64**	326	141.55	100
259.	circular	*adj*	*B2*	**132.65**	447	194.09	100

260.	continuous	*adj*	*B2*	**131.53**	559	242.72	100
261.	nucleus*	*n*		**131.48**	715	310.45	100
262.	substitute	*v/n*	*B2*	**130.60**	300	130.26	100
263.	approximately	*adv*	*B1*	**130.24**	260	112.89	100
264.	arise	*v*	*C1*	**129.78**	290	125.92	100
265.	atomic*	*adj*	*B2*	**129.10**	687	298.29	75
266.	plot	*n*	*B2*	**128.76**	340	147.63	100
267.	mole*	*n*		**128.41**	586	254.44	100
268.	uniform	*adj*	*A2*	**128.40**	522	226.65	100
269.	slightly	*adv*	*B2*	**128.16**	271	117.67	100
270.	integral	*adj*	*C1*	**128.00**	1004	435.93	100
271.	environment	*n*	*B1*	**127.29**	428	185.84	100
272.	spring	*n*	*A2*	**126.18**	700	303.94	100
273.	meter	*n*		**124.69**	325	141.11	100
274.	completely	*adv*	*B1*	**124.69**	248	107.68	100
275.	gene	*n*	*C1*	**123.87**	1105	479.79	50
276.	derivative	*n*		**123.83**	864	375.15	100
277.	fluid	*n*	*C2*	**123.64**	690	299.6	100
278.	correct	*adj*	*A2*	**123.06**	285	123.75	100
279.	magnetic	*adj*	*C1*	**122.26**	1422	617.43	100
280.	correspond	*v*	*B2*	**121.83**	276	119.84	100
281.	cylinder	*n*		**121.78**	530	230.13	100
282.	entire	*adj*	*B2*	**121.32**	227	98.563	100
283.	imagine	*v*	*B1*	**121.26**	245	106.38	100
284.	extend	*v*	*B2*	**120.83**	251	108.98	100
285.	upper	*adj*	*B1*	**120.23**	295	128.09	100
286.	presence	*n*	*B2*	**119.98**	255	110.72	100
287.	phase	*n*	*B2*	**119.67**	572	248.36	100
288.	muscle	*n*	*B2*	**118.77**	800	347.36	75
289.	mixture	*n*	*B2*	**118.70**	506	219.7	100
290.	significant	*adj*	*B2*	**118.61**	303	131.56	100
291.	proportional	*adj*		**118.18**	325	141.11	100
292.	typically	*adv*	*B1*	**116.87**	262	113.76	100
293.	equivalent	*adj*	*C1*	**115.56**	284	123.31	100
294.	absorb	*v*	*B2*	**114.47**	407	176.72	100
295.	characteristic	*n*	*B2*	**114.31**	239	103.77	100
296.	layer	*n*	*B2*	**113.72**	405	175.85	100
297.	reverse	*adj*	*B2*	**112.94**	273	118.54	100
298.	conclude	*v*	*C1*	**112.58**	222	96.392	100
299.	prevent	*v*	*B1*	**112.31**	290	125.92	100
300.	linear	*adj*		**112.24**	372	161.52	100
301.	corresponding	*v/adj*	*B2*	**111.87**	234	101.6	100
302.	membrane*	*n*		**111.48**	780	338.67	100
303.	balance	*v/n*	*B2*	**108.81**	331	143.72	100

304.	contribute	*v*	*B2*	**108.61**	250	108.55	100
305.	diameter	*n*		**107.65**	366	158.92	100
306.	growth	*n*	*B2*	**107.58**	556	241.41	100
307.	external	*adj*	*B2*	**107.58**	352	152.84	100
308.	ability	*n*	*B1*	**107.28**	258	112.02	100
309.	pattern	*n*	*B1*	**107.17**	437	189.74	100
310.	construct	*v/n*	*B2*	**106.99**	217	94.221	100
311.	tube	*n*	*B1*	**106.98**	422	183.23	100
312.	initially*	*adv*	*B2*	**105.84**	237	102.9	100
313.	edge*	*n*	*B1*	**103.91**	275	119.4	100
314.	bacterium	*n*		**103.86**	656	284.83	75
315.	dissolve	*v*	*C1*	**103.80**	414	179.76	100
316.	observation	*n*	*B2*	**103.79**	230	99.866	100
317.	shell	*n*	*B2*	**103.42**	478	207.55	100
318.	upward	*adj*	*C1*	**102.99**	314	136.34	100
319.	electrical	*adj*	*B1*	**102.76**	301	130.69	100
320.	sketch*	*v/n*	*C1*	**102.49**	313	135.9	100
321.	reflect	*v*	*B2*	**102.07**	453	196.69	100
322.	distribution	*n*	*C1*	**101.75**	362	157.18	100
323.	scale	*n*	*B2*	**101.48**	319	138.51	100
324.	outer	*adj*	*B2*	**101.41**	279	121.14	100
325.	device	*v/n*	*B2*	**100.98**	271	117.67	100
326.	ring	*n*	*A2*	**99.84**	412	178.89	100
327.	root	*n*	*B2*	**99.45**	590	256.18	100
328.	rotate	*v*		**99.14**	462	200.6	100
329.	chain	*n*	*A2*	**99.08**	443	192.35	100
330.	rod*	*n*		**98.65**	535	232.3	100
331.	enzyme	*n*		**98.56**	589	255.74	75
332.	symbol	*n*	*B2*	**98.29**	248	107.68	100
333.	yield	*v*	*C2*	**98.21**	270	117.23	100
334.	explore	*v*	*B1*	**97.83**	160	69.472	100
335.	multiply	*v*		**96.75**	214	92.918	100
336.	angular	*adj*		**96.57**	699	303.5	75
337.	fraction	*n*	*C2*	**96.48**	283	122.88	100
338.	theorem	*n*		**96.16**	631	273.98	75
339.	twice	*adj*	*A2*	**96.09**	184	79.892	100
340.	approximate	*v/adj*	*B2*	**95.75**	259	112.46	100
341.	encounter	*v*	*B2*	**94.95**	181	78.59	100
342.	overall	*adj*	*B2*	**94.72**	280	121.58	100
343.	respectively	*adv*	*C1*	**94.20**	188	81.629	100
344.	image	*n*	*B2*	**93.26**	685	297.43	100
345.	circuit	*n*	*C1*	**93.26**	901	391.21	100
346.	recall	*v*	*B2*	**93.22**	178	77.287	100
347.	iron*	*n*	*B1*	**92.04**	398	172.81	100

348.	slope*	*n*	*B2*	**91.73**	461	200.17	100
349.	response	*n*	*B2*	**91.63**	398	172.81	100
350.	sound	*adj/n*	*A2*	**91.52**	683	296.56	100
351.	atmosphere	*n*	*B1*	**91.44**	304	132	100
352.	mechanism	*n*	*C1*	**89.38**	348	151.1	75
353.	tissue	*n*	*B1*	**88.54**	531	230.56	100
354.	sodium	*n*		**88.37**	368	159.78	75
355.	orbital*	*adj*		**88.02**	1123	487.6	100
356.	segment	*n*		**88.00**	342	148.5	100
357.	measurement	*n*	*B2*	**87.29**	218	94.655	100
358.	sequence	*v/n*	*C1*	**87.18**	430	186.71	100
359.	displacement	*n*	*C1*	**86.95**	419	181.93	75
360.	interaction	*n*	*C1*	**86.94**	268	116.37	100
361.	simplify	*v*	*C1*	**86.55**	180	78.156	100
362.	scientist	*n*	*B1*	**86.15**	245	106.38	100
363.	weak	*adj*	*B1*	**85.89**	399	173.25	100
364.	consistent	*adj*	*C2*	**85.84**	174	75.55	100
365.	evolve	*v*	*C1*	**85.52**	340	147.63	75
366.	phenomenon	*n*	*C1*	**85.44**	192	83.366	100
367.	bind	*v*	*C2*	**85.31**	369	160.22	100
368.	disease	*n*	*B1*	**84.96**	443	192.35	75
369.	ice	*n*	*A2*	**84.66**	370	160.65	100
370.	expand	*v*	*B2*	**84.49**	222	96.392	100
371.	extremely	*adv*	*B1*	**84.41**	192	83.366	100
372.	nutrient	*n*		**84.36**	475	206.24	75
373.	commonly	*adv*	*C1*	**83.74**	159	69.038	100
374.	ray	*n*	*B2*	**83.64**	619	268.77	100
375.	label	*v/n*	*B1*	**83.45**	201	87.274	100
376.	string	*n*	*B2*	**81.98**	689	299.16	100
377.	gravitational*	*adj*		**81.62**	440	191.05	100
378.	mechanical*	*adj*	*B2*	**81.45**	261	113.33	100
379.	stable	*adj*	*C1*	**80.62**	272	118.1	100
380.	fuel	*v/n*	*B1*	**79.97**	333	144.59	100
381.	signal	*n*	*B2*	**79.50**	338	146.76	100
382.	recognize	*v*	*B1*	**79.42**	171	74.248	100
383.	cool	*v/adj*	*A2*	**79.40**	269	116.8	100
384.	transport	*v*	*B1*	**79.22**	346	150.23	100
385.	downward	*adj*	*C1*	**79.17**	230	99.866	100
386.	rock*	*n*	*B1*	**79.11**	306	132.86	100
387.	radiation	*n*	*C1*	**78.91**	426	184.97	100
388.	principal	*n*	*B1*	**78.74**	233	101.17	100
389.	detect	*v*	*C1*	**78.56**	233	101.17	100
390.	fundamental	*adj*	*C2*	**78.39**	214	92.918	100
391.	arrow*	*n*	*B2*	**78.15**	229	99.431	100
392.	variation	*n*	*B2*	**78.04**	186	80.761	100

393.	pure	*adj*	*B1*	**77.79**	300	130.26	75
394.	ocean*	*n*	*B1*	**77.68**	309	134.17	100
395.	approximation	*n*		**77.64**	295	128.09	75
396.	summarize	*v*	*C1*	**77.63**	153	66.432	100
397.	excess	*n*	*C1*	**77.50**	225	97.695	100
398.	instant	*adj*	*B2*	**77.46**	259	112.46	75
399.	tangent*	*n*		**77.15**	453	196.69	75
400.	dimension	*n*	*B2*	**76.96**	220	95.524	100
401.	unknown	*adj*	*B1*	**76.75**	202	87.708	100
402.	smooth	*adj*	*B1*	**76.75**	285	123.75	75
403.	assumption	*n*	*C1*	**76.43**	189	82.063	100
404.	experimental	*adj*	*C2*	**76.28**	165	71.643	100
405.	similarly	*adv*	*C1*	**76.07**	136	59.051	100
406.	beam*	*n*	*B2*	**75.76**	450	195.39	100
407.	visible	*adj*	*B2*	**75.17**	183	79.458	100
408.	agent	*n*	*B2*	**75.10**	304	132	75
409.	heart	*n*	*A2*	**74.61**	391	169.77	100
410.	primary	*adj*	*B2*	**73.88**	316	137.21	100
411.	enclose	*v*	*B2*	**73.57**	242	105.08	100
412.	configuration	*n*		**73.43**	343	148.93	75
413.	friction	*n*		**73.24**	471	204.51	100
414.	satisfy	*v*	*B2*	**72.84**	209	90.747	100
415.	conduct	*v/n*	*B2*	**72.58**	257	111.59	100
416.	hole*	*n*	*B1*	**72.54**	247	107.25	100
417.	distribute	*v*	*B2*	**71.62**	163	70.774	100
418.	brain	*n*	*A2*	**71.51**	420	182.36	100
419.	copper*	*n*	*B2*	**71.45**	334	145.02	100
420.	exceed	*v*	*C1*	**70.75**	133	57.748	100
421.	simultaneously	*adv*	*B2*	**70.74**	136	59.051	100
422.	skin	*n*	*B1*	**70.21**	315	136.77	100
423.	extreme	*adj*	*B2*	**70.19**	215	93.353	100
424.	synthesize	*v*		**70.02**	295	128.09	50
425.	eliminate	*v*	*C1*	**69.95**	139	60.354	100
426.	wavelength*	*n*	*C2*	**69.91**	581	252.27	75
427.	polar*	*adj*		**69.69**	400	173.68	100
428.	displace	*v*	*C1*	**68.94**	174	75.55	100
429.	valid	*adj*	*B2*	**68.85**	146	63.393	100
430.	hint	*v/n*	*B2*	**68.83**	174	75.55	100
431.	branch	*n*	*B1*	**68.76**	266	115.5	100
432.	deliver	*v*	*B1*	**68.64**	271	117.67	100
433.	assign	*v*	*C1*	**68.63**	185	80.327	100
434.	specify	*v*	*B2*	**68.62**	155	67.301	100
435.	shift	*v/n*	*B2*	**68.60**	193	83.8	100
436.	inner	*adj*	*B2*	**68.46**	215	93.353	100

437.	vessel*	*n*		**68.10**	268	116.37	100
438.	absolute	*adj*	*B2*	**67.98**	240	104.21	100
439.	interact	*v*	*B2*	**67.91**	144	62.525	100
440.	accelerate	*v*	*C1*	**67.83**	265	115.06	100
441.	partial	*adj*	*B2*	**67.58**	302	131.13	100
442.	attract	*v*	*B1*	**67.57**	196	85.103	100
443.	engine	*n*	*A2*	**67.48**	377	163.69	100
444.	distinguish	*v*	*B2*	**67.44**	146	63.393	100
445.	react	*v*	*B2*	**67.39**	330	143.29	75
446.	medium	*n*	*B1*	**67.35**	299	129.83	100
447.	nuclear	*adj*	*B2*	**66.93**	344	149.36	100
448.	following	*v/adj*	*A2*	**66.88**	145	62.959	100
449.	deter	*v*		**66.59**	111	48.196	100
450.	spherical*	*adj*		**66.58**	271	117.67	100
451.	synthesis	*n*		**66.43**	253	109.85	100
452.	reactant	*n*		**66.23**	445	193.22	50
453.	separation	*n*	*B2*	**65.91**	196	85.103	75
454.	genetic	*adj*	*C1*	**65.65**	416	180.63	100
455.	respond	*v*	*B2*	**65.63**	180	78.156	100
456.	coefficient	*n*		**65.24**	285	123.75	75
457.	leaf	*n*	*B1*	**64.53**	407	176.72	100
458.	isolate	*v*		**64.53**	154	66.867	75
459.	thick	*adj*	*B1*	**64.20**	188	81.629	100
460.	burn	*v*	*B1*	**63.74**	196	85.103	100
461.	stimulate	*v*	*B2*	**63.58**	359	155.88	75
462.	male	*n*	*B1*	**63.43**	494	214.49	75
463.	voltage*	*n*		**63.26**	589	255.74	100
464.	verify	*v*	*C1*	**62.92**	124	53.841	100
465.	consequently	*adv*	*B2*	**62.67**	117	50.801	75
466.	mix	*v*	*A2*	**62.39**	176	76.419	100
467.	laboratory	*n*	*B1*	**62.34**	133	57.748	100
468.	readily	*adv*	*B2*	**61.90**	123	53.406	100
469.	depth	*n*	*B1*	**61.87**	176	76.419	100
470.	notation	*n*		**61.80**	161	69.906	75
471.	compose	*v*	*B2*	**61.75**	190	82.498	100
472.	bound	*v/adj*	*B2*	**61.69**	361	156.75	100
473.	conversion	*n*	*C2*	**61.48**	240	104.21	75
474.	occupy	*v*	*B2*	**61.21**	165	71.643	100
475.	rotation	*n*		**61.09**	260	112.89	100
476.	numerical	*adj*		**61.07**	140	60.788	100
477.	battery	*n*	*A2*	**61.02**	553	240.11	100
478.	ionic*	*adj*		**60.73**	358	155.44	50
479.	mate*	*v/n*	*B1*	**60.72**	206	89.445	100
480.	female	*n*	*B1*	**60.71**	438	190.18	75
481.	percentage	*n*	*B2*	**60.66**	135	58.617	100

482.	exhibit	*v/n*	*C1*	**60.61**	149	64.696	100
483.	loop*	*n*		**60.58**	549	238.37	75
484.	vapor*	*n*		**60.43**	387	168.03	75
485.	conservation	*n*	*B2*	**60.33**	199	86.405	100
486.	division	*n*	*B2*	**60.25**	369	160.22	100
487.	mathematical	*adj*	*B2*	**60.22**	123	53.406	100
488.	boundary	*n*	*C1*	**59.86**	227	98.563	100
489.	gravity	*n*		**59.67**	176	76.419	100
490.	triangle	*n*	*B2*	**59.61**	211	91.616	100
491.	planet	*n*	*B1*	**59.38**	243	105.51	100
492.	error	*n*	*B2*	**59.22**	232	100.73	100
493.	intermediate	*adj*	*B1*	**59.16**	160	69.472	100
494.	essentially	*adv*	*B2*	**59.11**	130	56.446	100
495.	capacity	*n*	*B2*	**59.08**	258	112.02	100
496.	stretch	*v*	*B2*	**59.05**	185	80.327	100
497.	composition	*n*	*B1*	**58.49**	226	98.129	100
498.	evolution	*n*	*B2*	**58.28**	353	153.27	75
499.	interior	*n*	*B2*	**58.23**	179	77.721	100
500.	domain	*n*	*C1*	**58.07**	422	183.23	100
501.	lower	*adj*	*A2*	**57.71**	139	60.354	100
502.	precisely	*adv*	*B2*	**57.49**	113	49.064	100
503.	emit	*v*	*C2*	**57.48**	200	86.84	100
504.	molar*	*adj*		**57.04**	336	145.89	75
505.	width	*n*	*B2*	**57.04**	198	85.971	100
506.	proceed	*v*	*C1*	**56.91**	136	59.051	100
507.	consume	*v*	*B2*	**56.45**	183	79.458	100
508.	slide	*n*	*B2*	**56.41**	244	105.94	75
509.	cube	*n*		**56.35**	206	89.445	100
510.	bone*	*n*	*B1*	**55.51**	361	156.75	100
511.	aqueous*	*adj*		**55.51**	372	161.52	75
512.	chemistry	*n*	*A2*	**55.33**	209	90.747	75
513.	capture	*v/n*	*B2*	**54.98**	208	90.313	100
514.	precise	*adj*	*B2*	**54.88**	106	46.025	100
515.	pump*	*v/n*	*B1*	**53.92**	227	98.563	100
516.	conductor	*n*	*B2*	**53.90**	432	187.57	50
517.	sunlight*	*n*	*B2*	**53.78**	218	94.655	100
518.	modify	*v*	*C1*	**53.75**	133	57.748	100
519.	destroy	*v*	*B1*	**53.72**	166	72.077	100
520.	sheet	*n*	*A2*	**53.62**	181	78.59	100
521.	container	*n*	*B2*	**53.28**	253	109.85	100
522.	oxide*	*n*		**53.13**	340	147.63	75
523.	collision	*n*		**52.92**	430	186.71	100
524.	trace	*v/n*	*B2*	**52.30**	128	55.577	100
525.	favor	*v/n*	*B1*	**52.29**	170	73.814	100

526.	transform	*v*	*B2*	**52.24**	124	53.841	100
527.	percent	*n*	*B1*	**52.23**	276	119.84	100
528.	terminal	*adj*	*B2*	**52.19**	258	112.02	100
529.	atmospheric	*adj*		**51.65**	162	70.34	100
530.	decay	*v/n*	*B2*	**51.58**	278	120.71	100
531.	exam	*n*	*A2*	**51.45**	100	43.42	100
532.	resemble	*v*	*C1*	**51.21**	129	56.012	100
533.	storage	*n*	*B2*	**51.13**	126	54.709	100
534.	manufacture	*v*	*B2*	**51.13**	137	59.485	100
535.	seed*	*n*	*B2*	**51.12**	387	168.03	100
536.	compute	*v*		**50.85**	129	56.012	75
537.	forest	*n*	*A2*	**50.77**	384	166.73	75
538.	stem	*n*		**50.72**	319	138.51	75
539.	possess	*v*	*C1*	**50.37**	119	51.67	75
540.	oxidation*	*n*		**50.34**	466	202.34	50
541.	orbit	*n*		**50.32**	286	124.18	100
542.	biological*	*adj*	*B2*	**50.31**	177	76.853	75
543.	suspend	*v*	*B2*	**50.24**	132	57.314	100
544.	insect	*n*	*A2*	**50.22**	270	117.23	75
545.	finite	*adj*	*C2*	**50.19**	177	76.853	100
546.	momentum	*n*	*C2*	**50.02**	510	221.44	100
547.	switch	*v*	*B1*	**49.97**	267	115.93	100
548.	integrate	*v*	*C1*	**49.70**	190	82.498	100
549.	cylindrical	*adj*		**49.69**	175	75.985	100
550.	periodic	*adj*		**49.68**	246	106.81	100
551.	differentiate	*v*	*C1*	**49.65**	189	82.063	100
552.	scientific	*adj*	*B1*	**49.62**	166	72.077	100
553.	freely	*adv*	*B2*	**49.58**	99	42.986	100
554.	coil	*n*		**49.46**	413	179.32	100
555.	reproduce	*v*	*C1*	**49.14**	250	108.55	100
556.	geometry	*n*		**49.09**	164	71.209	75
557.	chromosome*	*n*		**49.06**	657	285.27	50
558.	steel*	*n*	*B2*	**48.86**	167	72.511	100
559.	behave	*v*	*B1*	**48.78**	110	47.762	100
560.	hypothesis	*n*	*C2*	**48.63**	189	82.063	100
561.	denote	*v*		**48.58**	115	49.933	100
562.	regardless	*adj*	*C1*	**48.21**	98	42.551	100
563.	column	*n*	*B2*	**48.09**	179	77.721	100
564.	continuously	*adv*	*B2*	**48.00**	98	42.551	100
565.	electromagnetic	*n*		**47.83**	243	105.51	100
566.	acidic*	*adj*		**47.72**	239	103.77	50
567.	transmit	*v*	*C1*	**47.68**	166	72.077	100
568.	reduction	*n*	*B2*	**47.66**	324	140.68	100
569.	summary	*n*	*B2*	**47.10**	96	41.683	100
570.	critical	*adj*	*B2*	**47.08**	188	81.629	100

571.	transition	*n*	*C2*	**46.95**	280	121.58	75
572.	substitution	*n*	*C1*	**46.90**	214	92.918	100
573.	discovery	*n*	*B2*	**46.87**	124	53.841	100
574.	selection	*n*	*B2*	**46.86**	281	122.01	75
575.	soil	*n*	*B2*	**46.69**	298	129.39	75
576.	tank*	*n*	*C1*	**46.60**	221	95.958	100
577.	primarily	*adv*	*B2*	**46.53**	114	49.499	100
578.	intensity	*n*	*C2*	**46.30**	341	148.06	100
579.	cation*	*n*		**46.11**	226	98.129	100
580.	prediction	*n*	*B2*	**46.09**	92	39.946	100
581.	screen	*n*	*A2*	**46.08**	246	106.81	100
582.	lake	*n*	*A2*	**46.07**	170	73.814	100
583.	cord	*n*		**45.91**	276	119.84	75
584.	convenient	*adj*	*B1*	**45.84**	92	39.946	100
585.	pole	*n*	*C2*	**45.67**	260	112.89	100
586.	roughly	*adv*	*B2*	**45.40**	119	51.67	100
587.	adjust	*v*	*B2*	**45.19**	94	40.815	100
588.	arbitrary	*adj*	*C2*	**44.97**	105	45.591	75
589.	cancel	*v*	*B1*	**44.88**	110	47.762	75
590.	farther	*adj*	*A2*	**44.83**	100	43.42	100
591.	rapid	*adj*	*B2*	**44.71**	105	45.591	100
592.	cellular*	*adj*		**44.68**	208	90.313	75
593.	tension	*n*	*B2*	**44.56**	285	123.75	100
594.	induce	*v*		**43.89**	337	146.32	100
595.	combustion	*n*		**43.71**	221	95.958	75
596.	rectangle	*n*	*C1*	**43.70**	247	107.25	100
597.	tail	*n*	*B2*	**43.67**	156	67.735	100
598.	abundant	*adj*		**43.66**	145	62.959	75
599.	bright	*adj*	*A2*	**43.47**	183	79.458	75
600.	tract*	*n*		**43.44**	177	76.853	75
601.	adjacent	*adj*	*C2*	**43.37**	145	62.959	100
602.	structural	*adj*	*C2*	**43.30**	150	65.13	75
603.	weigh	*v*	*B1*	**43.15**	155	67.301	100
604.	extract	*v*	*B2*	**43.11**	119	51.67	75
605.	isolated	*adj*	*C1*	**42.98**	145	62.959	75
606.	rectangular	*adj*	*B2*	**42.98**	158	68.603	75
607.	regulate	*v*	*C1*	**42.91**	220	95.524	75
608.	thickness	*n*		**42.88**	149	64.696	100
609.	ionization*	*n*		**42.79**	389	168.9	50
610.	sufficiently	*adv*	*C1*	**42.78**	84	36.473	100
611.	bend	*v*	*B2*	**42.67**	135	58.617	100
612.	resultant	*adj*		**42.57**	197	85.537	75
613.	arc	*n*		**42.55**	162	70.34	100
614.	vertically	*adv*		**42.54**	122	52.972	100

615.	partially	*adv*	*C1*	**42.46**	103	44.722	100
616.	reproduction	*n*	*C1*	**42.37**	249	108.12	50
617.	exact	*adj*	*B1*	**42.35**	89	38.644	100
618.	geometric	*adj*		**42.25**	154	66.867	75
619.	outward	*adj*		**42.03**	119	51.67	100
620.	expose	*v*	*B2*	**41.90**	121	52.538	100
621.	electricity	*n*	*A2*	**41.90**	124	53.841	100
622.	researcher	*n*	*B2*	**41.85**	193	83.8	75
623.	tendency	*n*	*C1*	**41.79**	121	52.538	100
624.	axe*	*n*		**41.66**	180	78.156	100
625.	negligible	*adj*	*C2*	**41.59**	106	46.025	100
626.	predator*	*n*	*C1*	**41.57**	251	108.98	50
627.	functional	*adj*	*B2*	**41.56**	170	73.814	100
628.	infinite	*adj*	*C2*	**41.46**	144	62.525	75
629.	melt	*v*	*B2*	**41.37**	165	71.643	100
630.	climate*	*n*	*B1*	**41.37**	211	91.616	100
631.	moon	*n*	*A2*	**41.34**	173	75.116	100
632.	wheel	*n*	*A2*	**41.28**	252	109.42	75
633.	frame	*v/n*	*B1*	**41.23**	302	131.13	100
634.	crystal	*n*	*C1*	**41.09**	212	92.05	75
635.	orientation	*n*	*C2*	**41.09**	108	46.893	100
636.	radial	*adj*		**40.82**	206	89.445	100
637.	observer	*n*	*C2*	**40.38**	376	163.26	100
638.	differential	*adj*		**40.33**	192	83.366	100
639.	receptor*	*n*		**40.21**	368	159.78	50
640.	diversity	*n*	*C1*	**40.20**	256	111.15	50
641.	characterize	*v*		**40.15**	86	37.341	100
642.	silver	*n*	*A2*	**40.14**	201	87.274	100
643.	amplitude	*n*		**40.05**	319	138.51	75
644.	stationary	*adj*		**39.88**	129	56.012	75
645.	secondary	*adj*	*B1*	**39.53**	240	104.21	75
646.	pathway*	*n*		**39.49**	162	70.34	75
647.	symmetry	*n*		**39.43**	159	69.038	100
648.	uniformly	*adv*		**39.41**	132	57.314	100
649.	evolutionary*	*adj*		**39.32**	225	97.695	50
650.	prey*	*n*	*C2*	**39.16**	230	99.866	75
651.	rocket	*n*	*B2*	**39.13**	160	69.472	75
652.	inverse	*adj*		**39.12**	191	82.932	75
653.	apparatus	*n*		**39.08**	104	45.157	75
654.	insert	*v*	*C1*	**39.00**	121	52.538	100
655.	classify	*v*	*C1*	**38.84**	87	37.775	100
656.	dash	*n*	*B2*	**38.81**	96	41.683	100
657.	surroundings	*n*	*B2*	**38.67**	173	75.116	75
658.	conserve	*v*		**38.65**	140	60.788	100

659.	alternate	*v*	*C1*	**38.61**	128	55.577	100
660.	integration	*n*	*C1*	**38.47**	245	106.38	100
661.	nerve	*n*	*C1*	**38.42**	226	98.129	75
662.	tip	*n*	*B1*	**38.40**	140	60.788	100
663.	neglect	*v*	*C1*	**38.34**	100	43.42	100
664.	attraction	*n*	*B1*	**38.27**	135	58.617	100
665.	host	*v/n*	*B2*	**38.10**	212	92.05	75
666.	input	*v/n*	*B2*	**37.70**	146	63.393	100
667.	fiber*	*n*		**37.62**	225	97.695	75
668.	radioactive*	*adj*		**37.54**	292	126.79	100
669.	solar	*adj*	*B2*	**37.49**	131	56.88	100
670.	thermal	*adj*		**37.47**	218	94.655	100
671.	glucose	*n*		**37.06**	292	126.79	50
672.	boil	*v*	*A2*	**36.85**	176	76.419	100
673.	depict	*v*	*C2*	**36.80**	78	33.867	100
674.	revolution	*n*	*B2*	**36.49**	129	56.012	100
675.	integer	*v*		**36.31**	154	66.867	75
676.	expansion	*n*	*B2*	**36.26**	214	92.918	100
677.	categorize	*v*		**36.21**	89	38.644	75
678.	horizontally	*adv*		**36.18**	112	48.63	100
679.	altitude	*n*		**36.11**	120	52.104	100
680.	pulse	*n*	*C1*	**35.88**	295	128.09	75
681.	transmission	*n*	*C2*	**35.75**	104	45.157	100
682.	strand	*n*	*C2*	**35.69**	267	115.93	100
683.	reasoning	*n*	*C2*	**35.60**	68	29.525	100
684.	lung*	*n*	*B2*	**35.56**	206	89.445	100
685.	analogous	*adj*	*C2*	**35.53**	81	35.17	100
686.	complicated	*adj*	*B1*	**35.53**	64	27.789	100
687.	microscopic	*adj*		**35.51**	111	48.196	75
688.	vacuum	*v/n*		**35.40**	101	43.854	100
689.	accomplish	*v*	*C1*	**35.34**	66	28.657	100
690.	solvent*	*n*		**35.30**	249	108.12	50
691.	transformation	*n*	*C1*	**35.28**	125	54.275	100
692.	gaseous	*adj*		**35.14**	157	68.169	50
693.	preceding	*v/adj*	*C2*	**35.13**	71	30.828	100
694.	array	*n*		**35.07**	88	38.209	100
695.	immune*	*adj*	*C2*	**35.07**	288	125.05	50
696.	generation	*n*	*B1*	**34.92**	179	77.721	100
697.	cubic	*adj*		**34.90**	135	58.617	100
698.	specialized	*v/adj*	*B2*	**34.86**	127	55.143	75
699.	chamber*	*n*		**34.78**	112	48.63	100
700.	rigid	*adj*	*C2*	**34.74**	148	64.261	100
701.	chemist	*n*	*A2*	**34.73**	117	50.801	75
702.	algebraic	*adj*		**34.61**	112	48.63	75
703.	instantaneous	*adj*		**34.59**	182	79.024	100

704.	dot	*n*	*B1*	**34.26**	101	43.854	100
705.	counterclockwise	*adj*		**34.21**	125	54.275	100
706.	unchanged	*adj*	*B2*	**34.19**	69	29.96	100
707.	contraction	*n*		**34.16**	185	80.327	100
708.	dense	*adj*	*B2*	**34.09**	99	42.986	100
709.	cable	*n*	*B1*	**34.09**	175	75.985	100
710.	similarity	*n*	*B2*	**34.08**	98	42.551	100
711.	random	*n*	*C1*	**33.86**	173	75.116	100
712.	capacitor	*n*		**33.70**	712	309.15	75
713.	gland*	*n*		**33.59**	247	107.25	75
714.	compress	*v*		**33.59**	151	65.564	100
715.	covalent	*adj*		**33.48**	271	117.67	50
716.	rubber	*n*	*A2*	**33.35**	86	37.341	100
717.	core	*n*	*C2*	**33.31**	122	52.972	100
718.	filter	*v/n*	*B2*	**33.30**	112	48.63	100
719.	neutral	*adj*	*C1*	**33.23**	120	52.104	75
720.	soluble	*adj*		**33.22**	148	64.261	75
721.	parameter	*n*		**33.20**	117	50.801	100
722.	distant	*adj*	*B2*	**33.18**	89	38.644	100
723.	nervous	*adj*	*B1*	**33.06**	198	85.971	50
724.	cone	*n*		**33.04**	154	66.867	75
725.	diverse	*adj*	*B2*	**33.00**	137	59.485	50
726.	infection*	*n*	*B2*	**32.50**	150	65.13	75
727.	diffuse*	*v*		**32.46**	158	68.603	75
728.	originate	*v*	*C2*	**32.44**	67	29.091	75
729.	insulate	*v*		**32.38**	115	49.933	100
730.	steady	*adj*	*B2*	**32.17**	93	40.38	75
731.	disorder	*n*	*C1*	**31.90**	169	73.379	75
732.	efficiency	*n*	*B2*	**31.89**	147	63.827	100
733.	solute*	*n*		**31.68**	282	122.44	75
734.	bulb	*n*	*B2*	**31.62**	179	77.721	100
735.	centimeter	*n*		**31.47**	81	35.17	100
736.	accumulate	*v*	*C2*	**31.38**	84	36.473	100
737.	fusion*	*n*		**31.37**	113	49.064	75
738.	projection	*n*	*C1*	**31.27**	105	45.591	100
739.	cluster*	*v/n*		**31.25**	106	46.025	75
740.	photosynthesis*	*n*		**31.16**	194	84.234	75
741.	mirror	*v/n*	*A2*	**31.15**	442	191.92	100
742.	oscillate	*v*		**31.07**	156	67.735	75
743.	overlap	*v*	*C2*	**31.02**	115	49.933	100
744.	electronic	*adj*	*B1*	**31.02**	97	42.117	100
745.	bubble	*n*	*C1*	**30.89**	91	39.512	100
746.	cavity	*n*		**30.81**	173	75.116	75
747.	exponential	*adj*		**30.74**	216	93.787	100
748.	stomach	*n*	*A2*	**30.63**	148	64.261	50

749.	metallic	*adj*		**30.44**	125	54.275	100
750.	proof	*n*	*B2*	**30.31**	103	44.722	100
751.	torque*	*n*		**30.23**	308	133.73	50
752.	obey	*v*	*B2*	**30.16**	63	27.354	100
753.	exit	*v/n*	*A2*	**29.98**	83	36.038	100
754.	activate	*v*		**29.93**	133	57.748	100
755.	tween*	*n*		**29.85**	74	32.131	50
756.	triple	*adj*		**29.79**	135	58.617	100
757.	clockwise	*adj*		**29.75**	98	42.551	100
758.	fuse*	*v/n*		**29.67**	114	49.499	75
759.	toxic	*adj*	*B2*	**29.50**	98	42.551	75
760.	parabola	*n*		**29.36**	171	74.248	50
761.	definite	*adj*	*B2*	**29.26**	137	59.485	100
762.	strip	*n*	*C1*	**29.18**	120	52.104	100
763.	orient	*v*		**28.90**	70	30.394	100
764.	pipe	*n*	*B1*	**28.88**	157	68.169	100
765.	emission	*n*	*C1*	**28.85**	130	56.446	75
766.	dipole	*n*		**28.64**	305	132.43	50
767.	fossil	*n*		**28.63**	189	82.063	100
768.	neutron	*n*		**28.47**	199	86.405	100
769.	frictionless	*adj*		**28.42**	172	74.682	50
770.	collide	*v*		**28.41**	106	46.025	100
771.	reversible	*adj*	*C2*	**28.37**	161	69.906	100
772.	pond	*n*	*B2*	**28.29**	72	31.262	100
773.	physics	*n*	*A2*	**28.12**	72	31.262	100
774.	ionize	*v*		**28.03**	133	57.748	75
775.	oscillation	*n*		**28.01**	177	76.853	100
776.	attain	*v*	*C1*	**28.00**	52	22.578	100
777.	span	*n*	*C2*	**27.89**	73	31.696	100
778.	intestine*	*n*		**27.79**	158	68.603	50
779.	intersection*	*n*		**27.77**	87	37.775	75
780.	leak	*v*	*B2*	**27.76**	66	28.657	100
781.	flux	*n*		**27.72**	390	169.34	75
782.	spontaneously	*adv*		**27.66**	89	38.644	100
783.	spectrum	*n*	*C1*	**27.58**	169	73.379	75
784.	diffusion	*n*		**27.55**	157	68.169	100
785.	hypothetical	*adj*		**27.55**	72	31.262	100
786.	hollow	*adj*	*C2*	**27.44**	86	37.341	75
787.	standing	*v/adj*	*A2*	**27.34**	104	45.157	100
788.	accompanying	*v/adj*	*B1*	**27.29**	118	51.235	100
789.	kidney*	*n*	*C2*	**27.17**	191	82.932	75
790.	melting	*v/n*	*B2*	**27.11**	133	57.748	100
791.	positively	*adv*	*B2*	**26.87**	87	37.775	75
792.	mechanic	*adj*	*A2*	**26.86**	103	44.722	100
793.	intersect	*v*		**26.83**	89	38.644	75

794.	remainder	*n*		**26.80**	59	25.618	100
795.	airplane	*n*		**26.67**	75	32.565	100
796.	disperse	*v*		**26.64**	101	43.854	100
797.	fragment	*n*		**26.48**	108	46.893	100
798.	filament	*n*		**26.42**	164	71.209	75
799.	denominator	*n*		**26.04**	92	39.946	75
800.	biologist	*n*		**26.02**	108	46.893	100
801.	physicist	*n*		**25.97**	51	22.144	100
802.	converge	*v*		**25.93**	226	98.129	100
803.	metric	*adj*		**25.88**	52	22.578	100
804.	seawater	*n*		**25.81**	90	39.078	100
805.	probability	*n*	*C1*	**25.77**	240	104.21	100
806.	elastic	*adj*		**25.61**	117	50.801	75
807.	spin	*v/n*	*C1*	**25.29**	137	59.485	100
808.	thermodynamics	*n*		**25.27**	123	53.406	75
809.	lifetime	*n*	*B2*	**25.23**	64	27.789	100
810.	seal	*v/n*	*B2*	**25.23**	99	42.986	75
811.	respiratory	*adj*		**25.01**	148	64.261	75
812.	incorrect	*adj*	*B1*	**25.00**	53	23.013	100
813.	laser	*n*	*B2*	**24.99**	126	54.709	75
814.	penetrate	*v*		**24.99**	75	32.565	75
815.	capillary	*n*		**24.91**	252	109.42	75
816.	rearrange	*v*	*B2*	**24.90**	47	20.407	100
817.	index	*n*	*C1*	**24.81**	201	87.274	100
818.	crop	*n*	*B1*	**24.60**	149	64.696	50
819.	spontaneous	*adj*		**24.59**	188	81.629	75
820.	projectile	*n*		**24.54**	184	79.892	75
821.	bulk	*n*	*C1*	**24.45**	73	31.696	75
822.	steam	*n*	*B2*	**24.43**	125	54.275	100
823.	experimentally	*adv*		**24.36**	67	29.091	75
824.	metabolic	*adj*		**24.21**	102	44.288	75
825.	absorption	*n*		**24.17**	80	34.736	100
826.	compression	*n*		**24.07**	109	47.328	100
827.	heating	*v/n*	*A2*	**23.99**	90	39.078	75
828.	numerator	*n*		**23.99**	63	27.354	75
829.	drift	*v/n*	*C2*	**23.96**	107	46.459	75
830.	conceptualize	*v*		**23.95**	69	29.96	50
831.	vibrate	*v*		**23.90**	119	51.67	100
832.	harmful	*adj*	*B2*	**23.82**	75	32.565	100
833.	float	*v*	*B1*	**23.81**	86	37.341	100
834.	grain	*n*	*C2*	**23.81**	106	46.025	75
835.	vibration	*n*		**23.78**	110	47.762	75
836.	barrier	*n*	*B2*	**23.77**	93	40.38	75
837.	graphical	*adj*		**23.76**	71	30.828	100
838.	scalar	*adj*		**23.71**	150	65.13	50

839.	inject	*v*		**23.69**	72	31.262	75
840.	synthetic	*adj*		**23.61**	78	33.867	75
841.	unstable	*adj*		**23.59**	62	26.92	100
842.	oxidize	*v*		**23.59**	202	87.708	75
843.	bacterial	*adj*		**23.49**	103	44.722	75
844.	digest	*v*	*C1*	**23.48**	108	46.893	75
845.	satellite	*n*	*B2*	**23.47**	132	57.314	100
846.	compact	*adj*		**23.42**	63	27.354	100
847.	infect	*v*	*C1*	**23.30**	115	49.933	50
848.	catalyze	*v*		**23.23**	77	33.433	50
849.	expel	*v*		**23.13**	78	33.867	100
850.	skeleton	*n*	*B2*	**22.99**	146	63.393	50
851.	subtract	*v*		**22.95**	58	25.183	100
852.	trigger	*v*	*C1*	**22.92**	97	42.117	75
853.	observed	*v/adj*	*B2*	**22.92**	66	28.657	75
854.	deduce	*v*	*C2*	**22.90**	51	22.144	100
855.	detector	*n*		**22.88**	57	24.749	100
856.	stability	*n*	*C1*	**22.84**	93	40.38	100
857.	baseball	*n*	*A2*	**22.83**	102	44.288	100
858.	proportionality	*n*		**22.83**	52	22.578	75
859.	balanced	*adj*	*B2*	**22.79**	120	52.104	75
860.	elevation	*n*		**22.75**	66	28.657	100
861.	dilute	*v*		**22.71**	119	51.67	50
862.	symmetric	*adj*		**22.60**	68	29.525	50
863.	reflection	*n*	*B2*	**22.48**	189	82.063	100
864.	discharge	*v/n*		**22.36**	104	45.157	100
865.	logarithm	*n*		**22.36**	100	43.42	75
866.	empirical	*adj*	*C2*	**22.32**	88	38.209	75
867.	hydrocarbon	*n*		**22.32**	131	56.88	75
868.	dominant	*adj*	*C1*	**22.32**	120	52.104	100
869.	subscript	*n*		**22.11**	56	24.315	50
870.	collectively	*adv*		**22.11**	61	26.486	75
871.	quantum	*n*		**22.10**	191	82.932	50
872.	plausible	*adj*	*C2*	**22.07**	107	46.459	75
873.	harmonic	*adj*		**22.06**	244	105.94	50
874.	reservoir	*n*		**22.06**	183	79.458	100
875.	adaptation	*n*	*C1*	**22.05**	114	49.499	50
876.	decompose	*v*		**21.97**	74	32.131	75
877.	resistor	*n*		**21.88**	411	178.46	50
878.	independently	*adv*	*B2*	**21.82**	51	22.144	100
879.	carrier	*n*		**21.77**	117	50.801	75
880.	gradient	*n*		**21.73**	136	59.051	75
881.	pendulum*	*n*		**21.62**	139	60.354	50
882.	adapt	*v*	*B2*	**21.56**	84	36.473	75
883.	immerse	*v*		**21.54**	66	28.657	100

884.	evaporate	*v*		**21.53**	59	25.618	75
885.	straight-line	*adj*	*A2*	**21.49**	44	19.105	75
886.	conversely	*adv*		**21.48**	55	23.881	100
887.	stimulus	*n*	*C2*	**21.46**	148	64.261	50
888.	electrostatic	*adj*		**21.41**	129	56.012	50
889.	catalyst	*n*		**21.34**	129	56.012	75
890.	isotope	*n*		**21.29**	186	80.761	100
891.	shrink	*v*	*B2*	**21.25**	54	23.447	100
892.	neuron	*n*		**21.23**	344	149.36	50
893.	circumference	*n*		**21.13**	51	22.144	100
894.	freezing	*v/adj*	*B1*	**21.10**	91	39.512	75
895.	interference	*n*	*C1*	**21.10**	270	117.23	100
896.	quantitative	*adj*		**21.04**	49	21.276	50
897.	mathematically	*adv*		**21.01**	47	20.407	75
898.	visualize	*v*		**20.96**	41	17.802	100
899.	marine*	*n*		**20.95**	92	39.946	75
900.	shaded	*v/adj*		**20.93**	52	22.578	100
901.	pollen*	*n*		**20.91**	209	90.747	50
902.	evaporation	*n*		**20.87**	63	27.354	75
903.	backward	*adj*	*C2*	**20.68**	43	18.671	100
904.	calculator	*n*	*B1*	**20.63**	75	32.565	100
905.	rope	*n*	*B2*	**20.61**	141	61.222	75
906.	fractional	*adj*		**20.50**	69	29.96	75
907.	curvature*	*n*		**20.48**	144	62.525	100
908.	coating	*v/n*		**20.44**	54	23.447	75
909.	equator	*n*		**20.41**	92	39.946	100
910.	static	*adj*		**20.38**	105	45.591	100
911.	inhibit	*v*		**20.33**	112	48.63	75
912.	revolve	*v*		**20.31**	154	66.867	75
913.	trigonometric	*adj*		**20.25**	110	47.762	75
914.	decomposition	*n*		**20.25**	116	50.367	75
915.	required	*v*	*B1*	**20.19**	46	19.973	100
916.	interactive	*adj*	*B2*	**20.19**	60	26.052	50
917.	resonance	*n*		**20.15**	161	69.906	75
918.	concentrated	*v/adj*	*B1*	**20.15**	72	31.262	75
919.	spacecraft	*n*		**20.10**	134	58.183	75
920.	brake	*v/n*	*B1*	**20.07**	91	39.512	100
921.	mouse	*n*	*A2*	**20.02**	97	42.117	50
922.	terrestrial*	*adj*		**20.00**	100	43.42	50
923.	node*	*n*		**19.94**	189	82.063	75
924.	entropy	*n*		**19.92**	395	171.51	75
925.	incidence	*n*		**19.78**	86	37.341	100
926.	moist	*n*		**19.70**	84	36.473	50
927.	inversely	*adv*		**19.68**	41	17.802	75
928.	abundance	*adj*		**19.56**	90	39.078	50

929.	radian	*n*		**19.54**	97	42.117	50
930.	precipitate	*v*		**19.47**	178	77.287	50
931.	frog	*n*	*B1*	**19.40**	101	43.85	50
932.	inherit*	*v*	*C2*	**19.31**	131	56.88	50
933.	boiling	*v/n*	*A2*	**19.31**	102	44.288	75
934.	arbitrarily	*adv*		**19.24**	47	20.407	100
935.	metabolism	*n*		**19.14**	77	33.433	50
936.	specified	*v/adj*	*B2*	**19.10**	50	21.71	100
937.	aquatic*	*adj*		**19.00**	88	38.209	50
938.	tropical*	*adj*	*B2*	**18.99**	107	46.459	75
939.	protective	*adj*	*B2*	**18.94**	52	22.578	75
940.	microscope	*n*		**18.94**	66	28.657	75
941.	derivation	*n*		**18.92**	45	19.539	100
942.	bounce*	*v/n*	*B2*	**18.87**	58	25.183	100
943.	magnet	*n*		**18.86**	141	61.222	75
944.	liver*	*n*	*B2*	**18.86**	88	38.209	75
945.	rotational	*adj*		**18.77**	118	51.235	50
946.	inward*	*adj*		**18.64**	44	19.105	75
947.	junction	*n*		**18.64**	109	47.328	50
948.	precision	*n*		**18.60**	51	22.144	100
949.	embed	*v*		**18.58**	64	27.789	75
950.	truck*	*n*	*B1*	**18.53**	81	35.17	75
951.	composite	*adj/n*		**18.51**	69	29.96	100
952.	qualitative	*adj*		**18.48**	58	25.183	50
953.	donor	*n*	*C2*	**18.47**	94	40.815	50
954.	thermodynamic	*adj*		**18.39**	96	41.683	50
955.	decimal	*n*		**18.38**	96	41.683	100
956.	quotient	*n*		**18.36**	121	52.538	50
957.	partition	*n*		**18.31**	133	57.748	75
958.	skeletal	*adj*		**18.26**	114	49.499	50
959.	dimensional	*adj*		**18.26**	47	20.407	100
960.	align	*v*		**18.23**	59	25.618	100
961.	breakdown	*n*	*B2*	**18.22**	79	34.302	75
962.	randomly	*adv*	*C1*	**18.21**	38	16.5	100
963.	violet	*n*		**18.17**	65	28.223	75
964.	condense	*adj*		**18.15**	65	28.223	75
965.	surrounding	*n/adj*	*B1*	**18.08**	45	19.539	100
966.	repel*	*v*		**18.07**	58	25.183	75
967.	binding	*v/n*	*C2*	**18.05**	84	36.473	75
968.	rewrite	*v*	*B2*	**18.04**	37	16.065	75
969.	bloodstream	*n*		**18.00**	86	37.341	50
970.	astronaut*	*n*		**17.99**	71	30.828	100
971.	valve*	*n*		**17.89**	61	26.486	100
972.	interstitial	*adj*		**17.88**	151	65.564	50

973.	incident	*n*	*B2*	**17.87**	171	74.248	50
974.	infinitesimal	*adj*		**17.86**	60	26.052	75
975.	invade	*v*	*B2*	**17.80**	83	36.038	75
976.	hemisphere	*n*		**17.78**	92	39.946	75
977.	trajectory*	*n*		**17.76**	100	43.42	75
978.	freshwater	*n*		**17.66**	71	30.828	75
979.	indicator	*n*		**17.59**	109	47.328	100
980.	diverge	*v*		**17.56**	129	56.012	75
981.	spiral	*adj*	*C2*	**17.55**	46	19.973	100
982.	encode	*v*		**17.43**	102	44.288	50
983.	generalize	*v*	*C1*	**17.23**	40	17.368	100
984.	physician	*n*		**17.18**	64	27.789	75
985.	conceptual	*adj*		**17.18**	48	20.842	75
986.	stabilize	*v*		**17.10**	61	26.486	75
987.	saturated	*v/adj*		**17.02**	84	36.473	75
988.	snake	*n*	*A2*	**17.00**	91	39.512	50
989.	quadratic	*adj*		**16.94**	73	31.696	75
990.	differentiation	*n*		**16.90**	84	36.473	75
991.	mold	*n*		**16.84**	67	29.091	50
992.	feather*	*n*	*B2*	**16.83**	55	23.881	100
993.	regenerate	*v*		**16.80**	57	24.749	50
994.	schematic	*adj*		**16.79**	37	16.065	75
995.	absent	*adj*	*B1*	**16.75**	81	35.17	100
996.	pore	*n*		**16.70**	92	39.946	50
997.	bullet	*n*	*B2*	**16.69**	96	41.683	100
998.	numerically	*adv*		**16.65**	35	15.197	75
999.	insoluble	*adj*		**16.64**	70	30.394	75
1000.	violate	*v*	*C2*	**16.58**	36	15.631	100
1001.	signify	*v*		**16.55**	47	20.407	75
1002.	physiology	*n*		**16.53**	117	50.801	75
1003.	radiate	*v*		**16.52**	69	29.96	75
1004.	respiration	*n*		**16.50**	125	54.275	50
1005.	droplet	*n*		**16.50**	44	19.105	75
1006.	solubility	*n*		**16.50**	168	72.945	50
1007.	discrete	*adj*		**16.44**	39	16.934	100
1008.	midpoint	*n*		**16.43**	49	21.276	75
1009.	harmless	*adj*	*B2*	**16.43**	52	22.578	75
1010.	liberate	*v*		**16.41**	54	23.447	75
1011.	disrupt	*v*	*B2*	**16.37**	47	20.407	75
1012.	maximize	*v*	*C2*	**16.31**	47	20.407	100
1013.	biochemical	*adj*		**16.31**	48	20.842	50
1014.	transparent	*adj*	*B2*	**16.28**	53	23.013	75
1015.	binary	*adj*		**16.23**	65	28.223	100
1016.	wedge	*n*		**16.23**	88	38.209	100
1017.	linearly	*adv*		**16.18**	39	16.934	75

1018.	impulse	*n*	*C2*	**16.15**	67	29.091	75
1019.	precipitation	*n*		**16.12**	127	55.143	50
1020.	triangular	*adj*		**16.05**	55	23.881	100
1021.	helix	*n*		**16.05**	135	58.617	100
1022.	quadrant	*n*		**16.03**	98	42.551	50
1023.	generator	*n*		**15.97**	67	29.091	100
1024.	reciprocal	*adj*		**15.85**	45	19.539	75
1025.	eject	*v*		**15.85**	50	21.71	75
1026.	invert	*v*		**15.82**	47	20.407	100
1027.	migrate	*v*		**15.76**	43	18.671	75
1028.	favorable	*adj*	*B2*	**15.73**	40	17.368	100
1029.	coordination	*n*		**15.70**	101	43.854	50
1030.	digit	*n*		**15.65**	53	23.013	100
1031.	sponge	*n*		**15.62**	100	43.42	50
1032.	moisture	*n*		**15.48**	49	21.276	100
1033.	attachment	*n*	*B2*	**15.40**	53	23.013	75
1034.	defense	*v*		**15.40**	86	37.341	50
1035.	indefinitely	*adv*	*C2*	**15.40**	32	13.894	100
1036.	macroscopic	*adj*		**15.32**	51	22.144	50
1037.	nest*	*n*	*C2*	**15.30**	83	36.038	100
1038.	slab	*n*		**15.23**	119	51.67	100
1039.	hybrid	*adj*		**15.21**	112	48.63	75
1040.	concentric	*adj*		**15.21**	44	19.105	100
1041.	urine	*n*		**15.19**	141	61.222	50
1042.	microbe*	*n*		**15.16**	119	51.67	50
1043.	incomplete	*adj*		**15.16**	44	19.105	100
1044.	intermolecular	*adj*		**15.16**	101	43.854	50
1045.	infinitely	*adv*	*C2*	**15.09**	54	23.447	75
1046.	lateral	*adj*		**15.04**	77	33.433	75
1047.	lightning	*n*	*B1*	**14.93**	40	17.368	75
1048.	ultraviolet	*n*		**14.91**	41	17.802	75
1049.	microwave	*n*		**14.87**	67	29.091	75
1050.	artery	*n*		**14.86**	110	47.762	75
1051.	athlete	*n*	*B1*	**14.86**	56	24.315	50
1052.	infectious	*adj*	*C2*	**14.82**	68	29.525	50
1053.	diffraction	*n*		**14.82**	184	79.892	100
1054.	antenna*	*n*		**14.80**	106	46.025	50
1055.	inertia*	*n*		**14.67**	221	95.958	75
1056.	byproduct	*n*		**14.63**	42	18.236	50
1057.	fertilizer	*n*		**14.60**	71	30.828	75
1058.	electrically	*adv*		**14.59**	38	16.5	75
1059.	graphically	*adv*		**14.53**	36	15.631	75
1060.	conduction	*n*		**14.44**	77	33.433	75
1061.	resistant	*n*		**14.40**	58	25.183	75

1062.	cross-section	*n*		**14.39**	70	30.394	100
1063.	donate	*v*	*B2*	**14.37**	53	23.013	75
1064.	cartesian	*adj*		**14.37**	103	44.722	75
1065.	defective	*adj*	*C2*	**14.33**	109	47.328	50
1066.	compartment	*n*		**14.32**	58	25.183	75
1067.	bundle	*n*	*C2*	**14.29**	60	26.052	75
1068.	flask	*n*		**14.27**	64	27.789	75
1069.	submerge	*v*		**14.27**	43	18.671	100
1070.	reactive	*adj*		**14.23**	64	27.789	75
1071.	modified	*v/adj*	*C1*	**14.17**	35	15.197	75
1072.	pesticide	*n*		**14.14**	47	20.407	50
1073.	fatty	*adj*	*C1*	**14.11**	79	34.302	50
1074.	snail	*n*		**14.08**	55	23.881	50
1075.	diagonal	*adj*		**14.08**	50	21.71	75
1076.	worm	*n*	*B2*	**14.07**	70	30.394	50
1077.	molarity	*n*		**13.97**	133	57.748	50
1078.	pulley*	*n*		**13.97**	119	51.67	50
1079.	digestion	*n*	*C1*	**13.92**	90	39.078	75
1080.	infected	*v/adj*	*C1*	**13.85**	81	35.17	50
1081.	muscular*	*adj*		**13.80**	69	29.96	75
1082.	deficiency	*n*	*C1*	**13.76**	58	25.183	100
1083.	bead*	*n*		**13.68**	55	23.881	100
1084.	polarize	*v*		**13.63**	108	46.893	50
1085.	equivalence	*n*		**13.61**	108	46.893	75
1086.	incoming	*v/n*		**13.58**	39	16.934	100
1087.	specialize	*v*	*B2*	**13.54**	56	24.315	50
1088.	completion	*n*	*C1*	**13.49**	82	35.604	50
1089.	starch	*n*		**13.48**	85	36.907	50
1090.	activation	*n*		**13.42**	86	37.341	50
1091.	radioactivity	*n*		**13.42**	45	19.539	100
1092.	hinge	*v/n*		**13.42**	62	26.92	75
1093.	electrolyte	*n*		**13.41**	108	46.893	75
1094.	inequality	*n*	*C2*	**13.40**	73	31.696	75
1095.	oppositely	*adv*		**13.34**	30	13.026	75
1096.	momentarily	*adv*		**13.32**	34	14.763	75
1097.	spider*	*n*	*B1*	**13.32**	73	31.696	50
1098.	neutralize	*v*		**13.29**	54	23.447	50
1099.	telescope*	*n*	*B2*	**13.19**	81	35.17	75
1100.	lining	*v/n*	*B2*	**13.18**	58	25.183	75
1101.	odor	*n*		**13.12**	51	22.144	75
1102.	lizard*	*n*		**13.05**	64	27.789	75
1103.	physiological*	*adj*		**13.04**	38	16.5	75
1104.	refrigerator	*n*		**13.02**	66	28.657	100
1105.	inorganic	*adj*		**13.00**	42	18.236	50
1106.	bee*	*n*	*B1*	**12.97**	86	37.341	75

1107.	corn*	*n*	*B1*	**12.94**	63	27.354	75
1108.	radially	*adv*		**12.78**	51	22.144	50
1109.	properties	*n*	*B1*	**12.77**	36	15.631	100
1110.	concave	*adj*		**12.77**	89	38.644	100
1111.	increment	*v/n*		**12.73**	36	15.631	100
1112.	whale*	*n*	*B1*	**12.71**	103	44.722	50
1113.	thrive	*v*	*C1*	**12.65**	52	22.578	50
1114.	unpaired	*adj*		**12.55**	88	38.209	75
1115.	symmetrical	*adj*		**12.53**	67	29.091	75
1116.	feedback	*n*	*B2*	**12.49**	107	46.459	50
1117.	complementary	*adj*		**12.45**	72	31.262	100
1118.	deviation	*n*		**12.42**	47	20.407	100
1119.	rational	*adj*	*C1*	**12.25**	107	46.459	50
1120.	farthest	*adj*	*A2*	**12.25**	33	14.329	100
1121.	repulsion	*n*		**12.24**	68	29.525	75
1122.	deflect	*v*		**12.22**	61	26.486	75
1123.	web	*n*	*A2*	**12.18**	56	24.315	50
1124.	bladder*	*n*		**12.15**	56	24.315	75
1125.	graphite	*n*		**12.14**	78	33.867	75
1126.	ingest	*v*		**12.09**	37	16.065	75
1127.	refraction	*n*		**12.03**	204	88.576	75
1128.	applied	*v/adj*	*B1*	**12.02**	59	25.618	50
1129.	optical	*adj*		**12.02**	84	36.473	75
1130.	calculated	*v*	*B2*	**11.99**	33	14.329	75
1131.	microorganism	*n*		**11.99**	55	23.881	50
1132.	exponentially	*adv*		**11.95**	41	17.802	100
1133.	foil	*n*		**11.93**	36	15.631	75
1134.	nonpolar	*adj*		**11.86**	71	30.828	75
1135.	excrete	*n*		**11.83**	56	24.315	50
1136.	implant	*v/n*		**11.83**	53	23.013	75
1137.	parasitic	*adj*		**11.72**	69	29.96	50
1138.	cardiac	*adj*		**11.66**	87	37.775	100
1139.	destructive	*adj*		**11.66**	54	23.447	75
1140.	elapse	*v*	*C2*	**11.64**	30	13.026	75
1141.	dimensionless	*adj*		**11.62**	31	13.46	50
1142.	condensed	*adj*		**11.58**	33	14.329	75
1143.	spinal	*adj*		**11.57**	91	39.512	50
1144.	generalization	*n*	*C1*	**11.51**	31	13.46	100
1145.	postulate	*v*		**11.47**	57	24.749	75
1146.	elimination	*n*	*C2*	**11.43**	69	29.96	75
1147.	colorless	*adj*		**11.42**	40	17.368	75
1148.	planar	*adj*		**11.41**	44	19.105	75
1149.	capacitance	*n*		**11.39**	211	91.616	50
1150.	vein*	*n*	*C1*	**11.33**	85	36.907	50
1151.	infrared	*n*		**11.29**	49	21.276	75

1152.	upright	*adj*	*B2*	**11.24**	69	29.96	75
1153.	latitude	*n*		**11.17**	39	16.934	75
1154.	buffer	*v/n*		**11.17**	177	76.853	50
1155.	translational	*adj*		**11.14**	64	27.789	50
1156.	elongate*	*v*		**11.11**	36	15.631	75
1157.	pregnancy	*n*	*C1*	**11.10**	63	27.354	75
1158.	elephant	*n*	*A2*	**11.01**	85	36.907	50
1159.	accelerator	*n*		**10.99**	37	16.065	75
1160.	cyclic	*adj*		**10.97**	49	21.276	75
1161.	multiplication	*n*		**10.96**	35	15.197	100
1162.	illuminate	*v*		**10.94**	54	23.447	75
1163.	shark*	*n*		**10.93**	52	22.578	50
1164.	saturate	*v*		**10.92**	38	16.5	75
1165.	homogeneous	*adj*		**10.88**	46	19.973	100
1166.	micrograph	*n*		**10.87**	42	18.236	50
1167.	watery	*adj*		**10.83**	45	19.539	50
1168.	rupture*	*v/n*		**10.83**	30	13.026	75
1169.	parabolic	*adj*		**10.78**	40	17.368	50
1170.	reactor	*n*		**10.75**	58	25.183	75
1171.	superposition	*n*		**10.66**	48	20.842	50
1172.	outermost	*adj*		**10.63**	37	16.065	100
1173.	elementary	*adj*	*B1*	**10.51**	52	22.578	75
1174.	buoyant	*adj*		**10.51**	63	27.354	100
1175.	conductivity	*n*		**10.48**	45	19.539	75
1176.	subunits	*n*		**10.45**	69	29.96	50
1177.	ellipse	*n*		**10.42**	121	52.538	50
1178.	nutrition*	*n*	*C1*	**10.40**	36	15.631	50
1179.	lightbulb	*n*		**10.34**	66	28.657	50
1180.	fetus*	*n*		**10.31**	62	26.92	50
1181.	endpoint	*n*		**10.28**	77	33.433	50
1182.	nucleic	*adj*		**10.28**	44	19.105	50
1183.	algebra	*n*		**10.27**	30	13.026	75
1184.	dissociate	*v*		**10.22**	51	22.144	50
1185.	continuity	*n*	*C2*	**10.22**	55	23.881	75
1186.	logarithmic	*adj*		**10.20**	47	20.407	100
1187.	magnification	*n*		**10.18**	79	34.302	50
1188.	endangered	*adj*	*B2*	**10.17**	45	19.539	50
1189.	prefix	*n*	*B2*	**10.17**	156	67.735	75
1190.	recycle	*v*	*B1*	**10.14**	40	17.368	75
1191.	arctic*	*adj*		**10.10**	49	21.276	75
1192.	pea*	*n*	*B1*	**10.09**	50	21.71	75
1193.	spacing*	*v/n*		**10.05**	34	14.763	75
1194.	semicircle*	*n*		**10.04**	45	19.539	50
1195.	predatory*	*adj*		**10.00**	41	17.802	50

D. LIST OF MULTI-WORD UNITS

Item	Freq.	Relative frequency	ARF	DOCF
1. time interval	464	201.46	121.37	4
2. kinetic energy	564	244.88	99.75	4
3. electric field	904	392.51	96.29	4
4. magnetic field	912	395.98	77.94	4
5. straight line	169	73.37	69.96	4
6. potential energy	389	168.90	58.30	3
7. chemical reaction	195	84.66	54.80	4
8. hydrogen atom	269	116.79	49.00	3
9. surface area	150	65.12	46.66	4
10. internal energy	269	116.79	45.26	3
11. maximum value	167	72.51	43.93	3
12. water molecule	211	91.61	40.63	3
13. rate of change	155	67.30	40.43	4
14. amino acid	269	116.79	38.33	2
15. force act	174	75.55	37.38	2
16. carbon atom	285	123.74	36.77	3
17. center of mass	302	131.12	36.00	2
18. gravitational force	227	98.56	33.55	2
19. positive charge	154	66.86	33.30	3
20. total energy	147	63.82	32.94	3
21. amount of energy	109	47.32	32.78	4
22. constant speed	123	53.40	32.56	4
23. blood cell	207	89.87	32.55	2
24. boiling point	171	74.24	32.52	4
25. negative sign	77	33.43	30.72	3
26. function of time	127	55.14	30.70	3
27. numerical value	62	26.92	29.32	3
28. high temperature	81	35.17	29.24	3
29. same direction	89	38.64	28.92	4
30. potential difference	317	137.64	28.28	2
31. nervous system	163	70.77	28.09	2
32. ideal gas	269	116.79	27.57	3
33. cross-sectional area	102	44.28	26.79	3
34. molar mass	160	69.47	26.24	2
35. negative value	57	24.74	26.05	3
36. particle move	111	48.19	25.24	3
37. chemical bond	88	38.20	24.82	3
38. side of the equation	58	25.18	24.69	4

39.	speed of light	121	52.53	24.65	3
40.	net force	127	55.14	24.08	3
41.	minimum value	92	39.94	23.79	3
42.	following statement	44	19.10	23.47	4
43.	covalent bond	183	79.45	23.45	2
44.	negative charge	121	52.53	23.44	3
45.	hydrogen bond	186	80.76	23.23	2
46.	circle of radius	66	28.65	23.07	2
47.	periodic table	156	67.73	23.00	3
48.	organic molecule	96	41.68	22.64	3
49.	energy transfer	132	57.31	22.62	3
50.	charge density	144	62.52	22.06	3
51.	initial value	76	32.99	22.05	3
52.	melting point	120	52.10	21.90	2
53.	angular speed	200	86.83	21.86	2
54.	oxygen atom	87	37.77	21.34	3
55.	blood vessel	103	44.72	21.23	4
56.	initial velocity	87	37.77	20.85	3
57.	number of electrons	77	33.43	20.67	3
58.	chemical equation	149	64.69	20.66	2
59.	cell wall	126	54.70	20.52	2
60.	sphere of radius	65	28.22	20.50	3
61.	significant figure	130	56.44	20.44	2
62.	red blood cell	114	49.49	20.43	2
63.	tangent line	119	51.66	20.36	3
64.	average value	104	45.15	20.02	4
65.	conservation of energy	55	23.88	19.81	4
66.	electric charge	73	31.69	19.59	2
67.	immune system	112	48.63	19.57	2
68.	sound wave	153	66.43	19.53	4
69.	specific heat	190	82.49	19.49	4
70.	atomic mass	170	73.81	19.31	3
71.	positive value	49	21.27	18.97	3
72.	mechanical energy	122	52.97	18.93	2
73.	atomic number	126	54.70	18.80	3
74.	direction of motion	50	21.70	18.75	3
75.	line segment	103	44.72	18.66	3
76.	temperature increase	51	22.14	18.51	4
77.	unit area	49	21.27	18.49	4
78.	energy change	67	29.09	18.36	3
79.	differential equation	114	49.49	18.31	3
80.	absolute value	56	24.31	18.28	3
81.	liquid water	65	28.22	18.26	3
82.	electromagnetic wave	127	55.14	18.23	3
83.	organic compound	94	40.81	18.23	2

84. total mass	57	24.74	18.14	4
85. isolated system	86	37.34	18.03	3
86. visible light	46	19.97	17.97	3
87. number of atoms	73	31.69	17.56	2
88. velocity vector	102	44.28	17.48	2
89. air resistance	75	32.56	17.42	2
90. use datum	56	24.31	17.28	3
91. atmospheric pressure	83	36.03	17.24	4
92. initial speed	105	45.59	17.15	2
93. body temperature	75	32.56	16.94	3
94. constant value	38	16.49	16.91	3
95. average speed	71	30.82	16.82	4
96. unit volume	46	19.97	16.59	4
97. unit length	79	34.30	16.48	3
98. charged particle	103	44.72	16.43	4
99. coordinate system	53	23.01	16.39	3
100. temperature change	71	30.82	16.16	4
101. law of conservation	46	19.97	16.10	4
102. electric current	51	22.14	16.09	3
103. constant rate	43	18.67	16.08	4
104. oxygen gas	50	21.70	15.98	3
105. cell membrane	67	29.09	15.73	4
106. equilibrium position	81	35.17	15.72	2
107. constant acceleration	67	29.09	15.68	2
108. circular orbit	67	29.09	15.49	3
109. much work	81	35.17	15.42	3
110. external force	86	37.34	15.40	2
111. dashed line	38	16.49	15.22	3
112. vertical line	53	23.01	15.14	3
113. water vapor	48	20.84	15.01	3
114. forces act	90	39.07	14.94	3
115. object move	57	24.74	14.93	2
116. continuous function	111	48.19	14.90	2
117. law of thermodynamics	84	36.47	14.57	3
118. back of the book	43	18.67	14.50	3
119. light source	51	22.14	14.46	4
120. maximum height	57	24.74	14.40	3
121. constant velocity	59	25.61	14.37	2
122. key concept	42	18.23	14.31	2
123. number of moles	78	33.86	14.23	2
124. angular momentum	173	75.11	14.09	3
125. blue line	31	13.46	14.08	3
126. right triangle	41	17.80	14.02	2
127. electric force	101	43.85	13.89	2
128. cross section	50	21.70	13.85	3

129. circular path	82	35.60	13.78	2
130. wave function	175	75.98	13.76	2
131. electric potential	146	63.39	13.74	3
132. partial pressure	94	40.81	13.73	2
133. gas molecule	55	23.88	13.55	2
134. solution contain	51	22.14	13.51	3
135. population growth	98	42.55	13.37	2
136. electromagnetic radiation	56	24.31	13.36	3
137. immune response	108	46.89	13.33	2
138. overall reaction	71	30.82	13.19	2
139. net charge	64	27.78	13.18	3
140. maximum speed	49	21.27	13.17	2
141. third law	49	21.27	13.15	3
142. unit vector	74	32.13	13.03	2
143. plane perpendicular	26	11.28	12.97	3
144. transfer of energy	46	19.97	12.92	3
145. molecular mass	70	30.39	12.88	2
146. energy level	116	50.36	12.87	3
147. horizontal line	43	18.67	12.81	3
148. cell type	59	25.61	12.79	2
149. white blood cell	61	26.48	12.75	2
150. initial condition	49	21.27	12.45	3
151. nerve cell	52	22.57	12.41	2
152. vector sum	48	20.84	12.37	2
153. sodium chloride	46	19.97	12.34	3
154. different color	25	10.85	12.21	4
155. first quadrant	78	33.86	12.20	2
156. constant temperature	48	20.84	12.16	4
157. electron density	111	48.19	12.12	2
158. gas pressure	72	31.26	12.00	2
159. double bond	104	45.15	11.97	2
160. instant of time	32	13.89	11.94	3
161. sulfuric acid	48	20.84	11.87	3
162. magnitude of the force	40	17.36	11.80	3
163. given value	31	13.46	11.79	3
164. heat capacity	128	55.57	11.78	3
165. unit time	27	11.72	11.76	4
166. radioactive decay	66	28.65	11.74	4
167. small intestine	90	39.07	11.73	2
168. charge distribution	86	37.34	11.64	3
169. natural logarithm	40	17.36	11.62	3
170. intermolecular force	79	34.30	11.49	2
171. hydrogen ion	43	18.67	11.46	2
172. horizontal surface	59	25.61	11.41	3
173. vapor pressure	157	68.16	11.37	2

174. coordinate axe	60	26.05	11.32	2
175. total force	39	16.93	11.29	3
176. moment of inertia	204	88.57	11.27	2
177. nitrogen atom	37	16.06	11.27	3
178. definite integral	95	41.24	11.18	2
179. functional group	89	38.64	11.13	2
180. metal atom	53	23.01	10.93	2
181. physical property	44	19.10	10.84	3
182. molecular formula	74	32.13	10.80	2
183. gravitational field	48	20.84	10.72	2
184. positive number	36	15.63	10.69	2
185. exponential function	87	37.77	10.40	3
186. conversion factor	58	25.18	10.40	2
187. position vector	68	29.52	10.34	2
188. strong acid	110	47.76	10.33	2
189. human population	54	23.44	10.30	2
190. constant pressure	83	36.03	10.28	2
191. light ray	98	42.55	10.25	3
192. acetic acid	62	26.92	10.21	2
193. light beam	46	19.97	10.15	2
194. life span	38	16.49	10.15	2
195. side of equation	29	12.59	10.13	3
196. circular motion	49	21.27	10.06	2
197. outer radius	26	11.28	9.96	2
198. reference frame	72	31.26	9.92	2
199. horizontal component	36	15.63	9.87	2
200. electric circuit	35	15.19	9.84	3
201. charged ion	29	12.59	9.80	
202. particle of mass m	34	14.76	9.79	3
203. hydrogen gas	32	13.89	9.78	3
204. sphere of radius r	26	11.28	9.69	2
205. smooth curve	64	27.78	9.66	3
206. magnetic force	145	62.95	9.63	2
207. energy of a system	45	19.53	9.62	2
208. plane of the page	30	13.02	9.46	2
209. total work	28	12.15	9.44	3
210. chemical property	38	16.49	9.37	2
211. net change	45	19.53	9.34	3
212. flowering plant	49	21.27	9.27	2
213. center of the circle	30	13.02	9.26	2
214. polar coordinate	93	40.38	9.22	3
215. average velocity	76	32.99	9.11	2
216. copper wire	34	14.76	9.08	4
217. biological molecule	56	24.31	9.04	2
218. constant volume	43	18.67	8.96	3

219. harmonic motion	106	46.02	8.94	2
220. chemical energy	49	21.27	8.93	3
221. basic solution	50	21.70	8.92	2
222. speed of sound	71	30.82	8.92	2
223. critical point	101	43.85	8.91	3
224. opposite sign	38	16.49	8.84	3
225. same mass	27	11.72	8.83	2
226. simple harmonic motion	103	44.72	8.82	2
227. total pressure	35	15.19	8.78	2
228. state of matter	28	12.15	8.73	2
229. radius of curvature	57	24.74	8.59	2
230. pure water	43	18.67	8.58	2
231. trigonometric function	68	29.52	8.55	3
232. interior point	46	19.97	8.52	2
233. same temperature	31	13.46	8.42	4
234. extreme value	102	44.28	8.38	3
235. total charge	57	24.74	8.38	3
236. gravitational potential energy	57	24.74	8.34	2
237. force of attraction	36	15.63	8.30	3
238. upper bound	42	18.23	8.28	3
239. end of the rod	27	11.72	8.21	2
240. time accord	31	13.46	8.20	2
241. vertical component	26	11.28	8.15	2
242. region of space	27	11.72	8.13	3
243. spinal cord	70	30.39	8.12	2
244. potential energy of the system	47	20.40	8.12	2
245. number of molecules	48	20.84	8.11	2
246. chemical change	31	13.46	8.10	2
247. wavelength of light	29	12.59	8.09	3
248. double helix	83	36.03	8.09	2
249. activation energy	72	31.26	8.08	2
250. decimal place	41	17.80	8.05	3
251. vertical plane	31	13.46	8.04	2
252. overall equation	43	18.67	8	2
253. electric field	47	20.40	7.98	2
254. ion concentration	30	13.02	7.96	3
255. unpaired electron	72	31.26	7.94	3
256. solid sphere	40	17.36	7.92	2
257. rate of energy	38	16.49	7.90	2
258. dipole moment	82	35.60	7.87	2
259. chemical formula	28	12.15	7.86	3
260. disk of radius	24	10.42	7.85	2
261. arc length	47	20.40	7.75	2
262. point charge	142	61.65	7.74	2
263. ionic bond	35	15.19	7.73	2

264. reverse reaction	38	16.49	7.62	2
265. muscle contraction	35	15.19	7.62	2
266. freezing point	51	22.14	7.50	2
267. scalar quantity	34	14.76	7.42	2
268. frame of reference	51	22.14	7.42	2
269. spherical shell	34	14.76	7.42	3
270. gas exchange	60	26.05	7.39	2
271. chlorine atom	28	12.15	7.39	2
272. rate of reaction	111	48.19	7.38	3
273. light wave	46	19.97	7.37	3
274. buoyant force	57	24.74	7.36	3
275. mass density	30	13.02	7.31	2
276. change in temperature	35	15.19	7.21	3
277. polar molecule	42	18.23	7.19	3
278. cartesian coordinate	41	17.80	7.17	3
279. constant force	26	11.28	7.16	2
280. diatomic molecule	30	13.02	7.14	2
281. second derivative	41	17.80	7.14	2
282. phase change	67	29.09	7.07	2
283. electrical signal	43	18.67	7	2
284. light intensity	48	20.84	6.92	3
285. limiting value	29	12.59	6.89	2
286. partial derivative	98	42.55	6.82	2
287. entropy change	93	40.38	6.78	2
288. final state	48	20.84	6.72	2
289. initial temperature	32	13.89	6.68	3
290. population size	51	22.14	6.62	2
291. nuclear reaction	39	16.93	6.62	3
292. falling object	27	11.72	6.52	2
293. exponential growth	45	19.53	6.52	2
294. billion year	26	11.28	6.51	3
295. dot product	33	14.32	6.47	2
296. free electron	30	13.02	6.45	2
297. ground state	41	17.80	6.43	2
298. resulting solution	31	13.46	6.41	2
299. carboxylic acid	58	25.18	6.40	2
300. active site	45	19.53	6.40	2
301. equal magnitude	28	12.15	6.36	2
302. axis of rotation	45	19.53	6.36	2
303. final temperature	45	19.53	6.34	2
304. first law of thermodynamics	30	13.02	6.29	3
305. vector quantity	36	15.63	6.26	2
306. value of the function	30	13.02	6.25	2
307. hydroxyl group	32	13.89	6.21	2
308. mass number	61	26.48	6.15	3

309. single bond	30	13.02	6.15	2
310. rotational motion	31	13.46	5.97	2
311. carrying capacity	57	24.74	5.91	2
312. heat of combustion	30	13.02	5.77	2
313. electron shell	31	13.46	5.66	2
314. gas law	33	14.32	5.55	3
315. mechanical wave	24	10.42	5.45	2
316. mole of gas	31	13.46	5.45	2
317. mean value theorem	47	20.40	5.44	2
318. point of water	28	12.15	5.33	3
319. mass of the system	28	12.15	5.32	2
320. transverse wave	40	17.36	5.27	2
321. base pair	57	24.74	5.24	2
322. radioactive nucleus	25	10.85	5.20	4
323. cross product	45	19.53	5.12	2
324. osmotic pressure	43	18.67	5.11	2
325. quantum number	82	35.60	4.86	2
326. reversible process	34	14.76	4.83	2
327. standing wave	93	40.38	4.81	2
328. final speed	26	11.28	4.81	2
329. wave front	39	16.93	4.78	2
330. theorem of calculus	28	12.15	4.70	2
331. time graph	48	20.84	4.69	2
332. diffraction pattern	77	33.43	4.66	3
333. methyl group	53	23.01	4.66	2
334. instantaneous rate	29	12.59	4.60	3
335. interference pattern	73	31.69	4.58	2
336. thin filament	38	16.49	4.56	2
337. angle of incidence	49	21.27	4.53	2
338. line integral	39	16.93	4.52	2
339. wave equation	26	11.28	4.50	3
340. number of protons	28	12.15	4.50	3
341. equilibrium condition	35	15.19	4.38	2
342. closed surface	49	21.27	4.38	2
343. change in internal energy	35	15.19	4.38	2
344. light travel	28	12.15	4.32	3
345. excited state	32	13.89	4.30	2
346. right-hand rule	28	12.15	4.25	2
347. barometric pressure	31	13.46	4.22	2
348. internal resistance	50	21.70	4.20	2
349. magnetic moment	81	35.17	4.19	2
350. magnitude of the magnetic field	38	16.49	4.18	2
351. instantaneous velocity	41	17.80	4.16	2
352. acceleration vector	38	16.49	4.09	2
353. triple point	26	11.28	3.95	2

354. number of significant figures	31	13.46	3.94	2
355. focal length	113	49.06	3.86	2
356. second law of thermodynamics	36	15.63	3.78	3
357. molecular shape	25	10.85	3.78	2
358. traveling wave	29	12.59	3.75	2
359. probability density	63	27.35	3.75	
360. point of inflection	31	13.46	3.75	2
361. adiabatic process	30	13.02	3.72	2
362. destructive interference	36	15.63	3.68	2
363. constructive interference	40	17.36	3.67	2
364. light of wavelength	29	12.59	3.55	2
365. thin plate	29	12.59	3.49	2
366. tuning fork	30	13.02	3.34	2
367. heat engine	66	28.65	3.29	2
368. equilibrium state	30	13.02	3.24	2
369. resultant wave	28	12.15	3.24	2
370. transformation equation	31	13.46	3.23	2
371. electrolytic cell	25	10.85	3.17	2
372. direction of the magnetic field	26	11.28	3.12	2
373. center of gravity	31	13.46	2.96	2
374. unit cell	136	59.05	2.92	2
375. change in entropy	44	19.10	2.73	2
376. radial node	27	11.72	2.70	2
377. number of microstates	36	15.63	2.66	2
378. elliptical orbit	26	11.28	2.38	2
379. power series	31	13.46	2.09	2

E. QUESTIONNAIRE (1) ON KEYWORDS

Please give your answer to the question:

To what extent is the word useful for your students in the science courses?

by choosing a number from 1 to 5 in the degree of usefulness column (1 is the LEAST useful and 5 is the MOST useful.

Item number	**Headword**	**Degree of Usefulness**				
		5	4	3	2	1
1.	equation					
2.	value					
3.	energy					
4.	result					
5.	produce					
6.	function					
7.	increase					
8.	constant					
9.	system					
10.	cell					
11.	determine					
12.	mass					
13.	force					
14.	occur					
15.	solution					
16.	contain					
17.	molecule					
18.	unit					
19.	surface					
20.	section					
21.	consider					
22.	cause					
23.	equal					
24.	reaction					
25.	speed					
26.	require					
27.	assume					
28.	base					
29.	direction					
30.	obtain					

31.	process					
32.	calculate					
33.	object					
34.	length					
35.	represent					
36.	distance					
37.	apply					
38.	rate					
39.	charge					
40.	state					
41.	remain					
42.	measure					
43.	positive					
44.	structure					
45.	depend					
46.	amount					
47.	earth					
48.	particle					
49.	position					
50.	define					
51.	condition					
52.	reach					
53.	allow					
54.	consist					
55.	chemical					
56.	curve					
57.	decrease					
58.	region					
59.	involve					
60.	volume					
61.	expression					
62.	center					
63.	graph					
64.	magnitude					
65.	motion					
66.	product					
67.	similar					
68.	law					
69.	reduce					
70.	ion					
71.	suppose					
72.	method					
73.	pressure					
74.	compare					
75.	radius					
76.	release					
77.	potential					

78.	interval					
79.	act					
80.	quantity					
81.	angle					
82.	initial					
83.	source					
84.	average					
85.	common					
86.	current					
87.	illustrate					
88.	density					
89.	approach					
90.	bond					
91.	solid					
92.	factor					
93.	velocity					
94.	human					
95.	material					
96.	component					
97.	certain					
98.	property					
99.	heat					
100.	indicate					
101.	express					
102.	relate					
103.	power					
104.	axis					
105.	substance					
106.	shape					
107.	situation					
108.	separate					
109.	liquid					
110.	step					
111.	relative					
112.	sum					
113.	solve					
114.	natural					
115.	relationship					
116.	formula					
117.	equilibrium					
118.	compound					
119.	vary					
120.	datum					
121.	horizontal					
122.	diagram					
123.	estimate					
124.	exist					

125.	direct					
126.	convert					
127.	vector					
128.	differ					
129.	vertical					
130.	specific					
131.	due					
132.	wave					
133.	directly					
134.	identify					
135.	divide					
136.	locate					
137.	replace					
138.	evaluate					
139.	connect					
140.	color					
141.	refer					
142.	acceleration					
143.	series					
144.	behavior					
145.	combine					
146.	observe					
147.	calculation					
148.	concept					
149.	remove					
150.	origin					
151.	rule					
152.	original					
153.	flow					
154.	coordinate					
155.	parallel					
156.	experiment					
157.	height					
158.	fix					
159.	attach					
160.	addition					
161.	derive					
162.	concentration					
163.	sample					
164.	associate					
165.	drop					
166.	complex					
167.	principle					
168.	exert					
169.	sphere					
170.	ratio					
171.	molecular					

172.	combination					
173.	variable					
174.	store					
175.	focus					
176.	generate					
177.	physical					
178.	maintain					
179.	internal					
180.	active					
181.	location					
182.	block					
183.	appropriate					
184.	additional					
185.	organism					
186.	theory					
187.	definition					
188.	species					
189.	perpendicular					
190.	population					
191.	wire					
192.	generally					
193.	surround					
194.	predict					
195.	rapidly					
196.	perform					
197.	portion					
198.	frequency					
199.	identical					
200.	multiple					
201.	cycle					
202.	resistance					
203.	central					
204.	typical					
205.	formation					
206.	respect					
207.	kinetic					
208.	weight					
209.	relatively					
210.	analyze					
211.	undergo					
212.	basic					
213.	circular					
214.	continuous					
215.	nucleus					
216.	substitute					
217.	approximately					
218.	arise					

219.	atomic					
220.	plot					
221.	mole					
222.	slightly					
223.	integral					

F. QUESTIONNAIRE (2) ON KEYWORDS

Please give your answer to the question:

To what extent is the word useful for your students in the science courses?

by choosing a number from 1 to 5 in the degree of usefulness column (1 is the LEAST useful and 5 is the MOST useful.

Item number	Headword	Degree of Usefulness				
		5	4	3	2	1
1.	environment					
2.	meter					
3.	completely					
4.	gene					
5.	derivative					
6.	fluid					
7.	magnetic					
8.	correspond					
9.	cylinder					
10.	entire					
11.	imagine					
12.	extend					
13.	upper					
14.	presence					
15.	phase					
16.	muscle					
17.	mixture					
18.	significant					
19.	proportional					
20.	typically					
21.	equivalent					
22.	absorb					
23.	characteristic					
24.	layer					
25.	reverse					
26.	conclude					
27.	prevent					
28.	linear					
29.	corresponding					
30.	membrane					
31.	balance					
32.	contribute					

33.	diameter					
34.	growth					
35.	external					
36.	ability					
37.	pattern					
38.	construct					
39.	tube					
40.	initially					
41.	edge					
42.	bacterium					
43.	dissolve					
44.	observation					
45.	shell					
46.	upward					
47.	electrical					
48.	sketch					
49.	reflect					
50.	distribution					
51.	scale					
52.	outer					
53.	device					
54.	root					
55.	rotate					
56.	rod					
57.	enzyme					
58.	symbol					
59.	yield					
60.	explore					
61.	multiply					
62.	angular					
63.	fraction					
64.	theorem					
65.	approximate					
66.	encounter					
67.	overall					
68.	respectively					
69.	image					
70.	circuit					
71.	recall					
72.	iron					
73.	slope					
74.	response					
75.	atmosphere					
76.	mechanism					
77.	tissue					
78.	sodium					
79.	orbital					

80.	segment					
81.	measurement					
82.	sequence					
83.	displacement					
84.	interaction					
85.	simplify					
86.	scientist					
87.	weak					
88.	consistent					
89.	evolve					
90.	phenomenon					
91.	bind					
92.	disease					
93.	expand					
94.	extremely					
95.	nutrient					
96.	commonly					
97.	ray					
98.	label					
99.	string					
100.	gravitational					
101.	mechanical					
102.	stable					
103.	fuel					
104.	signal					
105.	recognize					
106.	transport					
107.	downward					
108.	rock					
109.	radiation					
110.	principal					
111.	detect					
112.	fundamental					
113.	arrow					
114.	variation					
115.	pure					
116.	ocean					
117.	approximation					
118.	summarize					
119.	excess					
120.	instant					
121.	tangent					
122.	dimension					
123.	unknown					
124.	smooth					
125.	assumption					
126.	experimental					

127.	similarly					
128.	beam					
129.	visible					
130.	agent					
131.	primary					
132.	enclose					
133.	configuration					
134.	friction					
135.	satisfy					
136.	conduct					
137.	hole					
138.	distribute					
139.	copper					
140.	exceed					
141.	simultaneously					
142.	skin					
143.	extreme					
144.	synthesize					
145.	eliminate					
146.	wavelength					
147.	polar					
148.	displace					
149.	valid					
150.	hint					
151.	branch					
152.	deliver					
153.	assign					
154.	specify					
155.	shift					
156.	inner					
157.	vessel					
158.	absolute					
159.	interact					
160.	accelerate					
161.	partial					
162.	attract					
163.	distinguish					
164.	react					
165.	medium					
166.	nuclear					
167.	deter					
168.	spherical					
169.	synthesis					
170.	reactant					
171.	separation					
172.	genetic					
173.	respond					

174.	coefficient					
175.	leaf					
176.	isolate					
177.	thick					
178.	burn					
179.	stimulate					
180.	male					
181.	voltage					
182.	verify					
183.	consequently					
184.	laboratory					
185.	readily					
186.	depth					
187.	notation					
188.	compose					
189.	bound					
190.	conversion					
191.	occupy					
192.	rotation					
193.	numerical					
194.	ionic					
195.	mate					
196.	female					
197.	percentage					
198.	exhibit					
199.	loop					
200.	vapor					
201.	conservation					
202.	division					
203.	mathematical					
204.	boundary					
205.	gravity					
206.	triangle					
207.	planet					
208.	error					
209.	intermediate					
210.	essentially					
211.	capacity					
212.	stretch					
213.	composition					
214.	evolution					
215.	interior					
216.	domain					
217.	precisely					
218.	emit					
219.	molar					
220.	width					

G. QUESTIONNAIRE (3) ON KEYWORDS

Please give your answer to the question:

To what extent is the word useful for your students in the science courses?

by choosing a number from 1 to 5 in the degree of usefulness column (1 is the LEAST useful and 5 is the MOST useful.

Item number	**Headword**	**Degree of Usefulness**				
		5	4	3	2	1
1.	overlap					
2.	electronic					
3.	bubble					
4.	cavity					
5.	exponential					
6.	metallic					
7.	proof					
8.	cross-sectional					
9.	torque					
10.	obey					
11.	activate					
12.	tween					
13.	triple					
14.	clockwise					
15.	fuse					
16.	toxic					
17.	parabola					
18.	definite					
19.	strip					
20.	orient					
21.	pipe					
22.	emission					
23.	dipole					
24.	fossil					
25.	neutron					
26.	frictionless					
27.	collide					
28.	reversible					
29.	pond					
30.	ionize					
31.	oscillation					
32.	attain					
33.	span					

34.	intestine					
35.	intersection					
36.	leak					
37.	flux					
38.	spontaneously					
39.	spectrum					
40.	diffusion					
41.	hypothetical					
42.	hollow					
43.	accompanying					
44.	kidney					
45.	melting					
46.	positively					
47.	intersect					
48.	remainder					
49.	airplane					
50.	disperse					
51.	fragment					
52.	filament					
53.	denominator					
54.	biologist					
55.	physicist					
56.	converge					
57.	metric					
58.	seawater					
59.	probability					
60.	elastic					
61.	spin					
62.	thermodynamics					
63.	lifetime					
64.	seal					
65.	respiratory					
66.	incorrect					
67.	laser					
68.	penetrate					
69.	capillary					
70.	rearrange					
71.	index					
72.	crop					
73.	spontaneous					
74.	projectile					
75.	bulk					
76.	steam					
77.	experimentally					
78.	metabolic					
79.	absorption					
80.	compression					

81.	numerator					
82.	drift					
83.	conceptualize					
84.	vibrate					
85.	harmful					
86.	float					
87.	grain					
88.	vibration					
89.	barrier					
90.	graphical					
91.	scalar					
92.	inject					
93.	synthetic					
94.	unstable					
95.	oxidize					
96.	bacterial					
97.	digest					
98.	satellite					
99.	compact					
100.	infect					
101.	catalyze					
102.	expel					
103.	skeleton					
104.	subtract					
105.	trigger					
106.	observed					
107.	deduce					
108.	detector					
109.	stability					
110.	proportionality					
111.	balanced					
112.	elevation					
113.	dilute					
114.	symmetric					
115.	reflection					
116.	discharge					
117.	logarithm					
118.	empirical					
119.	hydrocarbon					
120.	dominant					
121.	subscript					
122.	collectively					
123.	quantum					
124.	plausible					
125.	harmonic					
126.	reservoir					
127.	adaptation					

128.	decompose					
129.	resistor					
130.	independently					
131.	carrier					
132.	gradient					
133.	pendulum					
134.	adapt					
135.	immerse					
136.	evaporate					
137.	conversely					
138.	stimulus					
139.	electrostatic					
140.	catalyst					
141.	isotope					
142.	shrink					
143.	neuron					
144.	circumference					
145.	freezing					
146.	interference					
147.	quantitative					
148.	mathematically					
149.	visualize					
150.	marine					
151.	shaded					
152.	pollen					
153.	evaporation					
154.	backward					
155.	calculator					
156.	rope					
157.	fractional					
158.	curvature					
159.	coating					
160.	equator					
161.	static					
162.	inhibit					
163.	revolve					
164.	trigonometric					
165.	decomposition					
166.	required					
167.	interactive					
168.	resonance					
169.	concentrated					
170.	spacecraft					
171.	brake					
172.	terrestrial					
173.	node					
174.	entropy					

175.	incidence					
176.	moist					
177.	inversely					
178.	abundance					
179.	radian					
180.	precipitate					
181.	frog					
182.	inherit					
183.	arbitrarily					
184.	metabolism					
185.	specified					
186.	aquatic					
187.	tropical					
188.	protective					
189.	microscope					
190.	derivation					
191.	bounce					
192.	magnet					
193.	liver					
194.	rotational					
195.	inward					
196.	junction					
197.	precision					
198.	embed					
199.	truck					
200.	composite					
201.	qualitative					
202.	donor					
203.	thermodynamic					
204.	decimal					
205.	quotient					
206.	partition					
207.	skeletal					
208.	dimensional					
209.	align					
210.	breakdown					
211.	randomly					
212.	violet					
213.	condense					
214.	surrounding					
215.	repel					
216.	binding					
217.	rewrite					
218.	bloodstream					
219.	astronaut					
220.	valve					

H. QUESTIONNAIRE (4) ON KEYWORDS

Please give your answer to the question:

To what extent is the word useful for your students in the science courses?

by choosing a number from 1 to 5 in the degree of usefulness column (1 is the LEAST useful and 5 is the MOST useful.

Item number	**Headword**	**Degree of Usefulness**				
		5	4	3	2	1
1.	proceed					
2.	consume					
3.	slide					
4.	cube					
5.	bone					
6.	aqueous					
7.	capture					
8.	precise					
9.	pump					
10.	conductor					
11.	sunlight					
12.	modify					
13.	destroy					
14.	container					
15.	oxide					
16.	collision					
17.	trace					
18.	favor					
19.	transform					
20.	percent					
21.	terminal					
22.	atmospheric					
23.	decay					
24.	resemble					
25.	storage					
26.	manufacture					
27.	seed					
28.	compute					
29.	stem					
30.	possess					
31.	oxidation					
32.	orbit					
33.	biological					

34.	suspend					
35.	finite					
36.	momentum					
37.	switch					
38.	integrate					
39.	cylindrical					
40.	periodic					
41.	differentiate					
42.	scientific					
43.	freely					
44.	coil					
45.	reproduce					
46.	geometry					
47.	chromosome					
48.	steel					
49.	behave					
50.	hypothesis					
51.	denote					
52.	regardless					
53.	column					
54.	continuously					
55.	electromagnetic					
56.	acidic					
57.	transmit					
58.	reduction					
59.	summary					
60.	critical					
61.	transition					
62.	substitution					
63.	discovery					
64.	selection					
65.	soil					
66.	tank					
67.	primarily					
68.	intensity					
69.	cation					
70.	prediction					
71.	cord					
72.	convenient					
73.	pole					
74.	roughly					
75.	adjust					
76.	arbitrary					
77.	cancel					
78.	rapid					
79.	cellular					
80.	tension					

81.	induce					
82.	combustion					
83.	rectangle					
84.	tail					
85.	abundant					
86.	tract					
87.	adjacent					
88.	structural					
89.	weigh					
90.	extract					
91.	isolated					
92.	rectangular					
93.	regulate					
94.	thickness					
95.	ionization					
96.	sufficiently					
97.	bend					
98.	resultant					
99.	arc					
100.	vertically					
101.	partially					
102.	reproduction					
103.	exact					
104.	geometric					
105.	outward					
106.	expose					
107.	researcher					
108.	tendency					
109.	axe					
110.	negligible					
111.	predator					
112.	functional					
113.	infinite					
114.	melt					
115.	climate					
116.	frame					
117.	crystal					
118.	orientation					
119.	radial					
120.	observer					
121.	differential					
122.	receptor					
123.	diversity					
124.	characterize					
125.	amplitude					
126.	stationary					
127.	secondary					

128.	pathway					
129.	symmetry					
130.	uniformly					
131.	evolutionary					
132.	prey					
133.	rocket					
134.	inverse					
135.	apparatus					
136.	insert					
137.	classify					
138.	dash					
139.	surroundings					
140.	conserve					
141.	alternate					
142.	integration					
143.	nerve					
144.	tip					
145.	neglect					
146.	attraction					
147.	host					
148.	input					
149.	fiber					
150.	radioactive					
151.	solar					
152.	thermal					
153.	glucose					
154.	depict					
155.	revolution					
156.	integer					
157.	expansion					
158.	categorize					
159.	horizontally					
160.	altitude					
161.	pulse					
162.	transmission					
163.	strand					
164.	reasoning					
165.	lung					
166.	analogous					
167.	complicated					
168.	microscopic					
169.	vacuum					
170.	accomplish					
171.	solvent					
172.	transformation					
173.	gaseous					
174.	preceding					

175.	array					
176.	immune					
177.	generation					
178.	cubic					
179.	specialized					
180.	chamber					
181.	rigid					
182.	algebraic					
183.	instantaneous					
184.	dot					
185.	counterclockwise					
186.	unchanged					
187.	contraction					
188.	dense					
189.	cable					
190.	similarity					
191.	random					
192.	capacitor					
193.	gland					
194.	compress					
195.	covalent					
196.	core					
197.	filter					
198.	neutral					
199.	soluble					
200.	parameter					
201.	distant					
202.	nervous					
203.	cone					
204.	diverse					
205.	infection					
206.	diffuse					
207.	originate					
208.	insulate					
209.	steady					
210.	disorder					
211.	efficiency					
212.	solute					
213.	bulb					
214.	centimeter					
215.	accumulate					
216.	fusion					
217.	projection					
218.	cluster					
219.	photosynthesis					
220.	oscillate					

I. QUESTIONNAIRE (5) ON KEYWORDS

Please give your answer to the question:

To what extent is the word useful for your students in the science courses?

by choosing a number from 1 to 5 in the degree of usefulness column (1 is the LEAST useful and 5 is the MOST useful.

Item number	**Headword**	**Degree of Usefulness**				
		5	4	3	2	1
1.	interstitial					
2.	incident					
3.	infinitesimal					
4.	invade					
5.	hemisphere					
6.	trajectory					
7.	freshwater					
8.	indicator					
9.	diverge					
10.	spiral					
11.	encode					
12.	generalize					
13.	physician					
14.	conceptual					
15.	stabilize					
16.	saturated					
17.	quadratic					
18.	differentiation					
19.	mold					
20.	feather					
21.	regenerate					
22.	schematic					
23.	absent					
24.	pore					
25.	bullet					
26.	numerically					
27.	insoluble					
28.	violate					
29.	signify					
30.	physiology					
31.	radiate					
32.	respiration					
33.	droplet					

34.	solubility					
35.	discrete					
36.	midpoint					
37.	harmless					
38.	liberate					
39.	disrupt					
40.	maximize					
41.	biochemical					
42.	transparent					
43.	binary					
44.	wedge					
45.	linearly					
46.	impulse					
47.	precipitation					
48.	triangular					
49.	helix					
50.	quadrant					
51.	generator					
52.	reciprocal					
53.	eject					
54.	invert					
55.	migrate					
56.	favorable					
57.	coordination					
58.	digit					
59.	sponge					
60.	moisture					
61.	attachment					
62.	defense					
63.	indefinitely					
64.	macroscopic					
65.	nest					
66.	slab					
67.	hybrid					
68.	concentric					
69.	urine					
70.	microbe					
71.	incomplete					
72.	intermolecular					
73.	infinitely					
74.	lateral					
75.	lightning					
76.	ultraviolet					
77.	microwave					
78.	artery					
79.	athlete					
80.	infectious					

81.	diffraction					
82.	antenna					
83.	inertia					
84.	byproduct					
85.	fertilizer					
86.	electrically					
87.	graphically					
88.	conduction					
89.	resistant					
90.	cross-section					
91.	donate					
92.	cartesian					
93.	defective					
94.	compartment					
95.	bundle					
96.	flask					
97.	submerge					
98.	reactive					
99.	modified					
100.	pesticide					
101.	fatty					
102.	snail					
103.	diagonal					
104.	worm					
105.	molarity					
106.	pulley					
107.	digestion					
108.	infected					
109.	muscular					
110.	deficiency					
111.	bead					
112.	polarize					
113.	equivalence					
114.	incoming					
115.	specialize					
116.	completion					
117.	starch					
118.	activation					
119.	radioactivity					
120.	hinge					
121.	electrolyte					
122.	inequality					
123.	oppositely					
124.	momentarily					
125.	spider					
126.	neutralize					
127.	telescope					

128.	lining					
129.	odor					
130.	lizard					
131.	physiological					
132.	refrigerator					
133.	inorganic					
134.	bee					
135.	corn					
136.	radially					
137.	properties					
138.	concave					
139.	increment					
140.	whale					
141.	thrive					
142.	unpaired					
143.	symmetrical					
144.	feedback					
145.	complementary					
146.	deviation					
147.	rational					
148.	repulsion					
149.	deflect					
150.	bladder					
151.	graphite					
152.	ingest					
153.	refraction					
154.	applied					
155.	optical					
156.	calculated					
157.	microorganism					
158.	exponentially					
159.	foil					
160.	nonpolar					
161.	excrete					
162.	implant					
163.	parasitic					
164.	cardiac					
165.	destructive					
166.	elapse					
167.	dimensionless					
168.	condensed					
169.	spinal					
170.	generalization					
171.	postulate					
172.	elimination					
173.	colorless					
174.	planar					

175.	capacitance					
176.	vein					
177.	infrared					
178.	upright					
179.	latitude					
180.	buffer					
181.	translational					
182.	elongate					
183.	pregnancy					
184.	accelerator					
185.	cyclic					
186.	multiplication					
187.	illuminate					
188.	shark					
189.	saturate					
190.	homogeneous					
191.	micrograph					
192.	watery					
193.	rupture					
194.	parabolic					
195.	reactor					
196.	superposition					
197.	outermost					
198.	elementary					
199.	buoyant					
200.	conductivity					
201.	subunits					
202.	ellipse					
203.	nutrition					
204.	lightbulb					
205.	fetus					
206.	endpoint					
207.	nucleic					
208.	algebra					
209.	dissociate					
210.	continuity					
211.	logarithmic					
212.	magnification					
213.	endangered					
214.	prefix					
215.	recycle					
216.	arctic					
217.	pea					
218.	spacing					
219.	semicircle					
220.	predatory					

J. QUESTIONNAIRE ON MULTI-WORD UNITS

Please give your answer to the question:

To what extent is the word useful for your students in the science courses?

by choosing a number from 1 to 5 in the degree of usefulness column (1 is the LEAST useful and 5 is the MOST useful.

Item number	**Item**	**Degree of Usefulness**				
		5	4	3	2	1
1.	time interval					
2.	kinetic energy					
3.	electric field					
4.	magnetic field					
5.	straight line					
6.	potential energy					
7.	chemical reaction					
8.	using equation					
9.	hydrogen atom					
10.	surface area					
11.	internal energy					
12.	maximum value					
13.	water molecule					
14.	rate of change					
15.	amino acid					
16.	force act					
17.	carbon atom					
18.	center of mass					
19.	gravitational force					
20.	positive charge					
21.	total energy					
22.	same value					
23.	amount of energy					
24.	constant speed					
25.	blood cell					
26.	boiling point					
27.	si unit					
28.	negative sign					
29.	function of time					
30.	numerical value					
31.	high temperature					
32.	same direction					
33.	potential difference					

34.	nervous system					
35.	ideal gas					
36.	cross-sectional area					
37.	molar mass					
38.	negative value					
39.	particle move					
40.	chemical bond					
41.	side of the equation					
42.	speed of light					
43.	net force					
44.	minimum value					
45.	following statement					
46.	covalent bond					
47.	negative charge					
48.	hydrogen bond					
49.	circle of radius					
50.	periodic table					
51.	height h					
52.	organic molecule					
53.	energy transfer					
54.	charge density					
55.	initial value					
56.	melting point					
57.	angular speed					
58.	oxygen atom					
59.	blood vessel					
60.	initial velocity					
61.	number of electrons					
62.	chemical equation					
63.	cell wall					
64.	sphere of radius					
65.	significant figure					
66.	red blood cell					
67.	red blood					
68.	tangent line					
69.	average value					
70.	conservation of energy					
71.	electric charge					
72.	immune system					
73.	sound wave					
74.	specific heat					
75.	atomic mass					
76.	positive value					
77.	mechanical energy					
78.	atomic number					
79.	length l					
80.	direction of motion					

81.	line segment					
82.	temperature increase					
83.	unit area					
84.	energy change					
85.	differential equation					
86.	absolute value					
87.	liquid water					
88.	electromagnetic wave					
89.	organic compound					
90.	total mass					
91.	isolated system					
92.	visible light					
93.	number of atoms					
94.	velocity vector					
95.	air resistance					
96.	use datum					
97.	atmospheric pressure					
98.	initial speed					
99.	body temperature					
100.	constant value					
101.	average speed					
102.	unit volume					
103.	unit length					
104.	charged particle					
105.	coordinate system					
106.	temperature change					
107.	law of conservation					
108.	electric current					
109.	constant rate					
110.	oxygen gas					
111.	cell membrane					
112.	equilibrium position					
113.	constant acceleration					
114.	circular orbit					
115.	much work					
116.	external force					
117.	dashed line					
118.	vertical line					
119.	water vapor					
120.	forces act					
121.	object move					
122.	continuous function					
123.	same speed					
124.	law of thermodynamics					
125.	back of the book					
126.	light source					
127.	maximum height					

128.	constant velocity					
129.	given point					
130.	key concept					
131.	number of moles					
132.	red arrow					
133.	angular momentum					
134.	blue line					
135.	right triangle					
136.	electric force					
137.	cross section					
138.	circular path					
139.	wave function					
140.	electric potential					
141.	partial pressure					
142.	gas molecule					
143.	solution contain					
144.	first law					
145.	population growth					
146.	electromagnetic radiation					
147.	immune response					
148.	overall reaction					
149.	net charge					
150.	maximum speed					

K. APPROVAL OF THE METU HUMAN SUBJECTS ETHICS COMMITTEE

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr
www.ueam.metu.edu.tr

01 ARALIK 2022

Konu: Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Prof. Dr. Ayşegül DALOĞLU

Danışmanlığını yürüttüğünüz Şebnem Çiçek Demirci'nin ***"Türkiye'deki bir devlet üniversitesindeki mühendislik fakültesi öğrencilerinin sözcük bilgisi ihtiyaçlarının derlem tabanlı analizi"*** başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülerek **600-ODTUİAEK-2022** protokol numarası ile onaylanmıştır.

Bilgilerinize saygılarımla sunarım.

Prof. Dr. Sibel KAZAK BERUMENT
Başkan

Prof.Dr. İ.Semih AKÇOMAK
Üye

Dr. Öğretim Üyesi Müge GÜNDÜZ
Üye

Dr. Öğretim Üyesi Şerife SEVİNÇ
Üye

Dr. Öğretim Üyesi Murat Perit ÇAKIR
Üye

Dr. Öğretim Üyesi Süreyya ÖZCAN KABASAKAL
Üye

Dr. Öğretim Üyesi A. Emre TURGUT
Üye

L. CURRICULUM VITAE

PERSONAL INFORMATION

Name, Last Name: Şebnem ÇİÇEK DEMİRCİ

Nationality: Turkish

Date of Birth: 23/05/1979

Place of Birth: Istanbul, TURKEY

Marital Status: Married

Place of Residence: Ankara

Phone number: +90 532 606 87 22

E-mail address: demircisebnem@hotmail.com

EDUCATION

2005-2008	Hacettepe University	Department of Translation and Interpretation (MA)
1997-2001	Hacettepe University	Department of Translation and Interpretation (BA)
1994-1997	Eskişehir Atatürk High School	
1986- 1994	Private Seymen Primary School	

OCCUPATIONAL EXPERIENCE

2004-present	Ankara University	Instructor
2002-2003	Eskişehir Final Dergisi Dershaneleri	English teacher
2001-2002	Private Çağdaş Primary School	English teacher

M. TURKISH SUMMARY / TÜRKÇE ÖZET

TÜRKİYE'DEKİ BİR DEVLET ÜNİVERSİTESİ'NDEKİ MÜHENDİSLİK FAKÜLTESİ ÖĞRENCİLERİNİN SÖZCÜK BİLGİSİ İHTİYAÇLARININ DERLEM TABANLI ANALİZİ

Giriş

Fen dersleri mühendislik fakültesi müfredatının önemli bir bölümünü oluşturmaktadır. Bu çalışma, bir devlet üniversitesindeki birinci sınıf mühendislik öğrencilerinin fen derslerindeki sözcük bilgisi ihtiyaçlarını belirleme gereksiniminden ortaya çıkmıştır. Bu ihtiyaçları belirlemenin müfredat geliştirme, materyal tasarlama ve ölçme-değerlendirme geliştirme alanlarında faydalı olacağı düşünülmektedir. Bu çalışma, fen derslerinde kullanılan ders kitaplarından oluşturulan, nesnel derlem verilerini kullanarak bir sözcük listesi oluşturmayı hedeflemektedir. Bu amaçla, birinci sınıf mühendislik öğrencileri tarafından alınan fizik, kimya, matematik ve biyoloji derslerinde kullanılan ders kitaplarından bir derlem oluşturulmuş ve bu derlem üzerinde anahtar sözcük analizi yapılmıştır. Derlem verileri ile oluşturulan listedeki sözcüklerin öğrenciler için ne ölçüde faydalı olduğu ile ilgili uzman görüşü alınmıştır. Nesnel, niceliksel derlem verisinin yanı sıra görüşme ve anketler ile öznel, niceliksel verilerden de yararlanan bu çalışma, pedagojik olarak uygun, derleme dayalı ve 1194 sözcükten oluşan bir hedef sözcük listesi sunmaktadır ve bu listenin yüksek öğrenim düzeyinde fen derslerini alan mühendislik öğrencileri için faydalı olacağı düşünülmektedir.

Yabancı dil öğreniminde sözcük bilgisinin önemli bir yer tutmaktadır ve Nation'a göre sıklığı yüksek sözcüklere odaklanmak dil yeterliği gelişiminde oldukça etkilidir (2006). Derlem, bir dile ait metinlerin elektronik olarak bir araya getirilmiş bütünü olarak tanımlanabilir ve derlemler sayesinde o dilde sıklıkla görülen kalıplara ulaşabilmek mümkün olur. Elektronik araçlara dayalı bir metin analiz yöntemi olarak derlem dilbilimi 1960-70'li yıllarda Brown ve LOB (Lancester-Oslo/Bergen) derlemlerinin geliştirilmesi ile başlamıştır (Gavioli, 2005). 1990'larda Sinclair (1987) tarafından geliştirilen Cobuild projesi alanda çığır açmıştır. Proje, sınıfta öğretilen İngilizcenin daha gerçekçi tanımlarını üretmeyi hedeflemiştir. Son yıllarda derlem dilbilimin kaydettiği ilerlemeye ragmen, dil öğretiminde ve öğreniminde derlem kullanımı halen sınırlıdır. Kennedy'e göre (2004) son otuz yılda derlem araştırmalarının yabancı dil müfredatları üzerinde neredeyse hiç etkisi olmamıştır. Ancak yine de, Biber ve Reppen (2002(tarafından da belirtildiği üzere derlemlerin ampirik analizleri İngilizcenin gerçek kullanımları konusundaki tanımlamalara önemli ölçüde katkıda bulunmuştur. Nelson'a göre (2004) son yirmi yılda elektronik derlemlerdeki ilerleme çevirim içi erişimi mümkün olan çok miktarda veriyi ulaşılabilir kılmıştır.

Derlem verisi belirli bir disipline özgü söylemlerin yinelenen özelliklerini anlamada etkili bir araçtır. Bu bağlamda, bir dilin belirli bir yönüne odaklanan özel derlemler (Bowker ve Pearson, 2002), özel amaçlı dil öğretiminde önemli bir kaynak değerindedir. Flowerdew'a göre (1993), özel bir derlemde en sık görülen sözcüklerden oluşan bir liste, özel amaçlı dil öğretimine yönelik bir izlenceye dahil edilecek içeriği seçmek için kullanılabilir. Flowerdew (1993) öğrencilerin okumakla yükümlü oldukları bir dizi biyoloji metnini ve katıldıkları derslerin transkriptlerini incelemiştir. Yüz bin sözcükten oluşan derleminde bazı sözcük türlerinin genel bir derleme kıyasla çok daha sık görüldüğünü tespit etmiştir ve bunun da izlence tasarlamada önemli bir role sahip olabileceği sonucuna varmıştır.

Her disiplinin kendine özgü bir söylemden ve sözcüklerden oluştuğu varsayımından hareketle, bu çalışma mühendislik alanındaki üniversite öğrencilerinin birinci sınıf düzeyinde aldıkları ders içeriklerinin sözcüksel özelliklerini derlem verisi yoluyla belirlemeyi hedeflemiştir.

Ankara Üniversitesi, Türkiye'nin en köklü ve çok sayıda öğrencisi olan üniversitelerinden biridir. Üniversite sınavını kazanarak bölümlerine kayıt yaptıran öğrenciler, şayet bölümlerinin eğitim dili İngilizce ise, İngilizce dil yeterliğine sahip olduklarını belgelemek durumundadırlar. Üniversitelerin yabancı diller yüksekokulları tarafından uygulanan İngilizce yeterlik sınavında başarılı olan öğrenciler bölümlerine devam ederken, başarılı olamayanlar ise hazırlık programında bir yıl İngilizce eğitimi alırlar.

Ankara Üniversitesi Yabancı Diller Yüksekokulu Hazırlık bölümünde kayıtlı toplam öğrenci sayısı 2022-2023 akademik yılında 995'tir ve bu öğrencilerin 554'ü mühendislik fakültesi öğrencileridir. Hazırlık programında en büyük grubu oluşturan mühendislik fakültesi öğrencileri diğer fakültelerin öğrencileri ile bir arada karma gruplarda eğitim almaktadır. Hazırlık eğitimi süresince genel İngilizce programına tabi olan bu öğrencilerin hazırlık programında öğrendikleri İngilizceyi bölümlerine geçtikleri ilk öğretim yılında ne gibi görevlerde ve nasıl kullanacaklarına dair bir bilgi bulunmamaktadır. Bu öğrenciler hazırlık programında bir genel İngilizce ders kitabı ile birlikte materyal geliştirme birimince hazırlanan destekleyici materyaller kullanmaktadırlar; ancak ders kitabı ve ek materyallerin içeriğinin ne ölçüde bu öğrencilerin bölümdeki derslerinde ihtiyaç duyacakları içerik ile ne ölçüde örtüştüğüne dair veri yoktur. Mühendislik öğrencilerinin bölüm derslerinde hangi sözcüksel bilgiye ihtiyaç duydukları ve bu sözcüklerin hazırlık programı kapsamında yer alıp almadığı bilinmemektedir. Hazırlık programında uygulanan ölçme değerlendirme bileşenleri de yine öğrencilerin hedef gerekliliklerine göre değil kullanılan ders kitabı ve materyal içeriklerine göre belirlenmektedir. Uygulanan sınavlarda sorgulanan sözcüklere öğrencilerin bölümlerinde ne ölçüde ihtiyaç duyacakları bilinmemektedir.

Gerek hazırlık programı gerekse bölümde verilen İngilizce dersleri müfredatına katkıda bulunma düşüncesinden hareketle bu çalışma mühendislik fakültesi öğrencilerinin sözcük düzeyinde hedef ihtiyaçlarını belirlemeyi amaçlamaktadır. Mühendislik fakültelerinde verilen birinci sınıf derslerinin çoğunluğunu Fizik, Kimya, Biyoloji ve Kalkülüs dersleri oluşturmaktadır. Bu nedenle, bu derslerin içerikleri incelenerek sözcük ihtiyaçlarının belirlenmesi hedeflenmiştir. Bu bağlamda, çalışmada aşağıdaki araştırma sorularına cevap aranmıştır:

1. Mühendislik birinci sınıf öğrencilerinin fen derslerindeki hedef sözcük ihtiyaçları nelerdir?
1.1. Mühendislik birinci sınıf öğrencilerinin hedef ihtiyaçları konusunda öğretim üyelerinin görüşleri nelerdir?

1.2. Mühendislik birinci sınıf öğrencilerinin fen derslerinde kullandıkları ders kitaplarının spesifik sözcük içerikleri nedir?

1.2.1. Mühendislik birinci sınıf öğrencilerinin fen derslerinde kullandıkları ders kitaplarındaki sözcüklerin sıklık özellikleri nedir?

1.2.2 Mühendislik birinci sınıf öğrencilerinin fen derslerinde kullandıkları ders kitaplarındaki anahtar sözcükler ve sözcük öbekleri nelerdir?

1.3. İngilizce hazırlık programının içeriği mühendislik birinci sınıf öğrencilerinin fen derslerinde ihtiyaç duydukları hedef sözcükleri ne ölçüde karşılamaktadır?

1.4. Fen derslerinde kullanılan ders kitaplarından oluşturulan bir derleme dayalı anahtar sözcük listesi, yaygın şekilde kullanılan sözcük listeleri ("*New General Service List", "New Academic Vocabulary List"* ve "*Science Word List")* ile ne ölçüde örtüşmektedir?

1.5. Fen derslerinde kullanılan ders kitaplarından oluşturulan derleme dayalı anahtar sözcük listesindeki sözcükleri öğrenmenin faydası konusunda öğretim üyelerinin görüşleri nelerdir?

Çalışmanın önemi

Bir çok öğretmene göre lisans öğrencilerinin Alana özel sözcük bilgisi sahibi olmasına yardımcı olmak önemlidir ve material geliştirme uzmanlarına kılavuzluk edecek ve öğrencilerin öğrenme süreçlerini planlamalarına yardımcı olacak anahtar sözcük listeleri geliştirme girişimleri olmuştur (Hyland and Tse, 2007). Fakat bu listeler genellikle genel veya akademik bağlamlarda sıklıkla karşılaşılan ve öğrencilere çalışmalarında yardımcı olabileceği düşünülen sözcüklerden oluşmaktadır. Ancak bu tür bir listenin her disiplin, alan veya tür için standart olabileceğini düşünmek bir yanılgıdır. Bu bağlamda, Hyland ve Tse (2007, p. 236-237) şu görüştedir:

> …öğrencilerin genel academik sözcük dağarcığına sahip olmalarının faydalı olup olmadığı daha tartışmalıdır çünkü bunun, kaydadeğer bir öğrenme çabasının karşılığında neredeyse hiç getirisi olmayabilir.

Bu çalışma, mühendislik öğrencilerinin akademik çalışmalarının önemli bir bölümünü oluşturan fen destlerine özgü spesifik bir sözcük profile olduğu ve jenerik sözcük listelerinin bu profili yansıtmayacağı varsayımına dayanmaktadır. Genel bir sözcük listesinin, specifik bir bağlamda spesifik bir grup öğrencinin spesifik ihtiyaçlarına cevap veremeyeceği düşünülmektedir. Bu nedenle, mühendislik öğrencilerinin birinci sınıf dersleri için spesifik sözcük ihtiyaçlarının belirlenmesi hedeflenmektedir. Bu ihtiyaçları yansıtan bir listening müfredat geliştirme, içerik belirleme, ölçme ve değerlendirme bileşenlerini oluşturma ve benzeri süreçlere olumlu katkısı olacağı varsayılmaktadır.

Araştırma Deseni

Bu çalışmada karma yöntem tekli vaka çalışması benimsenmiştir. Çalışma, gerçek hayatta spesifik bir grubun öğrenim ihtiyaçlarını, herhangi bir müdahale olmaksızın belirlemeyi hedeflediğinden bir vaka çalışması olarak nitelendirilebilir. Çalışmada niteliksel ve niceliksel veri toplama yöntemleri bir arada kullanılmıştır. Karma yöntem çalışmaları, araştırmacının tek bir çalışmada veri toplama, analiz etme ve bulguları raporlama için en az bir niteliksel ve bir niceliksel yöntem kullandığı çalışmalardır (Fielding & Fielding, 1986; Greene ve ark., 1989). Aşağıdaki tablo, bu çalışmada benimsenen araştırma tasarımının genel hatlarını ortaya koymaktadır.

<table>
<tr><th></th><th>Veri toplama ve analizi</th><th>Araştırma soruları</th></tr>
<tr><td>1. aşama</td><td>Öğretim üyeleri ile görüşme</td><td>1.1. Mühendislik birinci sınıf öğrencilerinin hedef ihtiyaçları konusunda öğretim üyelerinin görüşleri nelerdir?</td></tr>
<tr><td>2. aşama</td><td>Derlem oluşturma</td><td rowspan="4">1.2. Mühendislik birinci sınıf öğrencilerinin fen derslerinde kullandıkları ders kitaplarının spesifik sözcük içerikleri nedir?
1.2.1. Mühendislik birinci sınıf öğrencilerinin fen derslerinde kullandıkları ders kitaplarındaki sözcüklerin sıklık özellikleri nedir?
1.2.2 Mühendislik birinci sınıf öğrencilerinin fen derslerinde kullandıkları ders kitaplarındaki anahtar sözcükler ve sözcük öbekleri nelerdir?</td></tr>
<tr><td>3. aşama</td><td>Sıklık listesi oluşturma</td></tr>
<tr><td>4. aşama</td><td>Anahtar sözcük analizi ve anahtar sözcük listesi oluşturma</td></tr>
<tr><td>5. aşama</td><td>CEFR düzeyine göre sınıflandırma ve sözcük türü belirlemesi</td></tr>
</table>

6. aşama	Hedef sözcük şistesinin hazırlık programında kullanılan sözcük listesi ile karşılaştırılması	1.3. İngilizce hazırlık programının içeriği mühendislik birinci sınıf öğrencilerinin fen derslerinde ihtiyaç duydukları hedef sözcükleri ne ölçüde karşılamaktadır?
7. aşama	Hedef sözcük listesinin sıklıkla kullanılan jenerik sözcük listeleri ile karşılaştırılması	1.4. Fen derslerinde kullanılan ders kitaplarından oluşturulan bir derleme dayalı anahtar sözcük listesi, yaygın şekilde kullanılan sözcük listeleri ("*New General Service List", "New Academic Vocabulary List"* ve "*Science Word List")* ile ne ölçüde örtüşmektedir?
8. aşama	Hedef sözcük listesi ile ilgili öğretim üyelerinden anket yoluyla görüş alınması	1.5. Fen derslerinde kullanılan ders kitaplarından oluşturulan derleme dayalı anahtar sözcük listesindeki sözcükleri öğrenmenin faydası konusunda öğretim üyelerinin görüşleri nelerdir?

Çalışmanın ilk aşaması ihtiyaç analizi sürecinden oluşmaktadır. Mühendislik birinci sınıf öğrencilerinin fen dersleri için yükümlülüklerinin neler olduğunu ve öğrendikleri İngilizce ile yerine getirmeleri gereken görevleri ve bunun için ne gibi becerilere ihtiyaç duyduklarını belirlemek üzere ihtiyaç analizi yapılmıştır. İhtiyaç analizinde Hutchinson and Waters'ın (1987) modeli kullanılmıştır.

İhtiyaç analizinde yalnızca hedef gereklilikler belirlenmesi amaçlandığından öğrencilerin öğrenme tercihleri çalışmanın kapsamı dışında tutulmuş; öğretim üyelerinden veri toplanmıştır.

Çalışmanın birinci aşamasında, öğretim üyeleri ile yapılan yarı-yapılandırılmış görüşmeler yoluyla ders gereklilikleri ve öğrencilerin ihtiyaçları konusunda veri toplanmıştır. Mühendislik fakültesinde fizik, kimya, biyoloji ve kalkülüs derslerini veren 7 öğretim üyesi ile görüşmeler yapılmış, yapılan görüşmeler yazıya dökülmüş ve içerik analizi yöntemi ile incelenmiştir. İnceleme sonucunda ortaya çıkan temalar sınıflandırılmıştır.

Çalışmanın ikinci aşamasında mühendislik birinci sınıf öğrencilerinin fen derslerinde kullandıkları ders kitaplarından bir derlem oluşturulmuştur. Aşağıdaki tabloda, söz konusu kitaplar gösterilmiştir.

Zorunlu dersler	Kullanılan ders kitapları
Physics	*Physics for scientists and engineers*. R. A., & Jewett, J. W. (2018). Cengage learning. (6th Edition)
Calculus	*Thomas' Calculus*. Thomas, G. B., Weir, M. D., Hass, J., & Giordano, F. R. (2005). Addison-Wesley.
Chemistry	*General Chemistry: Principles and Modern Applications* Petrucci, R. H., Herring, F. G., & Madura, J. D. (2010). Pearson Prentice Hall.
Biology	*Biology: Life on Earth*. Audesirk, T., Audesirk, G., & Byers, B. E. (2001). Pearson Educación.

İlgili kitaplar elektronik ortamda txt. formatına dönüştürülmüş ve kitaplarda yer alan başlıklar, resimler, figürler, içindekiler bölümü ve benzeri içerik temizlenmiş ve standardize edilmiştir. Akabinde, dosyalar Sketch Engine programına yüklenerek bir derlem oluşturulmuştur.

Oluşturulan derlem üzerinde sıklık analizi yapılarak derlemde en sık karşılaşılan ifadeler belirlenmiştir. Oluşturulan liste, detaylı bir şekilde revize edilerek sözcük düzeyinde olmayan ifadeler, gramer unsurları, semboller, kısaltmalar ve benzeri içerik ayıklanmıştır.

Sonraki aşamada, oluşturulan derlem, genel bir derlem olan BNC derlemi ile karşılaştırılarak anahtar sözcük analizi yapılmıştır. Oluşturulan liste yine revize edilerek belirli bir sıklık düzeyinin altında kalan sözcükler ile diğer alakasız içerik çıkarılmıştır. Listedeki sözcükler CEFR düzeylerine göre sınıflandırılmış ve sözcük türü bilgisi eklenmiştir. A1 düzeyindeki sözcükler listeden çıkarılmıştır ve 1195 sözcüklük bir liste elde edilmiştir.

Yine aynı analiz sonucunda derlemde sıklıkla yer alan eşdizimli sözcükler sıklık verisine göre sıralanmış ve liste revize edilerek 379 maddelik bir liste elde edilmiştir.

Sıklık, dağılım ve anahtar kelime kriterlerine dayalı, derlem temelli bir liste oluşturduktan sonra bu liste hazırlık programında öğretilen sözcükler ile kıyaslanarak hazırlık eğitiminde kullanılan materyallerin mühendislik birinci sınıf öğrencilerinin ihtiyaçlarını ne ölçüde karşıladığı araştırılmıştır. Aynı liste dil öğretiminde yaygın olarak kullanılan kelime listeleri ile de kıyaslanarak listeler arasındaki benzerlik araştırılmıştır. Kıyaslamada AntWord Profiler programı kullanılmıştır.

Niceliksel verileri niteliksel veriler ile desteklemek amacıyla öğretmen görüşüne başvurulmuştur. Mühendislik birinci sınıf öğrencilerinin ihtiyaç duydukları sözcüklerin derleme dayalı oluşturulan listesi öğretmen görüşüne sunularak listedeki sözcükleri bilmenin öğrencilere ne ölçüde faydalı olacağına dair fikir istenmiştir. Listedeki sözcüklerden bir örneklem grubu oluşturulmamıştır çünkü hiçbir sözcüğün bir başka sözcüğü temsil edebilme durumu yoktur. Bu nedenle, liste beşe bölünerek beş ayrı anket hazırlanmış ve öğretmenlerden sözcükleri Likert ölçeğine göre 1 ile 5 arasında derecelendirmeleri istenmiştir. Ankete 13 öğretmen katılmıştır. Anket sonucunda çıkan verilere göre 1 ve 2 puan alan yani "faydasız" olarak nitelendirilen sözcükler listeden çıkarılmamış fakat listede işaretlenmiştir.

Sonuçlar

Bu bölümde, görüşmelerden elde edilen sonuçlar, derlem oluşturulması ve derlem üzerinde yapılan analizlerden elde edilen sonuçlar, ve oluşturulan kelime listesi içeriği hakkında öğretmenlere uygulanan anketlerden elde edilen sonuçlar açıklanmıştır.

Görüşme Sonuçları

Çalışmanın ilk araştırma sorusu mühendislik birinci sınıf öğrencilerinin aldıkları fizik, kimya, biyoloji ve kalkülüs derslerindeki hedef gereklilikleri belirlemeye yönelikti. Bu amaçla, dersleri veren öğretim üyeleri ile derslerin gereklilikleri ve öğrencilerin eksikliklerinin neler olduğunu belirlemeye yönelik görüşme yapıldı. Yedi öğretim üyesi ile yapılan görüşmelerden çıkan ortak temalar aşağıdaki tabloda özetlenmiştir.

İçerik Analizi sonuçları

	Temalar	Alt-kategoriler	N
Gereksinimler	Dersin gereklilikleri	▪ Sınav sorularını anlama ve yanıtlayabilme ▪ Yazılı materyalleri ve dersleri kavrayabilme ▪ Sunum yapabilme ▪ Eşitlikleri ve teoremleri okuyabilme	7 7 2 2
	Ders içeriği	▪ Ders kitapları ▪ Dersler ▪ Spesifik sözcükler	7 7 4
	İhtiyaç duyulan beceri ve alt-beceriler	▪ Genel yeterlik ▪ Sözcük bilgisi ▪ Dinleme ▪ Konuşma	7 4 4 4
Eksiklikler	Öğrencilerin karşılaştıkları zorluklar	▪ Uzun cümleleri ve kelimeleri anlama ▪ Sınav sorularını anlama ▪ Kendilerini ifade etme ▪ Sunum becerileri	7 4 3 2
Suggestions		▪ Daha iyi dil yeterliği ▪ Alana özgü sözcüklere aşinalık ▪ Bilimsel metin okuma ▪ Sunum becerileri	5 4 2 1
		▪ Özel amaçlı İngilizce müfredatı ▪ Genel İngilizce müfredatı	6 1

Öğretim üyelerine, dersin gerekliliklerinin neler olduğu, öğrencilerin neler yapmaları gerektiği ve ders hedeflerine ulaşmak için ne gibi beceri ve alt becerilere sahip olmaları gerektiği soruldu. Verilen yanıtlar arasında en sık bahsedilen konu derslerin ölçme değerlendirme boyutu ile ilgiliydi. Tüm öğretim elemanları sınav sorularının tam anlaşılmasının ve soruların yanıtlanabilmesine değindi. Öğrencilerin kendilerinden bekleneni kavrayabilmeleri ve sorulara tatminkar şekilde yanıt verebilmeleri gerektiği belirtildi. Bunu yapabilmek için de iyi bir dil yeterliğine ve sözcük bilgisine sahip olmalarının önemi vurgulandı.

Derslerin bir diğer gerekliliği ise yazılı materyallerin ve sözlü anlatımların tam anlaşılması olarak rapor edildi. Görüşülen tüm öğretim üyeleri ders kitaplarındaki yazılı içeriğin ve öğretmenlerin yaptığı sözlü anlatımların anlaşılmasının önemine değindi. Bunun da iyi bir dil yeterlik düzeyi, gramer ve kelime bilgisi ile mümkün olacağı belirtildi.

Görüşülen öğretim üyelerinden iki tanesi ders gereklilikleri arasında sunum becerilerinden bahsetti. Yine iki öğretim üyesi, fen derslerinin önemli unsurlarından olan denklem ve teorem okuma becerisinin önemini vurguladı. Bunun da geniş bir kelime dağarcığının yanısıra alana özgü kullanımlara hakim olunması gerektiği belirtildi.

Ders içeriğinin ders kitabına ve öğretmenlerin yaptıkları konu anlatımlarına dayalı olduğu belirtildi. Dört öğretim üyesi derslerin Alana özgü sözcükler ve kalıp kullanımlar içerdiğini ve öğrencilerin bunlara aşinalık kazanmasının içeriği daha iyi kavramalarını sağlayacağını vurguladı.

Ders gerekliliklerinin yerine getirebilmek için öğrencilerin iyi bir İngilizce düzeyine sahip olmalarının yanı sıra ders içeriğini anlamada ve ölçme değerlendirme uygulamalarını yerine getirmede sözcük bilgisinin önemi dört öğretim üyesi tarafından dile getirildi. Derslerde karşılaşacakları sözcüklere aşina olmalarının anlama ve ifade etmede olumlu bir role sahip olacağı belirtildi. Yine dört öğretim üyesi, derslerin büyük ölçüde sözlü anlatım yoluyla yapıldığını ve dolayısıyla dinleme becerisinin önemini vurguladı. Sözlü beceriler bakımından ise, öğrencilerin fikirlerini ifade edebilmeleri ve soru sorabilmeleri gerektiği vurgulandı.

Öğrencilerin karşılaştıkları zorluklar bakımından ise özellikle karmaşık cümle yapıları ve zor sözcükleri anlamada zorlandıkları, bunun da hedef içeriğin anlaşılmasına olumsuz etki ettiği belirtildi. Öğretim üyelerinden bazıları öğrencilerin sınav sorularını anlayamadığını ve Türkçe açıklama istediklerini belirttiler.

Öğrencilerin anlamada karşılaştıkları zorlukların yanı sıra üretime dayalı becerilerde de zorluk çektikleri belirtildi. Dört öğretim üyesi öğrencilerin kendilerini ifade edemediklerini bu nedenle de soru sormak istediklerinde bile sessiz kalmayı tercih ettiklerini ya da Türkçe konuşmaya çalıştıklarını söyledi.

İki öğretim üyesi öğrencilerin gerekli sunum becerilerine sahip olmadıklarını belirtti. Belirli bir konuda nasıl araştırma yapılacağını, belli başlı temel noktaların nasıl çıkarılacağını ve bir Powerpoint sunumunun nasıl hazırlanıp sunulacağının bilinmediği belirtildi.

Öğrencilerin eksikliklerini gidermeleri ve ders gerekliliklerini hakkıyla yerine getirebilmeleri için yapılabilecekler konusunda öğretim üyelerinin önerileri alındı. Öğretim üyelerinin çoğu öncelikle genel yeterlik düzeyinin iyileştirilmesi gerektiğini belirtti. Yazılı ve sözlü materyallerin daha iyi anlaşılması için hedef dil becerilerinin geliştirilmesi gerektiği vurgulandı. Ayrıca, bilimsel makalelerde yaygın olarak kullanılan spesifik sözcükleri bilmenin de metni anlamaya olumlu etkisi olacağı rapor edildi. Bu bağlamda, öğrencilerin derste öğreneceği bilimsel kavramları derinlemesine öğrenmiş olmalarının beklenmediği fakat daha az teknik ama Alana özgü sözcüklere aşina olunmasının önemli kavram ve metinlerin anlaşılmasına olumlu katkı sağlayacağı vurgulandı. Hazırlık programında kendi düzeylerine uygun bilimsel metinler de okunmasının hem sözcük öğrenimi bakımından hem de konulara aşinalık kazanma bakımından önemli olduğu söylendi.

Hazırlık eğitiminde mühendislik öğrencilerine yönelik özel amaçlı İngilizce programı (ESP) uygulanması konusunda fikirleri sorulduğunda, öğretim üyelerinin altısı bunun faydalı olacağını belirtti. Alana özgü içerik ile aşinalık kazanmanın öğrencilerin akademik performansını artıracağına dair görüş bildirildi. Bir öğretim üyesi aynı soruya olumsuz yanıt vererek, spesifik bir programa gerek olmadığını, genel İngilizce düzeylerinin artırılmasının yeterli olacağını belirtti.

Derlem oluşturma

Mühendislik fakültesi birinci sınıf öğrencilerinin fizik, kimya, biyoloji ve kalkülüs derslerinde kullandıkları ders kitapları format uyarlaması ve standardizaston sonrasında Sketch Engine programına yüklenerek 2,303,096 ifade, ve 1,898,324 kelimeden oluşan bir derlem oluşturuldu.

Hedef derlem her biri farklı alanlar olmak üzere dört alt derlemden oluştu. Aşağıdaki tablo her bir derlemin kaç ifadeden oluştuğunu ve bütün derleme oranını göstermektedir.

Alt-derlem	*İfade sayısı*	*Yüzde (%)*
Fizik	783,425	34
Biyoloji	591,391	25.7
Kimya	577,433	25.1
Kalkülüs	350,847	15.2

Hedef derlemin genel derlemden farklı olduğundan emin olmak amacıyla bir karşılaştırma yapıldı. Sketch Engine programı kullanılarak, oluşturulan Fen Kitapları Derlemi, genel bir derlem olan İngiliz Ulusal Derlemi (BNC) ile karşılaştırıldı ve 3.96 değeri elde edildi; bu değer iki derlem arasında önemli bir fark olduğunu göstermektedir.

Sıklık analizi

Mühendislik birinci sınıf öğrencilerinin kullandıkları fen kitaplarındaki içeriğin sözcüksel özelliklerini ve sıklık temsillerini araştırmak üzere oluşturulan Fen Kitapları Derlemi üzerinde sıklık analizi gerçekleştirildi. Analizde sıklık eşiği 50 olarak belirlendi ve toplam 2954 ögeden oluşan bir liste elde edildi. Derlemde sıklığı en yüksek olan gramer sözcükleri manuel olarak listeden çıkarıldı. Ayrıca, yalnızca tek bir alt derlemde bulunan ifadeler ile semboller, kısaltmalar, bağlaçlar, özel isimler ve hatalı girişler araştırmacı tarafından listeden çıkarıldı ve 1688 sözcükten oluşan nihai sıklık listesi oluşturuldu.

Anahtar sözcük analizi

Araştırma sorularından, mühendislik birinci sınıf öğrencilerinin kullandıkları fen kitaplarında hangi anahtar kelimeler ve kelime grupları sıklıkla yer almaktadır sorusuna yanıt aramak amacıyla hedef derlem üzerinde sıklık analizi yapılmıştır. Bir derlemde spesifik oluşumları tespit etmenin en yaygın yolu özel derlemi genel bir derlem ile kıyaslamaktır. Genel referans derlemde düşük sıklıkta fakat özel derlemde yüksek sıklıkta görülen ifadeler anahtar kelimeler olarak kabul edilir. Bu çalışmada oluşturulan Fen Kitapları Derlemi, BNC ile kıyaslanmıştır. Elde edilen sonuçlardan, belge sıklığı 2'nin altında olan, yani 2'den az alt derlemde bulunan ifadeler listeden çıkarılmıştır. Ayrıca liste kısaltmalar, özel isimler, gramer kullanımları bakımından revize edilmiş ve 1249 sözcük elde edilmiştir.

Sonraki aşamada listed yer alan sözcüklerin sözcük türü ve CEFR düzeyleri belirlenmiştir. CEFR düzeyleri belirlenirken "Text Inspector" sitesinden yararlanılmıştır. Sözcükler A1, A2, B1, B2, C1 ve C2 olarak etiketlenmiştir. Listede 450 adet sözcük ile ilgili düzey bilgisi bulunamamıştır. CEFR düzeylerinin belirlenmesinden sonra, liste yeniden gözden geçirilmiştir. Alana özgü spesifik bir anahtar sözcük listesi için A-1 düzeyi sözcüklerin çok basit olduğuna karar verilerek, 53 adet A-1 düzeyi sözcük araştırmacı tarafından listeden çıkarılmıştır. Nihai liste 1195 sözcükten oluşmuştur.

Oluşturulan Fen Kitapları Sözcük Listesi, ortalama indirgenmiş sıklık değeri (average reduced frequency-ARF- value) dikkate alınarak sıralanmıştır. ARF değeri en sık görülen ama aynı zamanda da derlem içinde dengeli bir dağılım göstermiş ifadeleri ortaya koyar. Diğer bir deyişle, sıklık ve dağılımı tek bir ölçüde bir araya toplar (Savický & Hlaváčová 2002). Listenin başında bulunan "point" sözcüğü ile ilgili sıklık verileri, sözcüğün hedef derlemde 4730 kere görülmüş, bir milyon sözcük içerisinde ise 2053.75 kez görülebildiğini ve 1647.80 ARF değeri ile en sık ve en dengeli dağılımı olan sözcük olduğunu ve dört alt derlemin hepsinde geçtiğini göstermektedir. Listede yer alan ilk 30 sözcük sıklık değerleri ile birlikte aşağıdaki tabloda gösterilmektedir.

	İfade	*Sözcük türü*	*CEFR düzeyi*	Sıklık	Nispi sıklık	Belge sıklığı	ARF
31.	*point*	*n*	*A2*	4730	2053.75	4	1647.80
32.	*equation*	*n*	*C1*	4524	1964.31	4	1547.45
33.	*form*	*v/n*	*A2*	3177	1379.44	4	1485.96
34.	*value*	*v/n*	*B1*	3925	1704.22	4	1455.36
35.	*energy*	*n*	*B1*	5630	2444.53	4	1385.48
36.	*result*	*v/n*	*B1*	2170	942.209	4	1202.46
37.	*call*	*v/n*	*A2*	2409	1045.98	4	1200.78
38.	*produce*	*v*	*B1*	2639	1145.84	4	1140.43
39.	*function*	*n*	*B2*	3888	1688.16	4	1056.58
40.	*move*	*v*	*A2*	2666	1157.57	4	1027.43
41.	*increase*	*v/n*	*B1*	2469	1072.03	4	1017.93
42.	*follow*	*v*	*A2*	1957	849.72	4	979.89
43.	*constant*	*adj*	*B2*	2514	1091.57	4	955.75

44.	*large*	*adj*	*A2*	1951	847.12	4	946.56
45.	*system*	*n*	*B1*	3056	1326.91	4	907.43
46.	*cell*	*n*	*B2*	5311	2306.02	4	879.46
47.	*determine*	*v*	*C1*	1876	814.55	4	874.27
48.	*describe*	*v*	*A2*	1646	714.69	4	864.24
49.	*mass*	*n*	*B2*	3384	1469.327	4	859.34
50.	*force*	*v/n*	*B2*	4023	1746.779	4	859.17
51.	*occur*	*v*	*B2*	1834	796.3194	4	852.65
52.	*solution*	*n*	*B1*	3048	1323.436	4	830.46
53.	*high*	*adj*	*A2*	1682	730.3213	4	765.28
54.	*contain*	*v*	*B1*	1583	687.3357	4	752.13
55.	*line*	*n*	*A2*	2327	1010.379	4	749.05
56.	*molecule*	*n*		3143	1364.685	4	741.08
57.	*unit*	*n*	*B1*	1682	730.3213	4	728.97
58.	*surface*	*n*	*B2*	2469	1072.035	4	726.46
59.	*section*	*n*	*B1*	1381	599.6276	4	723.97
60.	*consider*	*v*	*B1*	1315	570.9706	4	709.20

Toplam 1195 sözcükten, 450'sinin düzeyi belirlenmemiştir. 269 sözcük B2 düzeyi sözcüklerden oluşarak listenin çoğunluğunu oluşturmaktadır. Bunu 193 B1 düzeyi, 119 C1, 93 A2 ve 71 C2 düzeyi sözcük izlemektedir.

Sketch Engine yazılımındaki aynı araç kullanılarak çoklu sözcük kalıpları da çıkarılmıştır. Toplam liste 892 ifadeden oluşmuştur. Listenin revize edilmiş ve 2 alt-derlemden daha az alt derlemde yer alan ifadeler çıkarılmıştır. Nihai liste 379 ifadeden oluşmaktadır. Listedeki en sık görülen ilk 30 sözcük aşağıdaki tabloda gösterilmiştir.

İfade	Sıklık	Nispi Sıklık	Belge Sıklığı	ARF
30. time interval	464	201.46	4	121.37
31. kinetic energy	564	244.88	4	99.75
32. electric field	904	392.51	4	96.29
33. magnetic field	912	395.98	4	77.94
34. straight line	169	73.37	4	69.96
35. potential energy	389	168.90	3	58.30
36. chemical reaction	195	84.66	4	54.80
37. hydrogen atom	269	116.79	3	49
38. surface area	150	65.12	4	46.66
39. internal energy	269	116.79	3	45.26
40. maximum value	167	72.51	3	43.93
41. water molecule	211	91.61	3	40.63

42. rate of change	155	67.30	4	40.43
43. amino acid	269	116.79	2	38.33
44. force act	174	75.55	2	37.38
45. carbon atom	285	123.74	3	36.77
46. center of mass	302	131.12	2	36
47. gravitational force	227	98.56	2	33.55
48. positive charge	154	66.86	3	33.30
49. total energy	147	63.82	3	32.94
50. same value	66	28.65	3	32.88
51. amount of energy	109	47.32	4	32.78
52. constant speed	123	53.40	4	32.56
53. blood cell	207	89.87	2	32.55
54. boiling point	171	74.24	4	32.52
55. negative sign	77	33.43	3	30.72
56. function of time	127	55.14	3	30.70
57. numerical value	62	26.92	3	29.32
58. high temperature	81	35.17	3	29.24
30. potential difference	317	137.64	2	28.28

Kapsam analizi

Derlem verilerinden elde edilen hedef sözcük listesinin mühendislik öğrencilerinin aldığı İngilizce hazırlık programında ne ölçüde öğretildiğini araştırmak amacıyla çalışmada oluşturulan Fen Kitapları Sözcük Listesi, hazırlık programında kullanılan ders kitabının kelime içeriği ile karşılaştırılmıştır. AntWord Profiler programı kullanılarak yapılan analiz sonucunda listede yer alan sözcüklerin yalnızca yüzde 12.6'sının hazırlık programında öğretilen sözcükler ile örtüştüğü bulunmuştur. İki listede 151 sözcük ortak olarak yer almaktadır. Elde edilen bu değer göstermektedir ki hazırlık programındaki içerik mühendislik birinci sınıf öğrencilerinin fen dersleri için ihtiyaç duydukları içeriği karşılamamaktadır.

Oluşturulan spesifik listenin, dil öğretiminde yaygın biçimde kullanılan Genel İngilizce alanındaki sözcük listeleri ile ve Coxhead ve Hirsh (2007) tarafından geliştirilmiş bilim alanına özgü spesifik bir sözcük listesi olan Bilim Sözcükleri Listesi (Science Word List) ile karşılaştırılması sonucunda şu değerler elde edilmiştir.

Bulunan değerlerden anlaşıldığı üzere, spesifik olarak mühendislik birinci sınıf öğrencilerinin ihtiyaçlarına göre oluşturulan sözcük listesi ile dil öğretiminde yaygın olarak kullanılan genel sözcük listeleri düşük ölçüde örtüşebilmektedir. Bu da göstermektedir ki, akademik ya da genel olması fark etmeksizin mevcut sözcük listeleri hedef grubun sözcük ihtiyaçlarını karşılamamaktadır. Benzer şekilde, özel bir liste olan Bilim Sözcükleri listesi de spesifik olarak mühendislik birinci sınıf öğrencilerinin ders kitaplarının içeriğinde yer alan sözcüklerin çok küçük bir kısmını kapsamaktadır. Bu nedenle, spesifik bir grubun spesifik ihtiyaçlarına yönelik, objektif derlem verilerine dayalı bir sözcük listesi, uygun şekilde kullanıldığında, dil öğretimine kaydadeğer katkılar sağlayabilecektir.

Sözcük listesi	Kapsama değeri
New GSL (Brezina and Gablasava, 2015)	% 32.20
New Academic Vocabulary List (Gardner and Davies, 2014)	% 30.8
Science Word List (Coxhead and Hirsch, 2007)	% 13.30

Anket sonuçları

Çalışmada, objektif niceliksel verileri subjektif niteliksel veriler ile desteklemek amacıyla, öğretmen görüşlerine başvurulmuştur. Anahtar sözcük listesi 5'e bölünerek, 5 ayrı anket hazırlanmıştır. Öğretmenlerden, anketteki sözcükleri bilmenin öğrenciler için ne ölçüde faydalı olacağını Likert ölçeği ile 1-5 arasında derecelendirmeleri istenmiştir. Aşağıdaki tabloda sonuçlar özetlenmiştir.

	Ortalama Skor	*Pearson'ın r değeri*	*Ortalama indirgenmiş skor*	*3'ün altında skor alan ifade sayısı*
Anket 1	4.085	*0.099*	*327.971*	*29*
Anket 2	3.907	*-0.028*	*83.5*	*27*
Anket 3	3.600	*0.044*	*41.9*	*36*
Anket 4	3.732	*0.034*	*23.239*	*25*
Anket 5	3.932	*0.166*	*13.729*	*32*

Elde edilen bulgulara göre öğretmen görüşleri ile derlem verileri arasında önemli bir korelasyon yoktur. *.099, -.028, .044, .034, .166* korelasyon değerleri zayıf veya istatistiksel olarak önemsiz korelasyona işaret etmektedir. Bunun sebebi görüşlerine başvurulan öğretmen sayısının nispeten düşük olmasından kaynaklanıyor olabilir. Ancak yine de, bu sonuç derlem verileri ile sezgisel veriler arasında düşük korelasyon bulan diğer yapılmış çalışmalar ile uyumludur. Örneğin, Alderson (2007) çalışmasında *.67,* Schmitt ve Dunham (1999) ise *.53–.65* korelasyon deperini bulmuştur. Brzoza (2018), Lehçe ve İngilizce sözcüklerin objektif sıklık verisini L1 kullanıcılarının sıklık görüşleri ile karşılaştırmış ve iki değişken arasında zayıf korelasyon bulmuştur. Bununla birlikte objektif ve subjektif sıklık verileri arasında önemli korelasyon bulan çalışmalar da olmuştur. Örneğin, Okamato (2015) derleme dayalı sözcük sıklıkları ile anadili konuşanların sıklık görüleri arasındaki ilişkiyi incelemiş ve ikisi arasında yakın bir ilişki bulmuştur. McGee'ye (2008) göre derlem verileri ile sezgisel veriler arasındaki farklılık şaşırtıcı değildir çünkü farklı derlemler sözcük sıklıklarında farklılık gösterebilir ve bu yüzden hem derleme hem de sezgiye dayalı veriler faydalıdır. He ve Godfroid (019) COCA ve COCA Akademik Derlemindeki akademik sözcüklerin sıklığı ile bu sözcüklerin faydası konusundaki öğretmen algıları arasında orta düzey bir korelasyon bulmuştur. Dang ve diğerleri (2022) dört iyi bilinen sözcük listesinin yararlığını öğretmen algıları ve öğrenci sözcük bilgisi kullanarak araştırmıştır ve öğretmen grupları arasında güçlü korelasyon bulmuştur. Birbirlerinden farklılık gösteren bu çalışma sonuçları, sübjektif sezgisel sıklık verileri ve objektif derlem temelli sıklık verileri arasındaki ilişki konusunda daha fazla araştırmaya ihtiyaç duyulduğunu göstermektedir.

Öğretmen görüşlerine başvurulan sözcükler arasında 149 tanesi Likert ölçeğine göre 3'ün altında skor elde etmiştir, yani bu sözcükler öğretmenler tarafından faydalı bulunmamıştır. Bu sözcükler listeden çıkarılmamış, listed "*" ile gösterilmiştir; böylelikle bu sözcüklerin müfredata dahil edilmesi aşamasında daha fazla subjektif veri toplanabilir.

Sonuç

Niceliksel verilerin niteliksel uzman görüşü ile desteklendiği bu çalışmadan bir dizi pedagojik sonuç çıkabilir. Öncelikle, derlemden çıkarılan ve pedagojik olarak uygun sözcük listeleri hem öğretmenler hem de öğrenciler için değerlidir. Materyal geliştirenler ve kitap yazarlarının sözcüksel seçimleri rastlantısal ya da keyfi olabilmektedir. Bu denenle, gerçek kullanıma dayalı bir sözcük listesinin geliştirilmesi ve bunun müfredata entegre edilmesi kaydadeğer katkılar sağlayabilir. Her okul veya öğrenci grubunun kendine özgü amaç ve hedefleri vardır; İngilizce öğrenme sebepleri kaçınılmaz olarak farklılık gösterebilir; bu yüzden, ihtiyaca ve amaca göre geliştirilmiş bir müfredat öğrenci ihtiyaçlarına daha iyi hizöet eder. Türkiye bağlamında yüksek öğrenim düzeyinde, akademik çalışmalarını yapacak düzeyde İngilizce bilgisine sahip olmayan öğrenciler bölümlerine başlamadan önce bir yıl İngilizce hazırlık eğitimi alırlar. Bu programda genel İngilizce öğretilir ancak bölümlerine geçtiklerinde aşamlarına özgü spesifik bir akademik İngilizce ile çalışmalarını yaparlar. Farklı disiplinlerdeki öğrencilerin farklı ihtiyaçları olduğu bu çalışma kapsamında toplanan öğretmen görüşleri ile de desteklenmiştir, ki bu da bu öğrencilerin ihtiyaçlarına yönelik bir sözcük dağarcığının belirlenmesinin faydalı olacağını göstermektedir. Hyland ve Tse (2007) de bu bağlamda şöyle der: "Her bir disiğin veya ders içinde, öğrencilerin çalışmalarını başarmak için ve grup üyesi olarak aktılımda bulunmak için kullanabilecekleri özgün söylem yeterliklerini kazanmaları gerekir (248-249). Bu çalışmada, mühendislik birinci sınıf öğrencilerinin sözcük ihtiyaçlarını belirleyerek kendilerine uygun bir sözcük listesi oluşturulmuş ve buna dayalı olarak bu öğrencilere yönelik bir ders programı hazırlanabileceği öngörülmüştür.

Sıklık, psikolinguistik bir gerçektir ve dil öğretiminde kullanılacak olan her türlü sözcüksel içeriğin önemli bir unsuru olmalıdır. Dilin öğretildiği bağlamdan bağımsız olmak üzere, derlem verilerine başvurmak önemlidir. Derlem temelli, sıklık verilerine dayanan bir yaklaşım benimsemiş olan bu çalışma hem öğrenciler hem de öğretmenler için faydalı bir araç olabilir. Söz konusu öğrencilerin hedef sözcük ihtiyaçları çerçevesinde bir müfredat veya ders programı planlanabilir; ölçme değerlendirme uygulamaları yine bu çerçevede şekillendirilebilir, ve destekleyici materyaller sık kullanılan sözcüklerin olduğu bağlamlar kullanılarak geliştirilebilir. Öğrenciler için sık karşılaştıkları kelimeleri öğrenmek ve hatırlamak daha anlamlı ve kalıcı bir öğrenme sağlayacaktır.

Bu çalışmada oluşturulan sözcük listesinin niteliksel öğretmen görüşünden de faydalanmış olması, öğretilebilirlik özelliğine katkıda buunmaktadır. Öğretmenler tarafından gerekli görülmeyen sözcükler listed işaretlenmiştir. Böylece, liste kullanım amacına ve durumuna göre değişiklikler yapılmaya açık bir temel kaynak niteliğinde olduğundan hem program geliştirme hem de değerlendirme aşamalarında kullanılabilir.

Ayrıca, çalışmada oluşturulan derlem, öğrenciler tarafından kendi öğrenme süreçlerinde kullanılabilir. Hedef bağlamda sıklıkla kullanılan sözcük ve yapıları kendileri keşfedebilir, bu yapıların özellikleri ile ilgili çıkarımlarda bulunabilirler. Ders öğretmenleri yine derlem bağlamından faydalanarak material hazırlayabilir. Boulton'a (2016) göre, derlemler her tür pedagojik material hazırlamada faydalı olabilir.

Son olarak, hedef sözcük listesi her ne kadar mühendislik fakültesi öğrencilerine yönelik hazırlanmış olsa da, fizik, kimya, biyoloji ve matematik bölümlerindeki öğrenciler de bu tür bir Alana özel sözcük listesinden fayda sağlayabilir. Sözcükler, fizik, kimya, biyoloji ve kalkülüs ders kitaplarından oluşturulan derlemden çıkarıldığından, fen alanında herhangi bir disiplinde okuyan öğrenciler veya ders veren öğretmenler bu çalışmadan faydalanabilir. Yine benzer şekilde, oluşrurulan derlem de fen disiplinlerinde özel amaçlı İngilizce programında material geliştirmede kullanılabilir.

Öneriler

Bu çalışma, mühendislik fakültesi birinci sınıf öğrencilerinin hedef sözcük ihtiyaçlarını belirlemeyi ve derlem sıklık verileri ile temel bir envanter oluşturacak bir sözcük listesi geliştirmeyi hedeflemiştir. Çalışma ile ilgili bazı sınırlamalar aşağıda açıklanmıştır.

Çalışmanın niteliksel veri toplama aşamasında yer alan katılımcı sayısı sınırlıdır. Çıkarılan kelime listesindeki sözcükler ile ilgili görüşüne başvurulan öğretmen sayısının daha fazla olması daha sağlam bulgulara ulaşmayı mümkün kılardı. Bu çalışmaya katılacak öğretmenler yalnızca bu dersleri veren öğretmenler ile sınırlı tutulduğundan, küçük bir grup öğretim elemanı ile çalışmak mümkün olmuştur. Ancak, diğer fakültelerde de fen dersleri alanında ders veren öğretim elemanlarından görüş alınarak daha fazla veri toplama imkanı gözden geçirilebilir.

Diğer bir sınırlama, derlem verilerinin yalnızca yazılı metinlere dayalı olmasıdır. Derlemde, söz konusu derslerde kullanılan yazılı materyaller kullanılmıştır. Daha dengeli bir derlem olması bakımından sözlü içeriğe de yer verilmesi önemlidir, ancak bunun oldukça meşakatli ve zaman alan bir iş olduğu da akılda tutulmalıdır. Sözlü ders anlatımlarından örnekler, yazıya dökülerek derleme yüklenebilir, böylece sözlü dil özelliklerini yansıtabilecek veriler de elde edilebilirdi.

Ölçme ve değerlendirme uygulamaları öğrencilerin zorluk çektikleri bir alan olduğundan çalışma için önemli veriler sağlayabilirdi. Öğrencilerin, sınavlarda karşılaştıkları dil ve sözcük özellikleri çalışma için önemli bir kaynak olabilirdi, ancak bu tür verilerin toplanması gizlilik ve güvenlik bakımından risk oluşturduğundan mümkün değildir.

Bir diğer nokta da, oluşturulan hedef sözcük şistesinde yer alan sözcüklerin öğretimi konusunda İngilizce öğretmenlerinin fikri alınarak çalışmanın kapsamı genişletilebilirdi. Böyle bir sözcük listesine dayalı bir ders programı oluşturulacağı varsayılarak hangi sözcüklerin dahil edilebileceği konusunda öğretmen görüşüne başvurulabilirdi. Bu tür bir veri, ders izlencesinin planlamada rehberlik edebilirdi.

Çalışmanın kapsamını genişletecek bir başka nokta ise derlem verilerine dayalı örnek bir ders planıdır. Çıkarılan sözcük listesindeki sözcüklerin nasıl öğretilebileceğine dair bir ders planı oluşturulabilirdi. Derlemde yer alan bağlamlar kullanılarak, okuma metinleri, boşluk doldurma aktiviteleri veya yazma görevleri şeklinde bir içerik hazırlanması söz konusu listeden fayda sağlamak isteyenler için yönlendirici olabilirdi.

Son olarak, oluşturulan Fen Kitapları Sözcük Listesi, üniversitelerin hazırlık programlarında sıklıkla kullanılan İngilizce öğretimi ders kitaplarının sözcük listeleri ile kıyaslanarak, hedef sözcüklerin hangi kitaplarda daha büyük ölçüde öğretildiği incelenebilir.

N. THESIS PERMISSION FORM / TEZ İZİN FORMU

(Please fill out this form on computer. Double click on the boxes to fill them)

ENSTİTÜ / INSTITUTE

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences ☐

Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences ☒

Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics ☐

Enformatik Enstitüsü / Graduate School of Informatics ☐

Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences ☐

YAZARIN / AUTHOR

Soyadı / Surname : Çiçek Demirci
Adı / Name : Şebnem
Bölümü / Department : İngiliz Dili Öğretimi / English Language Teaching

TEZİN ADI / TITLE OF THE THESIS (**İngilizce** / English): TÜRKİYE'DE BİR DEVLET ÜNİVERSİTESİ'NDEKİ MÜHENDİSLİK FAKÜLTESİ ÖĞRENCİLERİNİN SÖZCÜK BİLGİSİ İHTİYAÇLARININ DERLEM TABANLI ANALİZİ / A CORPUS-BASED ANALYSIS OF VOCABULARY NEEDS OF ENGINNERING STUDENTS AT A STATE UNIVERSITY IN TURKEY

TEZİN TÜRÜ / DEGREE: **Yüksek Lisans** / Master ☐ **Doktora** / PhD ☒

1. **Tezin tamamı dünya çapında erişime açılacaktır. /** Release the entire work immediately for access worldwide. ☐

2. **Tez iki yıl süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of **two years. *** ☐

3. **Tez altı ay süreyle erişime kapalı olacaktır.** / Secure the entire work for period of **six months**. ***** ☒

***** *Enstitü Yönetim Kurulu kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir. / A copy of the decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.*

Yazarın imzası / Signature .... **Tarih** / Date ....
(Kütüphaneye teslim ettiğiniz tarih. Elle doldurulacaktır.)
(Library submission date. Please fill out by hand.)

Tezin son sayfasıdır. / This is the last page of the thesis/dissertation.