

Can the Essential Lexicon of Geology be Appropriately Represented in an Intuitively Written EAP Module?

Rahma Al - Mahrooqi

Sultan Qaboos University E-mail: mrahma@squ.edu.om

Saleh Al - Busaidi

Sultan Qaboos University E-mail: asad@squ.edu.om

Jayakaran Mukundan (corresponding author)

Universiti Putra Malaysia E-mail: jayakaranmukundan@yahoo.com

Touran Ahour

Universiti Putra Malaysia E-mail: torahour2@yahoo.com

Ng Yu Jin

Universiti Tenaga Nasional E-mail: yujin@uniten.edu.my

Received: March 21, 2011 Accepted: April 9, 2011 doi:10.5539/ells.v1n1p50

Abstract

This study aimed to find out to what extent an intuitively developed ESP module for Science majors, taught at Sultan Qaboos University is appropriately written in terms of lexicon when compared to a core Geology textbook. The module was developed based on key topics which appeared in the Geology textbook. This study will only be evaluating vocabulary and will not be looking at other aspects of material evaluation. The digitized pages of the Module (LANC 2050) and the Essentials of Geology Textbook were loaded into the Software of WordSmith 5.0 for analysis. The results revealed low percentage word coverage in the module as compared with the limited pedagogical word list that was developed for Geology. In addition, the high density and low consistency ratios for the module as compared to the textbook indicated the compactness of the module for teaching purposes. A very low percentage of the technical words that were related to Geology was discovered in the Module (LANC 2050).

Keywords: EAP, ESP, Module, Textbook, Geology, Science, WordSmith

1. Introduction

Knowledge of vocabulary is an important matter for the learning of a second or foreign language. Its importance increases when it comes to the areas of English for Specific Purposes (ESP) and English for Academic Purposes (EAP). For developing this knowledge, much more attention should be paid to the explicit and implicit ways of teaching the vocabulary. In some situations, direct teaching and learning are required, while, in other situations, indirect ways of learning, such as incidental learning in extensive reading, would suffice (Sutarsyah, Nation & Kennedy, 1994).

Encountering a word for enough number of times throughout the textbook would reflect the high frequency of that word which, in turn, would show its importance in learning and teaching. For example, knowledge of first 2000 words of general service list (GSL) (West, 1953) together with the academic word list (AWL) (Coxhead, 2000) would provide, according to Nation (1990, 2001), about 95% of the vocabulary required for the reading of academic texts essential for the ESL/EFL students' learning. A corpus-based content analysis of a textbook can make this decision easier for the teacher (Fox, 1998). However, the basic criterion in learning and teaching the vocabulary is the underlying purpose(s) on which they are based, i.e., whether the aim is to learn the words for general use (GE), academic use (EAP) or specialized use (ESP).

1.1 Background of vocabulary learning and teaching

Learning vocabulary in the ESL/EFL situations is one of the specific goals of learning in the language classroom. In the language classrooms and in a well-designed course, four important strands should be present (Nation, 2001). These strands approach learning new language items from four perspectives including: (a) comprehensible meaning-focused input through the listening and reading where about 95% of the running words are familiar to the learners. In other words, for learning to occur very few unknown words should be present in the input. (b) language-focused learning or form-focused instruction (Ellis, 1990) in which the focus is on the deliberate teaching

and learning of the language items. If the goal of the course is, for instance, teaching and learning vocabulary, the direct teaching and learning of vocabulary is needed. (c) meaning-focused output through the speaking and writing where the learners can strengthen their knowledge of previously met vocabulary; (d) fluency-focused development through activities in which they use and practice the items they already know. According to Nation (2001), in a language course approximately the same amount of time (i.e., 25% of the learning time) should be spent on any of these four strands.

In addition, Nation (2001) states that there are four types of vocabularies: the high frequency words, academic words, technical words and low-frequency words. High frequency words are words which are commonly used in a language textbook and they require high amount of attention by the teacher and students (Nation & Waring 1997; Fox, 1998). In the development of vocabulary list in different disciplines, priority has to be given to the General Service List (GSL) (West, 1953) comprising words that are of high frequency in all kinds of texts and which make up the most frequent 2000 words of English (Nation, 2001; Coxhead, 2000; Nation & Waring, 1997). The GSL is a list that is irreplaceable as it covers up to 87% of any academic texts (Nation, 2001; 1990; Coxhead, 2000) and teachers and learners alike should focus on comprehending the words in order to cope with academic reading materials at higher rates.

Besides academic words, technical words are important to learners who need the vocabulary to help in planning and achieving their specific goals in language use. Technical words are specialised vocabulary developed from specialised corpus, words which qualify in terms of frequency and relevance (Nation, 2001). "Strictly technical words are characterized by the absence of exact synonyms, resistance to semantic change, and a very narrow range; e.g., words such as urethane or vulcanise" (Mudraya, 2006, p.239). These words have low frequency but comprise a large number of words (Nation, 1990; 2001).

In academic texts, especially in specific fields, the frequency and coverage of academic and technical words are higher than in general English texts (Sutarsyah et al., 1994). Being ignorant about these words can pose challenging experiences to students who wish to learn more about field-specific vocabulary which can initiate the socialization of students into their discourse communities. Failing to be associated to their discourse community can hinder productive progress among students academically. Academic vocabulary occurs frequently in most kinds of academic texts; however, their frequency is lower than general service vocabulary (Worthington & Nation, 1996; Xue & Nation, 1984).

Having good knowledge of academic vocabulary has been emphasized for success at higher levels of education (Coxhead, 2000; Corson, 1997). Different corpus-based studies have been conducted with regard to the vocabulary of the texts in different areas and disciplines. Worthington and Nation (1996), for example, have investigated the feasibility of the texts that are used to sequence the target vocabulary of a course. Some studies have explored the technical vocabulary in specialized texts such as in anatomy and applied linguistics texts (Chung & Nation, 2003) or in the language of mathematics, science and English textbooks (Mukundan & Menon, 2005; 2007a; 2007b).

1.2 The need for ESP materials in the university

At Sultan Qaboos University, all science based colleges and specializations are taught in English, so proficiency in English has been deemed important for students' success in their majors since the inception of this university. From its start, the university has realized that almost all its students majoring in science based colleges might be in need for content area English tuition to prepare them for their majors and so it adopted ESP and EAP as the basis for its English modules. To teach these it established the Language Centre, which has been operating since 1986. In these modules, English is taught using context of the content related to students' majors. The learning outcomes of ESP are language-based (grammar, sentence structure, word forms, coherence, cohesion, the four skills, subject-area terminology, spelling and punctuation). The rationale is that EFL students are expected to perform at the same level in their university courses as native speakers of English, who commence their university education with a full command of general English, and who would have spent 12 years studying all subjects in English within a larger community in which English is the only medium of communication. The English for Science courses taught to College of Science students are credit courses designed with some input from the College of Science. Needs analyses were carried out informally and the colleges provided some input on how the English courses need to be structured and what skills or projects they should include. However, in terms of vocabulary, it was impressionistically included in the module by the college representatives based on their experience. One major drawback with the English for Science courses is that their content targeted general science because they were taught to all science majors. Therefore, the vocabulary and the topics might not match anyone specialization in particular. No thorough analysis of the science books was conducted.

1.3 Purpose of the study

The aim of this study is to compare the first of the two Modules (LANC 2050) which was intuitively written for

EAP with the Essentials of Geology Textbook as the Reference Corpus of the study for ESP course. The specific objectives of the study are: (a) to find out the similarities and differences of the words in the Module (LANC 2050) and the Essentials of Geology Textbook, their frequency range, density ratio, consistency ratio, and types of words; (b) to determine whether the Module (LANC 2050) can prepare the students for the ESP course in Geology.

1.4 Research questions

The two corpora were examined to seek the possible answers to the following questions:

- (1). To what extent are the 1,000 most frequently used specialised content words in the Module (LANC 2050) and the Essentials of Geology Textbook different?
- (2). With the creation of a limited pedagogical word list, to what extent do the words in the word list represent those in the EAP Module (LANC 2050)?

2. Methodology

This study compares two corpora, the Module (LANC 2050) and the Essentials of Geology Textbook. The Module (LANC 2050) is taught in the English for Science Course, the completion of which is one of the requirements for the students of Science in the College of Science in Sultan Qaboos University (SQU). It consists of six units with different sections covering topics related to *The Earth, Space, Matter and Materials, Technology, Biotechnology*, with the last unit as *Review and Practice*. The Module (LANC 2050) units are compared with eleven selected chapters of the Essentials of Geology Textbook that includes different topics such as *Introduction to Geology, Minerals: Building blocks of rocks, Igneous rocks, Volcanoes and other Igneous activity, Weathering and soils, Sedimentary rocks, Metamorphic rocks, Earthquakes and Earth's interior, Plate Tectonics, Geologic Time, and Earth history: A brief summary*. Most of the topics in the textbooks were used as the basis for the development of the module. The 1,000 most frequently used content words from both the Essentials of Geology Textbook and the Module (LANC 2050) will be compared to find out to the differences.

The keyword lists created in this study are based on the study carried out by Menon (2009), who created 'keyword lists' using the Word Smith software. Harvesting or building a word list can be done by comparing with a larger corpus which functions as a reference corpus (Nation, 2001). Thus, the limited pedagogical corpora for both the textbook and the module are compared with the British National Corpus (BNC), as a well-established corpus for word list building (Menon, 2009). WordSmith 5.0 was used for analyzing the data. In order to use this software all the related pages of the module and the textbook were digitized. These pages were scanned and converted into the text files. Then, the text files were fed into the software for further analysis.

3. Results and Discussion

The comparison of the output of WordSmith analysis for Module (LANC 2050) units and the Essentials of Geology chapters (see Table 1) revealed the number of running words, or tokens, in the textbook chapters (128 031) compared to the module (51 993). The types of the words were also higher in the textbook (8098) than those in the module (6006). The density ratio, that is the proportion of types to tokens, indicates the high intensity or compactness of the materials (Mukundan, 2007; Nation, 1990). The module has a density ratio reading of 11.5 compared to 6.4 of the text. It signifies that the density of the words in the Module (LANC 2050) is higher than the density of words in the Essentials of Geology Textbook chapters. In contrast, the consistency ratio, that is the proportion of tokens to types, is higher for the Essentials of Geology chapters (15.53) as compared to that of the Module (LANC 2050) units (8.71). It reflects the consistency on the introduction of new words in the Essentials of Geology Textbook and Module (LANC 2050). In other words, each new word is introduced after about 16 words interval in the textbook while its introduction is in every 9-word interval for the module. In this regard, the module is considered to be of high density with lower level of consistency interval. The compactness of the Module (LANC 2050) compared to the Essentials of Geology, may make the learning and acquiring of the new words rather difficult to the students. It requires the teacher to focus the attention of the students to these words, although the repetition of the words can help in this regard. Table 1 presents a summary of these findings:

Table 1:

The same pattern was found across the units and chapters of the Module (LANC 2050) and the Essentials of Geology Textbook (see Table 2). It was revealed that the mean density ratio of the module units (23) is much higher than that of the textbook chapters (17.91). It implies that, on average, the occurrence of the different words is more in the module compared to the textbook. This pattern would make the acquisition and learning of the new language more difficult to the students. Similarly, the mean consistency ratio or the interval unit of words for the introduction of each new word is lower (4.41) in the module compared to the textbook (5.67). This low interval in the module

indicates that every new word is introduced after four words, while in the textbook the introduction of each new word is after every six words. Therefore, learning new words would be easier in the pattern revealed in the textbook rather than the module.

Table 2:

In this regard, comparison of the density and consistency ratios across the units of the Module (LANC 2050) shows that Units 2 and 1 would be easier to learn followed by Units 3, 5, 4, and 6, respectively. Likewise, the Essentials of Geology Textbook chapters from the easiest to the most difficult would be chapters 3, 4, 16, 2, 6, 18, 15, 5, 7, 19, and 1, respectively. The best sequence for introducing the units of the module and the chapters of the textbook can be based on the difficulty level of the units or chapters. The more difficult ones would be better presented at the end of the books while bringing the easiest ones to the front. This would facilitate the acquisition of the new language more effectively.

3.1 Frequency range in the Module (LANC 2050) and the Essentials of Geology Textbook

Table 3 and Table 4 demonstrate the frequency range of the words in the Module (LANC 2050) and the Essentials of Geology Textbook. The tables show high numbers of the words belong to the low frequency words in both the module and the textbook. Low frequency words are words that appear only in the frequency range of 1 to 40. On the other hand, the high frequency words (words that appear more than 40 times) are very few in both of them. The number of the words for all the frequency range is more in the textbook than in the module. This is due to the high number of total types in the textbook.

Table 3:

Table 4:

The frequency range shows that very few words are repeated with high frequency, especially in the module. For instance, when 46 words are repeated between the range of 101-1000 times in the module, almost three times more words (149) are repeated in this range in the textbook. When the words are repeated frequently, their acquisition would be easier for the students (Mukundan, 2007; 2007a; 2007b, Nation, 1990; 2001). The same pattern is observed in the frequency range of both the module and the textbook. In order to differentiate what kinds of words comprise the module and the textbook, the researchers selected fifty words from the high frequency a word in their word lists (see Table 5). The selected words are the fifty most frequently used words in both the module and textbook. More function words are found in the module rather than in the textbook.

Table 5:

3.2 Comparability of the words in the Module (LANC 2050) and the Essentials of Geology Textbook

Among the wordlists for the two corpora, 3345 words are found to be comparable or overlap in both the module and the textbook (see Table 6). This comprises about 56% and 40% of the total types of the module and the textbook respectively. The rest includes the words that are different in both corpora. In other words, about 44% of the total types of the module are not found in the textbook. In the same vein, about 60% of the total types of the textbook are not found in the module.

Table 6:

Many of the high frequency content words in the textbook also occur in the module corpus, but with a much lower frequency. To further investigate the types of similar or different words introduced in the module as opposed to the textbook, the 1000 most frequently used specialised content words from both the textbook and module were identified. The specialised content words are words that are outside the GSL and the function words like 'AND', 'THE' and 'A' were removed from the lists as well. The definition used in this study is adapted from Coxhead's (2000) development of Academic Word List (AWL). The proper nouns and abbreviations identified are also removed from the lists. The specialised content words can also be regarded as words which appear very frequently in field-specific texts (Chung & Nation, 2003; Nation, 2001).

The Module (LANC 2050) is supposed to be tailored to the needs of students in the field of geology. Hence, the vocabulary introduced in the module should be almost similar to that of the textbook, especially the most frequently used content words. Appendix A indicates the similar content words which can be found in both the textbook and the module. From the result of the analysis, only 30% of the content words from the module are similar to the content words in the textbooks (considering the top 1,000 most frequently used content words of both materials). Hence, it proves that the module is done intuitively, not considering the important aspect of introduction of specialised geology vocabulary. The results should yield at least a higher percentage if the module is developed from data from

a corpus. After the analysis, the results revealed that the majority of the words that are in the textbook but are not found in the module are related to the technical geology words such as *earthquake, weathering, metamorphic, ions, eruption, volcano, shale, ore, intrusive, slate, richter, cinder, mica, magmatic, pumice, muscovite, chamber*.

3.3 Comparison of Key Words Derived from 'Keyword' Lists (Keyness)

Using Word Smith 5.0, the 'Keyword' function was utilized to generate the keywords list, after comparing with a larger corpus. Nation (2001) states one of the ways to create a word list is through the comparison of the study corpus with a much bigger corpus. The 'Keyword' function in the WordSmith 5.0 programme enables the researchers to create a word list by comparing the target corpora with a reference corpus. Menon (2009) in her study had created keyword lists for the field of English for Science and Technology. Menon (2009) also justified that using the British National Corpus (BNC) as a reference corpus is valid as the size of BNC is at least 5 times or more than the study corpus. In this study, the BNC is definitely much larger (more than five times) than the pedagogical module and the Geology corpora. The size of the BNC is almost 100 million running words as compared with the limited pedagogical corpora of the Module (LANC 2050) (51 993 tokens) and the Essentials of Geology Textbook (128 031 tokens). Hence, using the BNC as the reference corpus in the study is ideal. Only the words with the positive 'keyness' are included in the lists. The analysis results from the Word Smith 5.0 software yield the significant keywords of the Module (LANC 2050) (531 words) and the Essentials of Geology Textbook (1425 words). The words were not categorized into word families as learning words in the lexico-grammar context can yield better results in students' acquisition of the vocabulary (Ward, 1999; 2009; Mudraya, 2006). Out of the 531 keywords found in the Module (LANC 2050), only 155 words (29%) were found to be common with the 1425 keywords of the Geology corpus. The 155 common key words from both the module and the textbook are stipulated in Appendix B. The result shows that the intuitively produced module does not cover most of the essential vocabulary needed when students need to enrol in a Geology course. Knowing more field-specific words enables learners to associate into their discourse community much more effectively (Menon, 2009), hence it would be best for module developers to be informed by data regarding lexicon from analysis of field-specific corpora.

4. Implication and Conclusion

As a result of this research, we now have a limited pedagogic corpus of 1425 words. This would be useful when the Module (LANC 2050) is revised to make it resemble Geology texts commonly encountered by students. This study was conducted to find out the appropriateness of the Module (LANC 2050) as a preparatory ESP course for Geology majors. For this, the corpus-based content analysis of the words for each was explicated and the results were compared and discussed considering different criteria including the frequency of the words, density and consistency ratio, keyness, and coverage of the words.

The evaluation of the in-house produced Module (LANC 2050) revealed that the introduction of essential technical or field-specific words was not effectively executed throughout the module. Hence, it is not able to prepare students for assimilating into their majors. Most of the essential words found in the Essentials of Geology Textbook were non-existent in the Module (LANC 2050). Despite the meticulous process of developing a module, material writers should always take into consideration the essential vocabulary needed by learners which could be done by creating a field-specific corpus. Learners truly need the field-specific vocabulary to function well in their course and prepare them for their discourse community.

References

- Chen, Q., & Ge, G. (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, 502-514, doi:10.1016/j.esp.2007.04.003, <http://dx.doi.org/10.1016/j.esp.2007.04.003>
- Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language*, 15(2), 103-116. CiteSeer. Retrieved from <http://nflrc.hawaii.edu/rfl/October2003/chung/chung.pdf>
- Corson, D. (1997). The learning and use of academic English words. *Language Learning*, 47, 671-718, DOI: 10.1111/0023-8333.00025, <http://dx.doi.org/10.1111/0023-8333.00025>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34 (2), 213-238.
- Cowan, J R. (1974). Lexical and syntactic research for the design of EFL reading materials. *TESOL Quarterly*, 8(4): 389-399.
- Ellis, R. (1990). *Instructed second language acquisition*. Basil Blackwell: Oxford.
- Fox, G. (1998). Using corpora data in the classroom. In B.Tomlinson (Ed.), *Materials development in language teaching* (pp.25-43). Cambridge: Cambridge University Press.

- Hyland, K., & Tse, P. (2007). Is there an “academic vocabulary”? *TESOL Quarterly*, 41(2), 235–253.
- Lutgens, F.K. & Tarbuck, E.J. (2003). *Essentials of Geology* (8th Ed.). New Jersey: Prentice Hall.
- Menon, S. (2009). *Corpus-based analysis of lexical patterns in Malaysian secondary school Science and English for Science and Technology textbooks*. Unpublished doctoral dissertation, Serdang: Universiti Putra Malaysia, Malaysia.
- Mudraya, O. (2006). Engineering English: a lexical frequency instructional model. *English for Specific Purpose*, 25, 235-256. doi:10.1016/j.esp.2005.05.002, <http://dx.doi.org/10.1016/j.esp.2005.05.002>
- Mukundan, J. (2007). Irregularities in vocabulary load and distribution in same level textbooks written by different writers. *Indonesian Journal of English Language Teaching*, 3 (1), 99-118.
- Mukundan, J., & Menon, S. (2007a). The language of Mathematics, Science and English Language Textbooks: A comparative Corpus Based Study. In J. Mukundan, and S. Menon (Eds.), *ELT matters 4: Developments in English language learning and teaching* (pp.244-272). Selangor: Universiti Putra Malaysia Press.
- Mukundan, J., & Menon, S. (2007b). Lexical similarities and differences in the Mathematics, Science and English Language Textbooks. *Kata*, 9 (2), 91-111.
- Nation, P. (1990). *Teaching and learning vocabulary*. Boston, Massachusetts: Heinle & Heinle Publishers.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, P., & Kyongho, H. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System*, 23 (1), 35-41, doi:10.1016/0346-251X(94)00050-G, [http://dx.doi.org/10.1016/0346-251X\(94\)00050-G](http://dx.doi.org/10.1016/0346-251X(94)00050-G)
- Scott, M. (2008). *WordSmith Tools (version 5)*, Liverpool: Lexical Analysis Software.
- Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based case study. *RELC Journal*, 25(2): 34-50, doi: 10.1177/003368829402500203, <http://dx.doi.org/10.1177/003368829402500203>
- Trimble, L. (1985). *English for Science and technology: A discourse approach*. New York: Cambridge University Press.
- Wang, J., Liang, S.L. & Ge, G.C. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, 27, 442-458, doi:10.1016/j.esp.2008.05.003, <http://dx.doi.org/10.1016/j.esp.2008.05.003>
- Ward, J (1999). How large a vocabulary do EAP engineering students need? *Reading in a Foreign Language*, 12(2), 309-323.
- Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purpose*, 28, 170-182, doi:10.1016/j.esp.2009.04.001
- West, M. (1953a). *A General Service List of English Words*. London: Longman, Green & Co.
- Willis, J. (1998). Concordances in the classroom without a computer: Assembling and exploiting concordances of common words. In B.Tomlinson (Ed.), *Materials development in language teaching* (pp.44-67). Cambridge: Cambridge University Press.
- Worthington, D., & Nation, P. (1996). Using texts to sequence the introduction of new vocabulary in an EAP course. *RELC Journal*, 27(2), 1-11, doi: 10.1177/003368829602700201, <http://dx.doi.org/10.1177/003368829602700201>.
- Xue, G., & Nation, P. (1984). A university word list. *Language Learning and Communication*, 3, 215-229.

Table 1. Tokens, Types, Density Ratio, and Consistency Ratio in Geology Module (LANC 2050) and Essentials of Geology Textbook

	Module (LANC 2050)	Essentials of Geology Textbook
Tokens	51 993	128 031
Types	5969	8243
Density Ratio	11.5	6.4
Consistency Ratio	8.71	15.53

Table 2. Number of Tokens, Types, Density Ratio, and Consistency Ratio in the Module (LANC 2050) Units and Essentials of Geology Textbook Chapters

Source	Tokens (Total words)	Types (Different words)	Density Ratio (Type/token)	Consistency Ratio (Token/type)
Module (LANC 2050)				
Unit				
1	8460	1730	20	4.89
2	10 927	2194	20	4.98
3	10 320	2215	21	4.66
4	8032	1983	25	4.05
5	7108	1674	24	4.25
6	7146	1969	28	3.63
Mean	8665	1961	23	4.41
Essentials of Geology Textbook				
Chapter				
1	10 201	2077	20	4.91
2	10 732	1782	17	6.02
3	10 924	1717	16	6.36
4	15 924	2561	16	6.22
5	10 224	1914	19	5.34
6	11 962	2022	17	5.92
7	8066	1527	19	5.28
15	13 654	2426	18	5.63
16	13 752	2279	17	6.03
18	11 309	2006	18	5.64
19	11 283	2255	20	5.00
Mean	11 639	2051	17.91	5.67

Table 3. The Distribution of Low Frequency Words in the Module (LANC 2050) and the Essentials of Geology Textbook

Frequency	Number of Words in the Module (LANC 2050)	Number of Words in the Essentials of Geology Textbook
1	2,630	3,199
2	1,011	1,175
3	499	694
4	338	437
5	246	330
6	185	223
7	136	204
8	105	176
9	103	126
10	70	114
11-20	358	717
21-30	115	281
31-40	48	170

Table 4. The Distribution of High Frequency Words in the Module (LANC 2050) and the Essentials of Geology Textbook

Frequency Range	Number of Words in the Module (LANC 2050)	Number of Words in the Essentials of Geology Textbook
41-50	36	96
51-60	19	83
61-70	17	44
71-80	13	33
81-90	17	34
91-100	8	25
101 – 1000	46	149
1001 – 3600	6	8
3700 - 8200	0	2

Table 5. The first fifty high frequency words in the Module (LANC 2050) and the Essentials of Geology Textbook

No.	Module	Textbook	No.	Module	Textbook
1	the	the	26	will	have
2	of	of	27	your	most
3	and	and	28	one	minerals
4	a	a	29	at	when
5	to	in	30	not	into
6	in	is	31	he	was
7	is	that	32	task	they
8	that	are	33	write	earth
9	are	as	34	earth	can
10	it	by	35	but	more
11	on	rocks	36	what	not
12	for	from	37	when	time
13	as	this	38	text	magma
14	from	figure	39	these	than
15	you	rock	40	about	other
16	be	for	41	his	because
17	by	with	42	which	some
18	was	it	43	notes	one
19	or	an	44	other	plate
20	have	or	45	has	earth's
21	with	these	46	were	sedimentary
22	can	be	47	been	called
23	they	on	48	used	mineral
24	this	at	49	there	years
25	an	which	50	space	their

Table 6: The Number of Similar and Different Types of Words and Their Percentage in the Module (LANC 2050) and the Essentials of Geology Textbook

Wordlist	Similar words (%)	Different words (%)	Total types (%)
Module	3345 (56%)	2624 (44%)	5969 (100%)
Textbook	3345 (41%)	4898 (59%)	8243 (100%)

Appendices/Supplementary Materials:

Appendix A: Similar Specialised Content Words in Both Corpora

Words	Frequency in the Module (LANC 2050)	Frequency in the Essentials of Geology Textbook
ACCUMULATE	5	17
ACCUMULATED	4	7
ACCURATE	4	9
ACID	6	32
AFFECT	2	8
AFFECTED	4	6
AFFECTS	2	6
AFRICA	3	42
AID	2	20
ALGAE	2	12
ALPHA	8	15
ALTERED	2	17
ALUMINUM	2	41
AMERICA	5	80
APPARENT	3	12
APPROPRIATE	18	8
APPROXIMATELY	13	14
AREA	15	78
AREAS	11	52
ATMOSPHERE	56	80
ATOM	16	31
ATOMIC	18	46
ATOMS	23	72
ATTRIBUTED	2	6
AVAILABLE	7	23
AXIS	2	14

BACTERIA	9	11
BASALT	5	41
BEDROCK	3	14
BETA	3	13
BIOLOGICAL	10	25
BOMBS	2	10
BRIEF	18	18
CAPABLE	2	11
CARBON	21	91
CASTS	2	8
CATEGORY	2	6
CHEMICAL	44	210
CLIMATE	7	43
CLIMATIC	2	12
CLUES	4	17
COM	12	30
COMPACT	2	6
COMPLEX	9	47
COMPONENTS	7	28
COMPOUNDS	11	23
CONCEPT	8	8
CONCEPTS	2	12
CONCLUDE	3	8
CONSIDERABLE	4	19
CONSIST	3	28
CONSISTING	4	20
CONSISTS	4	38
CONSTANT	6	8
CONSTANTLY	4	7
CONSTRUCTION	6	12
CONTACT	5	28
CONTINENTAL	16	167
CONTINENTS	5	77
CONVECTION	3	23
CORE	14	70
COURTESY	3	20
CRATER	13	30
CREATE	17	22
CREATED	27	50
CREATING	5	17
CRETACEOUS	6	19
CRUST	31	185
CRYSTAL	8	33
CRYSTALLINE	4	40
CRYSTALS	11	98

CUBIC	4	12
CURRENTS	3	11
DATA	30	39
DEBRIS	3	31
DECADES	5	10
DEFINITE	4	11
DENSE	3	35
DENSITY	6	17
DEPOSITED	10	72
DEPOSITION	3	36
DEPOSITS	6	113
DERIVED	11	27
DESIGNED	9	7
DIAGRAM	3	13
DIAGRAMS	3	6
DIAMETER	11	10
DINOSAURS	48	36
DIOXIDE	10	41
DISSOLVED	3	37
DISTINCT	4	17
DRIFT	7	32
EARTH'S	3	268
ELECTRONS	6	62
ELEMENT	5	33
ELEMENTS	22	81
EMISSION	6	13
EMIT	3	14
EMITTED	5	16
ENERGY	48	94
ENGLAND	7	7
ENVIRONMENT	18	81
ENVIRONMENTAL	5	22
EQUIVALENT	3	14
ERA	5	57
ERAS	4	11
ESTABLISH	6	11
ESTIMATED	3	18
EUROPE	5	18
EVENTUALLY	9	69
EVIDENCE	29	81
EVIDENT	3	6
EVOLUTION	3	17
EVOLVED	5	24
EXPAND	9	13
EXPOSED	7	53

EXPOSURE	9	14
EXTINCT	7	14
EXTINCTION	12	23
EXTRACT	9	6
EXTRACTED	4	7
FEATURES	5	40
FINAL	8	7
FINALLY	5	11
FLUID	4	50
FLUIDS	5	27
FOCUS	3	50
FOSSIL	23	66
FOSSILS	40	98
FRAGMENTS	6	73
FUEL	4	7
FUNDAMENTAL	4	9
FURTHERMORE	4	14
GASEOUS	3	12
GENERATE	9	45
GENERATED	3	41
GEOLOGIC	3	134
GEOLOGICAL	19	42
GEOLOGIST	10	15
GEOLOGISTS	13	76
GEOLOGY	12	101
GIANT	7	7
GLOBAL	3	20
GLOBE	6	10
GRAVITY	20	30
GROUNDWATER	3	9
HUGE	12	26
HYDROGEN	13	15
IDENTIFIED	12	29
IDENTIFY	10	13
IDENTIFYING	5	7
IGNEOUS	8	236
IMAGE	3	7
IMPACT	18	42
INDEX	3	10
INDIVIDUAL	7	22
INITIAL	10	7
INTERNET	16	22
INVOLVED	3	8
ISOTOPES	4	29
KILOMETERS	5	155

LABORATORY	13	8
LANDFORMS	4	10
LAVA	7	150
LAVAS	7	38
LAYER	21	85
LAYERS	18	106
LIMESTONE	6	58
LINKED	4	7
LOCATION	3	18
LOCATIONS	4	22
MAGMA	4	303
MAGNETIC	13	53
MAJOR	6	124
MANTLE	6	195
MARGIN	3	17
MATRIX	3	11
MATURE	4	9
METALLIC	3	30
METEORITES	21	11
METER	3	11
METERS	10	81
METHOD	5	22
METHODS	9	18
MICROSCOPIC	10	9
MINERALOGY	3	12
MODIFIED	13	7
MOLTEN	10	34
NEGATIVELY	3	7
NEUTRAL	4	14
NEUTRON	4	9
NEUTRONS	12	25
NITROGEN	19	13
NORMAL	4	15
NUCLEI	10	22
NUCLEUS	21	48
OBTAINED	7	16
OBVIOUS	3	15
OCCUR	4	109
OCCURRED	9	63
OCCURS	4	88
OCEANIC	15	138
ORGANIC	3	49
ORGANISMS	21	67
OXYGEN	22	96
PARALLEL	3	29

PERCENT	5	87
PERIOD	17	108
PERIODS	8	35
PHENOMENA	7	7
PHYSICAL	10	64
PLANET	49	51
PLANETS	33	20
POSITIVE	7	13
POSITIVELY	4	8
PREHISTORIC	3	8
PRIMARY	7	31
PRIMITIVE	3	17
PROCESS	23	95
PROCESSES	6	140
PROTON	3	14
PROTONS	6	38
QUARTZ	5	100
RADIATION	24	8
RADIOACTIVE	12	46
RADIOACTIVITY	6	19
RANGE	5	59
REACTION	10	33
REACTIONS	5	11
REGION	7	45
REGIONS	6	54
RELEASE	3	7
RELEASED	12	22
REMOVAL	3	7
REMOVED	6	31
REQUIRE	4	7
RESEARCH	45	7
RESEARCHERS	4	29
RESERVOIR	11	9
RESISTANT	7	18
RIGID	3	19
ROLE	5	23
SECTION	36	47
SEDIMENTARY	17	266
SEDIMENTS	3	50
SEISMIC	4	54
SEQUENCE	7	26
SERIES	4	43
SIGNIFICANT	5	39
SILICA	5	99
SILICON	6	57

SIMILAR	15	75
SITE	7	32
SITES	6	15
SODIUM	3	52
SOLAR	26	18
SOURCE	48	59
SOURCES	20	19
SPECIES	32	12
SPECIFIC	13	33
SPHERICAL	3	10
STABLE	6	50
STRATA	6	46
STRUCTURE	21	107
STRUCTURES	4	112
SUCCESSION	5	8
SUCCESSIVE	3	12
SUMMARY	58	24
SURVIVE	9	7
SUSPENDED	3	8
TASK	154	7
TECHNIQUES	9	13
TECTONIC	10	11
TECTONICS	4	65
TERTIARY	5	20
THEORIES	11	8
THEORY	64	43
THERMAL	4	11
TINY	7	23
TOPIC	10	7
TRACE	4	16
TRACES	3	8
TRANSFER	4	7
TRANSPORT	4	23
TRANSPORTED	3	22
UNDERGROUND	3	14
UNSTABLE	6	19
URANIUM	14	28
VARIES	3	8
VAST	6	33
VEGETATION	4	24
VELOCITY	3	8
VISCOSITY	4	15
VITAL	5	8
VOLCANIC	5	235
VOLCANOES	13	105

VOLUME	3	33
WEAKER	3	7
WEB	5	39
WEBSITE	6	11
WEGENER	3	25
WIDESPREAD	3	13
WORLDWIDE	3	15

Appendix B: Similar Common Keywords Found in Both the Module LANC (2050) and the Essentials of Geology Textbook

ALPHA	COMPOSED	DISTANCES
ANIMALS	COMPOSITION	EACH
APATOSAURUS	COMPOUNDS	EARTH
ARE	CONTAIN	ELECTRICAL
ATMOSPHERE	CONTAINS	ELEMENTS
ATOM	CONTINENTAL	EMITTED
ATOMIC	COOLED	ENERGY
ATOMS	CORE	ERAS
BASALT	CRATER	EXAMPLES
BELOW	CREATED	EXTINCT
BILLION	CRETACEOUS	EXTINCTION
BIOLOGICAL	CRUST	FORM
CALLED	CRYSTALS	FORMED
CARBON	DECAY	FORMS
CHARACTERISTICS	DEPOSITED	FOSSIL
CHEMICAL	DIFFERENT	FOSSILS
CLAY	DINOSAURS	FOUND
COM	DIOXIDE	FROM
GAS	MELTS	RADIOACTIVITY
GASES	METALS	RESISTANT
GEOLOGIC	METEORITES	ROCK
GEOLOGICAL	METERS	ROCKS
GEOLOGIST	MICROSCOPIC	SAMPLES
GEOLOGISTS	MILLION	SCIENTIFIC
GEOLOGY	MINERAL	SCIENTISTS
GRAVITY	MINERALS	SEA
HEAT	MOLTEN	SEDIMENTARY
HEATED	MOTION	SEVERAL
HELIUM	NATURAL	SHAPE
HTTP	NEUTRONS	SILICA
HUMANS	NITROGEN	SOIL
HYDROGEN	NUCLEI	SOLAR
ICE	NUCLEUS	SOLID
IGNEOUS	OCEANIC	SOURCE
IMPACT	OF	STRATA
INTERNET	ORBITING	SUMMARY
IRON	ORGANISMS	SURFACE
ISOTOPES	ORIGIN	TECTONIC
KILOMETERS	OXYGEN	TEMPERATURE
KNOWN	PALEONTOLOGISTS	TEMPERATURES
LANDFORMS	PARTICLES	THEORY
LAVA	PLANET	THESE
LAVAS	PLANETS	THORIUM
LAYER	PLANT	TYPES
LAYERS	PLANTS	URANIUM
LIFE	PLATE	VAPOR
LIQUID	PLATES	VOLCANOES
MAGNETIC	PRODUCE	WATER
MANTLE	PROPERTIES	WEBSITE
MASS	PROTONS	WEGENER
MATERIAL	QUARTZ	YEARS
MATERIALS	RADIOACTIVE	