Sultan Qaboos University
Journal of Arts & Social Sciences

جامعة السلطان قابوس
مجلة الآداب والعلوم الاجتماعية

An Analysis of Lexicon in Omani English Language Teaching: Challenges from Teaching and Learning Perspectives

Rahma Al-Mahrooqi

Deputy Vice-Chancellor for
Postgraduate Studies and Research
Sultan Qaboos University
ralmahrooqi@gmail.com

Faisal Al-Maamari

Language Centre
Sultan Qaboos University
faisalf@squ.edu.om

Christopher Denman

Office of the Deputy Vice-Chancellor
for Postgraduate Studies and Research
Sultan Qaboos University
denman@squ.edu.om

Paper received: 20/09/2015
Paper accepted: 03/04/2016

An analysis of lexicon in Omani English language teaching: challenges from teaching and learning perspectives

Rahma Al-Mahrooqi, Faisal Al-Maamari & Christopher Denman

Abstract

Textbooks are indispensable in most formal education settings. However, many teachers believe weaknesses in textbooks can result in ineffective learning and teaching. This research uses a corpus-based approach to identify issues related to vocabulary with a particular focus on English textbooks in use in Omani schools. WordSmith Tools and RANGE concordance software are used to analyze these textbooks. Findings indicate that there are irregularities in vocabulary loading and that the types, tokens, density and consistency ratio show inconsistencies. Moreover, in terms of the coverage of GSL and AWL word lists, the textbooks fall short of the ideal range of 95% to 98% coverage suggested by Nation (1990, 2001). In fact, vocabulary from large percentages of the texts are not found in any of the essential word lists. The paper suggests recommendations for consistent analysis and guidance in analyzing textbooks in order to improve their quality in terms of vocabulary introduction and loading.

Keywords: lexicon, vocabulary loading, corpus-based approach, GSL/AWL.

دراسة تحليلية حول المفردات اللغوية في مجال تدريس اللغة الانجليزية في عمان: تحديات من وجهة النظر التعليمية والتعلمية

رحمة بنت ابراهيم المحروقي وفيصل بن سعيد المعمري وكريستوفر دينمان

مستخلص

تعتبر الكتب الدراسية ضرورية ولا غنى عنها في معظم أنظمة التعليم النظامي الموجودة عالميا. ولذا فإن كثيرا من المدرسين يعتقدون بأن أي ضعف في هذه الكتب قد يؤدي إلى التقليل من فرص نجاح وكفاءة عمليتي التعلم والتعليم. وقد استخدمت هذه الدراسة طريقة تحليل المحتوى عن طريق تحديد مواضيع تتعلق بالمفردات المستخدمة في الكتب الدراسية الخاصة باللغة الإنجليزية بسلطنة عمان. وللقيام بالتحليل استخدمت برامج "وورد سميث" و "رينج". وبينت النتائج أن هناك أنماط من عدم التوازن (خلل) في تقديم المفردات المطروحة في الكتب الخاصة بالصفوف المختلفة. كما أظهرت النتائج اختلال طرح وتحميل أنواع هذه الكلمات ومتعلقاتها ومدى كثافتها في الكتب المختلفة. كما أن الكتب تخفق في تغطية الكلمات الموجودة في قائمة الكلمات الأكاديمية والقائمة العامة للخدمات المعتمدتان عالميا حيث تتدنى نسبة التغطية عن ٩٥٪ و ٩٨٪ وهي النسب المقترحة من قبل نيشن (١٩٩٠ و ٢٠٠١) بل في الحقيقة إن كثيرا من المفردات المستخدمة في كثير من النصوص غير موجودة أصلا في قوائم المفردات المعتمدة. تقدم الورقة مقترحات لتحسين تحليل الكتب من أجل تطويرها وتحسين جودتها فيما يتعلق بطرح المفردات في الكتب الدراسية المختلفة الخاصة باللغة الانجليزية.

كلمات مفتاحية: معجم، تحميل المفردات، النهج القائم على تحليل المحتوى، قائمة الكلمات الأكاديمية، القائمة العامة للخدمات.

Introduction

Textbooks are indispensable for teaching and learning in most formal situations. They are a standard element in ELT and classrooms seem incomplete without them (Hutchinson & Torres, 1994). Nevertheless, relationships between textbooks and teachers are riddled with mistrust almost everywhere textbooks are used (Ansary & Babaii, 2002; Elis, 1997; Mukundan, 2003; Sheldon, 1988). By the same token, teachers firmly believe that textbook weakness can cause ineffective learning and teaching and hence produce outcomes different from those anticipated.

In Oman, English is considered an important language and is used across the private and public sectors. For this reason, initiatives taken by the Ministry of Education have sought to implement changes within English language teaching through the introduction of a revised curriculum and new textbooks. However, as with any educational reforms of this type, it is important for teachers and learners to be made explicitly aware of the pedagogical principles associated with the teaching approaches promoted, or else problems associated with a lack of awareness may occur (Al-Jardani, 2012). To the same extent, if textbooks are not thoroughly evaluated, there is a risk that these books might be ignored by instructors and may eventually be completely abandoned, or that a hidden curriculum may arise in which learning and teaching continue as they did before (Kennedy, 1987). Hence, the need for a systematic evaluation of Omani English textbooks arises.

In Oman, the reformed education system is called "Basic Education" and features Cycle 1 (grades 1-4) and Cycle 2 (grades 5-10), followed by "Post-Basic Education" (grades 11 and 12). Throughout the history of state-sponsored English instruction in Oman, the Ministry of Education has employed both commercial and locally produced textbooks and materials for the English curriculum. To date, around five curricula have been designed based on these in-house materials and two series of commercial textbooks (Al-Jardani, 2012). However, so far, very few studies have systematically examined Oman's currently used English textbooks.

The current study sought to address this issue through a computerized analysis of the vocabulary loading of the Omani Basic Education Curriculum's ELT textbooks "English for Me" and "Engage with English". The study itself, however, is part of a larger evaluation of these textbooks using two separate checklists. The first checklist focused on the support the textbooks offer for communicative teaching approaches, while the second examined the textbooks' components and use in the classroom as evaluated by teachers and supervisors. The entire evaluation process, moreover, is part of a nation-wide study aiming to investigate the reasons behind Omani school graduates' lack of English proficiency.

Objectives

The objectives of this study involved identifying probable issues around vocabulary teaching in Oman's English language textbooks. A systematic evaluation exercise was carried out on these books to discover to what extent they facilitate or hamper English learning in the classroom. The study also sought to identify features of the vocabulary load of these textbooks as vocabulary loads that increase more quickly than the lexical knowledge of students may result in issues with reading speed and comprehension.

Research questions

In order to meet the above objectives, the following questions were posited:

1. What are the characteristics of vocabulary loading in the Omani grade 1-12 English textbooks?
2. To what extent, if at all, do irregularities exist in the vocabulary loading in these textbooks?
3. What are the lexical properties of the Omani grade 1-12 English textbooks in terms of the GSL and AWL?
4. What are the keywords found in the Omani school English textbooks?

Definition of keywords

- Type vs. token

The number of tokens in a corpus is the total number of words. On the other hand, the number of types

is the number of unique words. Baker, Hardie, and McEnery (2006) state that these can be used to calculate the type/token ratio which has to do with lexical repetition in a text or corpus.

- Density ratio and consistency ratio

Text density is a means of calculating the proportion of lexical words in a text or corpus by expressing it as a percentage (Ure, 1971). It is calculated by dividing the number of lexical or content words (nouns, adjectives, adverbs and main verbs) by the total number of words (or tokens) before multiplying by one hundred (Stubbs, 1996, pp. 71–73). However, it can also be calculated by dividing the number of unique lexical words in the text by the total number of words and then multiplying by 100 (sometimes removing the function words first, sometimes not). The Text Consistency Ratio, often simply called the consistency ratio, measures the consistency of new vocabulary being introduced in the texts (Mukundan, 2007). It also measures how often a new word is introduced after specific running words: that is to say, it refers to new vocabulary introduction intervals. It can be calculated by dividing the number of tokens by the number of types found within the textbooks according to grades.

- English vocabulary – General Service List (GSL)

Michael West published a list of around 2,000 words entitled the General Service List (GSL) in 1953 for ESL teachers and learners. The list was intended to represent the most frequently-occurring words in English based on a written corpus. West claimed that a person who is aware of all the words on the GSL will understand around 80-85 percent of written texts and about 90-95 percent of colloquial speech. Although the GSL consists entirely of headwords, it is assumed that the language user is familiar with all forms of the word. For instance, for the GSL entry "be, assumed knowledge includes the forms of am, is, are, was etc.

- Academic Word List (AWL)

The Academic Word List (AWL) is based on the assumption that learners in English-medium academic environments need to be familiar with around 570 core pieces of vocabulary which account for about 10 percent of the words in any given academic text. For learners to successfully complete an academic course of study, they must necessarily be familiar with these words. The AWL is based on a corpus of written academic English. The AWL can be contrasted with technical vocabulary, or those words that are closely related to a specific field or area of study (Richards, Platt, Platt, & Candlin, 1992).

- Keywords

One issue associated with keyword analysis is that a word may appear to be key due to its high level of repetition in a small number of texts in a corpus. To help determine if a word is key and representative of the corpus as a whole, a dispersion plot across the corpus could be examined or a list of words that are central to multiple texts in the corpus could be calculated, therefore allowing for the potentially skewing nature of disproportionate representation.

- Review of the literature

Frequently used English words, academic words, technical, and low-frequency words are four kinds of vocabulary that have been identified thus far (Nation, 2001). High-frequency words can be defined as those words that are covered across subjects and can be encountered in different kinds of texts with a significant level of coverage. As outlined above, West's 1953 word list is often offered as perhaps the best example of a high-frequency word list (Coxhead, 2000; Nation, 2001), and it still continues to provide a significant level of coverage of running words in academic texts. In addition to high frequency word lists, academic vocabulary should also be focused on during core classes and learner self-study. As Coxhead (2000) states, the AWL consists of words that cover 28 subject areas across a variety of fields, including commerce, law, science and arts.

Academic vocabulary is sometimes referred to as sub-technical vocabulary as it includes formal rather than technical words (Nation, 2001). Cowan (1974, p. 391) defined sub-technical words as "context independent words which occur with high frequency

across disciplines". Trimble (1985, p.130) adds that this term "refers to those words that have one or more 'general' English meaning and which in technical contexts take on extended meanings (technical or specialised in some fashion)". A number of authors claim that sub-technical and/or academic vocabulary is often a source of difficulty for students (Cowan, 1974; Fox, 1998; Mudraya, 2006; Trimble, 1985; Wang, Liang, & Ge, 2008; Ward, 2009), and, along with technical vocabulary, can become an area of great concern and interest for learners in fields that require a large amount of specific vocabulary.

Technical words form the third type of vocabulary. Technical vocabulary is usually associated with studies in a specific field and/or for a specific purpose. Words from this category are related closely to the subject matter and represent specialised knowledge in a particular field (Chung & Nation, 2003).

The last type - low frequency words - are rarely used words with a wide range but poor coverage in academic texts (Chen & Ge, 2007). Low frequency words have been defined as "words of moderate frequency that did not manage to get into the high-frequency list, proper nouns and words rarely used" (Nation, 2001, p. 11).

Textbooks which are used for language teaching purposes have more recently undergone some analysis of their lexicon. While about 87% of running words in a text often come from the 2,000 high-frequency words, material writers have recently attempted to explore why some textbooks still do not feature a sufficient number of high-frequency words. Another issue with some current textbooks is that these 2,000 high frequency words may not be repeated enough in such a way as to support acquisition. This lack of repetition clashes with a very important principle in language learning which holds that students remember words which they encounter at least seven times over spaced intervals in meaningful contexts (Thornbury, 2002).

To begin reading authentic texts (Nation, 1990; Schmitt, Schmitt, & Clapham, 2001), research indicates that second language learners must know the 3,000 high-frequency words of the language (Nation & Waring, 1997). However, gaining access to 87% of the content of an ordinary text requires ESL learners to know the first 2,000 words (Nation, 1990). For these reasons, the present study attempted to explore the loading and distribution of the 2,000 high-frequency words in ELT textbooks in Oman.

High-frequency words comprise the majority of tokens in any discourse (Schmitt, 2000) as they are the most familiar words. Higher-frequency words are usually short words which learners are likely to encounter more often (Thornbury, 2002). Therefore, it is important that ESL learners learn them for two reasons. First, once high-frequency words are acquired, learners' acquisition of the less-frequently occurring words becomes an easier task through both intentional and incidental learning (Schmitt, 2000). Second, knowledge of the high-frequency words enables learners to more accurately predict the meaning of the less frequently occurring words.

The GSL, which is "a set of 2,000 words selected to be of the greatest general service to learners of English" (Bauman, n.d., para. 1), accounts for about 80%-85% of any given text. Learning these words is an oft-cited goal for second language learners (Nation & Waring, 1997; Schmitt, 2000). This list is comprised of headwords, each of which represents a word family, and, for many years, it has formed the basis for the development of graded readers and other leaning material.

Coxhead (2000, p. 214) asserts that "academic words (e.g. substitute, underlie, establish, inherent) are not highly salient in academic texts, as they are supportive of but not central to the topics of the texts in which they occur". The academic word list (AWL) consists of 570 word families often found across a variety of academic texts. As stated above, the AWL is intended to contain around 10% of the vocabulary encountered in written academic texts.

Academic vocabulary knowledge is an indispensable component of the academic reading abilities that have been directly associated with academic success, economic opportunity, and societal well-being. Gardner and Davies (2013) claim that academic vocabulary plays a pivotal role for native and non-native En-

glish language speakers at all grade levels, including primary, middle-school, secondary and higher education. Having less control over academic vocabulary may be the most influential discriminator in what Gardner and Davies describe as 'gate-keeping' education tests, including TOEFL and IELTS. Furthermore, these authors state that inadequate knowledge of academic vocabulary has been strongly linked with the existence of a 'gap' in academic achievement among certain groups of students.

Previous research

A recent trend in analyzing English textbooks is to use corpus-based techniques in order to explore a variety of linguistic issues. Researchers across a variety of fields now have access to the tremendous processing power of computers to apply a variety of corpus-based analysis techniques to corpora, or samples of authentic written text. There are several studies that have made a corpus-based international comparison of English textbooks. Ban (2002) compared the English textbooks of junior and senior high schools in Korea and Japan and of elementary schools in China, Singapore, the UK, and the US. Among her findings were that:

(a) Chinese textbooks have a higher frequency of 'be' verbs and 'to' clauses than other textbooks

(b) mean word length and the number of words per sentence are similar in Korean and Japanese textbooks at every grade

(c) the vocabulary level measured by the degree of difficulty as calculated by word type and number of words, both of whose calculations exploit the compulsory vocabulary list of Japanese junior high schools, drastically increases between junior and senior high school in Japan.

Ishikawa (2008) conducted a thorough study that contrasted high school first-year textbooks from Korea and Japan and concluded that the Japanese textbooks do not cover as much vocabulary as the Korean textbooks as they have:

(a) much less token-based vocabulary in total

(b) a lower lemma-based coverage of The Freiburg-LOB Corpus of British English (FLOB) and Freiburg-Brown Corpus of American English (Frown) corpora

(c) a lower Guiraud index, which is a measure of topic variety

(d) a lower vocabulary level based on the text coverage by the General Service List and the Academic Word List

(e) a higher frequency of parts of speech representing spoken language

Hsu (2009) adopted a corpus-based approach and analyzed General English (GE) reading textbooks from Taiwan. The results show that the GE textbooks can contribute to learning 49–415 interdisciplinary academic words. It is estimated that a GE textbook is capable of supplying learners with around 162–2001 new word families. It may be useful in preparing learners for an intermediate general English proficiency test by covering 24.55% to 65% of the vocabulary involved in the test.

Method

This study adopted a corpus-based approach to analyze Omani school English textbooks used in grades 1 to 12. These are the textbooks prescribed by the Ministry of Education for teaching and learning English throughout Oman. In order to analyze them, the researchers built the corpus in the absence of another that might have been useful.

WordSmith 5.0 (Scott, 2008), which is a software program that examines words and word patterns in texts, was employed for analysis. To use this software, all books must be made into readable digital versions. Pages were scanned using a scanner with an optical character recognition (OCR) function and converted into text file format (.txt) before being analyzed. Two WordSmith tools - WordList and Concord - were mainly used. The study employed West's (1953) GSL as a reference for the 2,000 most frequent words in analyzing the patterns of loading, distribution, and repetition of these words in the Omani textbooks.

Next, a program called RANGE (Heatley, Nation & Coxhead, 2002) was used to perform further analysis. RANGE software can help to determine and discriminate among GSL, AWL words, and technical and low frequency words which are used in books.

The keyword lists used in this study were based on Menon (2009). Menon created 'keyword lists' with the assistance of WordSmith. The construction of a word list can be achieved through the use of a larger corpus that acts as a point of reference (Nation, 2001). For this reason, the pedagogical corpora here were compared to the British National Corpus (BNC). The BNC is the most suitable corpus for use as a reference since it consists of 100 million tokens and thus is at least five times greater than the target corpus (Berber-Sardinha, 2002; McEnery, Xiao, & Tono, 2006).

Results & discussion

The study examines the properties of textbooks developed for Omani primary and secondary schools in terms of the following features: the number of running words (tokens), word variety (types), density ratio, and consistency ratio across textbooks for all 12 grades. The statistical data will help in deciding whether a book is suitable for its level in terms of vocabulary loading and distribution, and the level of difficulty for Omani students studying English as a foreign language. The 'cramp-ness' of a text and how frequently an item is introduced after every running word can be measured through the density and consistency ratio respectively.

The next part discusses the appropriateness of the vocabulary introduction and its characteristics, based on a comparison with similar words found in the GSL and AWL, both of which are believed to provide coverage of up to 87% for high frequency words and up to 8% for the academic vocabulary in any teaching text (Nation, 1990, 2001). However, Coxhead (2000) claims that her AWL can provide up to 10% of word coverage for any academic materials. Thus, having words from both lists is essential for learners in Oman to enhance their vocabulary knowledge and to meet their academic needs from grade 1 to grade 12. The keyword list in this study was generated by WordSmith 5.0 software which can show the 'representation-ness' of words found in the entire series of textbooks. All the findings will further be elaborated on and discussed in terms of the research questions outlined above.

What are the characteristics of vocabulary loading in the Omani grade 1 to grade 12 school English textbooks?

"Basic Education" is Oman's dominant education system. As stated above, it includes two main cycles called Cycle 1 (grades 1-4) and Cycle 2 (grades 5-10), in addition to grades 11 and 12 which are termed "Post-Basic Education". A summary of statistics for the series of books is shown in Table 1. Using concordance software, WordSmith 5.0 showed that the total number of running tokens found in the corpus was 651,995.

Table 1: Summary of the Statistics of Omani Grades 1-12 English Textbooks

Grade	Tokens	Types	DensityRatio (Types/Tokens * 100)	Consistency Ratio (Tokens/ Types)
1	5344	673	12.59	7.94
2	11442	1319	11.53	8.67
3	15794	1823	11.54	8.66
4	24859	2864	11.52	8.68
5	20896	1314	6.29	15.90
6	41152	2637	6.41	15.61
7	62159	4569	7.35	13.60
8	80546	6439	7.99	12.51
9	100955	7332	7.26	13.77
10	85514	7014	8.20	12.19
11	81920	6576	8.03	12.46
12	121414	8738	7.20	13.89

From the table it can be seen that, overall, as students move to higher grades, the more tokens and types can be found in their textbooks. Tokens refer to the number of running words in the texts and they do not exclude repeated words. Types, on the other hand, refer to the number of distinct words found

in a text. This is common because, as learners grow and learn at higher levels, they should be exposed to more vocabulary for their academic needs. The trend initially appears in the table as an uptrend, except that, starting at grade 5, it becomes irregular. The following section deals with these irregularities.

To what extent, if at all, do irregularities exist in the vocabulary loading of these textbooks?

Data interpretation will be presented in the form of graphs so that the trend can be easily distinguished and discussed. There are four graphs which are the tokens, types, density ratio and consistency ratio.

The pattern of the token representation is linear and trends upward starting from Grade 1 (see Figure 1). However, it dips in grade 5 before it increases again. In the grade 10 and 11 textbooks, the number of tokens drops to grade 8 level before reaching the 12,000 tokens mark for grade 12 textbooks. This trend should not occur if the textbooks are controlled and analyzed with concordance software. Moreover, it clearly shows inconsistency in the distribution of words, which is a common problem with locally written or government-commissioned textbooks, as they tend to be developed in an ad hoc manner (Mukundan, 2007, 2009). Exposure to a given number of words is essential and should be controlled by material developers so that students will not become overwhelmed by the sheer number of words (Tomlinson, 1998).

Figure 2 shows the distribution of types across the grades in the textbooks and reveals a trend similar to that of the pattern found in the tokens graph. The introduction of types in any academic textbook must be done systematically so as not to overburden students and affect their language acquisition negatively. It is evident again that inconsistency is found in the types distributions across the grades.

Another focus of this study is the density of the texts (see Figure 3). Density refers to the 'compactness' of a page, especially in terms of number of words it contains. This can be calculated by dividing the number of types by the number of tokens found in the textbooks or by referring to the Omani English Language Textbooks Corpus (OELTC). For the purpose of this study, the density ratio was calculated using the same formula except that percentages were used instead of decimals - i.e. the value of the density ratio was multiplied by 100. Analyzing data from the density percentage gives researchers a better estimation or readability value. For instance, if the density percentage ratio is 50%, it means that 50% of the texts are crammed with words. The density ratio measures text intensity. The higher the ratio or percentage, the more compact the page or textbook is.

Textbooks ought to be provided with more graphical presentation so that the pages look less threatening in terms of "white space". According to material development principles, the right amount of "white space" on a page is essential to facilitate better learning (Ng, 2011; Tomlinson, 1998). For the purpose of this study, the density ratio is assumed to provide the right amount of "white space".

Figure 3 shows that the density ratio decreases instead of increasing across the grades and it fluctuates with, for instance, a sudden drop from grades 4 to 5 before increasing again towards grade 8 and then moving in an up-down-up pattern onwards to grade 12. The density ratio for each grade should remain consistent or increase steadily to indicate that the words are distributed throughout the textbooks equally or introduced systematically. Clearly, this is not the case and the fluctuation once again suggests that the textbooks have been developed in an ad hoc manner. When the two extremes are compared, it is found that grade 1 pages are more text-compact than grade 12 pages. The results, based on expected density ratio comparison, should have been completely the opposite. This suggests that grade 1 learners have been overly exposed to words, which in turn may have had a negative impact on their learning of English even in this early stage of education. The results, therefore, show an unacceptable ratio of density to grades throughout the textbooks produced for Omani primary and secondary schools.

The Text Consistency Ratio measures the consistency of new vocabulary being introduced in the textbooks (Mukundan, 2007). It also measures how often a new

vocabulary item is introduced after specific running words, or new vocabulary introduction intervals. This can be calculated by dividing the number of tokens by the number of types found in the textbooks. For instance, if a textbook's consistency ratio is recorded as 10.11, the introduction of a new vocabulary item will be implemented after approximately 10.11 words, or at about a 10-word interval. If this is the case, students can read comfortably the first 10 words before encountering an unknown eleventh word. This measure is a good indicator of the consistency of vocabulary loading, especially when dealing with the types of vocabulary. Figure 4 reveals that the trend found in the analysis is unfavourable to lower-level students.

The values of the consistency ratio across the textbooks should be inversely proportional, meaning that, as the textbook levels increase, the consistency ratio value should decrease to indicate that the higher-level textbooks are more difficult; or, in other words, that more new vocabulary is introduced after fewer running words. From the results obtained, it is evident that the consistency ratio of the textbooks studied fluctuates, showing that the vocabulary loading of new words in the textbooks is not controlled well. Learners in grades 1-4 have been exposed to too many new words after an approximately 8-word interval, which may simply overwhelm and discourage them. Figure 4 reveals that, from grade 5 onwards, the data reading is justified because learners are exposed to new words more regularly as they progress to higher grades. However, from grades 8-12, the range of the consistency ratio is from 12.5 to 13.9 with a mean ratio of approximately 13, which shows that the beginning grade 8 students are consistently introduced to new vocabulary item after every 13 words. Stressing students with excessive amounts of new vocabulary in the early stages of their language study may simply prevent learning from taking place (Tomlinson, 1998).

What are the lexical characteristics of the Omani grades 1-12 school English textbooks in terms of the GSL and AWL?

Although the Omani Ministry of Education provides a comprehensive syllabus meant as a guide for all its educators, its vocabulary section lacks any mention of specifically-recommended vocabulary. Without a syllabus word list, teachers might not be fully aware of the types of vocabulary to be introduced. Despite valuable advice concerning the scope and sequence of instructional components in the syllabus, the absence of a word list is a key shortcoming of the textbooks. Hence, teachers will have different ideas about the key vocabulary to be taught and must rely solely on vague recommendations. Learners will be misdirected and confused and, worse, perhaps deprived of the functional or essential English vocabulary needed for academic purposes.

The GSL is regarded as the essential list of commonly-used English words which can provide reading coverage of up to 87% of an academic text (Nation 1990, 2001). According to Nation (2001) and Coxhead (2000), the list is essential for facilitating learning at the lower levels. Another word list, essential for academic purposes, is the AWL (Coxhead, 2000), which can provide coverage of up to 10% of an academic text. The coverage percentage claimed for these two lists is confirmed even in TOEFL materials (Ng et al., 2012) and knowing them is essential for learners before moving to technical language in their tertiary education (Ng et al., 2013).

In this study, the researchers can only *assume* that the GSL and AWL are the lists implicitly suggested in the Omani ELT syllabus based on its intended scope. Thus, the analysis needed to answer the research question here must constantly take into account the percentage of coverage for words found in the GSL and AWL. According to Nation (2001), in order to achieve the ideal coverage to read and listen at optimum levels, learners need to achieve 98% coverage of the target texts. He suggests 8,000-9,000 word families for effortless reading of a passage, while, for listening, 6,000-7,000 word families will suffice. However, achieving the recommended levels is an uphill task for EFL learners. Hence, achieving the 2,570 word families mentioned in the GSL and AWL seems a feasible task as this figure can provide up to 95% of text coverage for comprehension purposes, though

percentages will vary according to the nature or discipline of the textbooks. Learning only 2,570 word families instead of around 6,000 seems like reasonable target for educators and learners alike.

Figure 5 illustrates the coverage percentage of words in the GSL and AWL found across the grades 1-12 textbooks. From the graph, it appears that the GSL and AWL coverage provides a decent overall percentage, with the coverage of words in grade 6 reaching 91.88% and the lowest coverage reaching 84.35%. The results suggest that if Omani learners acquire the 2,000 most commonly-used English headwords in the GSL and the 570 headwords in the AWL, they will be guaranteed at least an 84% coverage of the English textbooks throughout grades 1-12. This seems both worthwhile and achievable.

Although the textbooks show a decent coverage of words in the GSL and AWL, it is still evident that the vocabulary loading in terms of types has not been effectively planned and introduced. The data still show inconsistency because the GSL should have been widely introduced at the lower levels – especially primary school – with a gradual decrease before the introduction of the AWL and other technical or specialised vocabulary in later stages. The percentage of words not in the lists should be reduced at the lower levels to facilitate better learning and more effective teaching.

What are the keywords found in Omani school English textbooks?

Identifying the keywords in textbooks can be done in various ways. For a limited corpus like the one studied, it can be done by comparisons with a larger corpus, often referred to as a reference corpus (Menon & Mukundan, 2010; Nation, 2001; Ng et al., 2012, 2013). The BNC is used as a particularly suitable model, because, as mentioned earlier, it is at least five times greater than the target corpus (100 million tokens compared to 651,995). The necessity of this ratio is stressed by Berber-Sardinha (2002) and McEnery, Xiao, and Tono (2006). The keywords found in this study are in their lexico-grammatical forms and learners are expected to find out more about them from a word list. It is effective to acquire vocabulary from lists, especially when learners translate and learn the words' context in their first language (Nation, 2001; Thornbury, 2004).

Nakata (2008) states that list learning can be likened to a "rehearsal" where new and repeated information can reach the long-term memory. Thus, the keywords identified in the current research can act as a guide for teachers and learners to supplement essential vocabulary that should have been taught or learned. The words are not categorised into families in case more in-depth scrutiny is needed as to collocation, structure, multi-word pairing and combination. The keywords list provided in Appendix A contains the words with significant frequency after 'filtering' and comparison with the reference corpus. It means that for a word to emerge as a keyword from this study, it must have positive 'keyness' characteristics.

This is determined by using the 'keyword' function in WordSmith 5.0 with the BNC as a reference corpus in order to determine which word in the target corpus occurs significantly more frequently than in the BNC. The output of the keywords identified serve as a supplementary list to the other two more important lists - i.e. the GSL and AWL. When the list is created after the filtering process, similar words found in both the GSL and AWL and among the keywords were removed. Abbreviations and proper nouns identified in the keywords list were also removed. The final keywords list is offered solely as a supplementary guide so that learners and educators may benefit from the analysis results.

Conclusion

The main objective of the study was to discover whether the prescribed ELT textbooks used in Oman's Basic Curriculum were developed systematically to ensure effective pedagogy among learners in terms of vocabulary. Based on the findings, it can be concluded that there are anomalies in vocabulary loading in that words are not introduced at appropriate levels. The types, tokens, density and consistency ratios show obvious inconsistency as they reveal ex-

treme values. It is suggested that the materials were not developed in a planned manner with comprehensive analysis and arrangement in terms of vocabulary. Furthermore, in terms of GSL and AWL word coverage, the entire series falls short of the ideal range of 95%-98% suggested by Nation (1990, 2001) as the materials are supposedly purely academic in nature. Acceptable percentages (the range was only from 8.11% to 15.65%) were not found in any of the recommended essential word lists.

The findings reveal the importance of creating awareness among material developers and publishers that anomalies will certainly exist if textbooks are written in an ad hoc manner. Concordance software can help writers and publishers to discover anomalies so that vocabulary loading can be effectively distributed and planned in terms of the types to be introduced. Also, from a concordance analysis keyword lists can be created to act as a supplement to facilitate better vocabulary teaching and learning, making it easier for teachers to decide which important words ought to be taught. With consistent monitoring and analysis of textbooks, their quality and effectiveness can undoubtedly be elevated in terms of vocabulary introduction and loading. Well-guided writers will produce well-guided learners and thus ensure an improved acquisition of the target language.

A major finding from this study is that a reference word list was most likely not used when the textbooks evaluated here were being developed. This absence can, as discussed above, seriously affect the quality of learning and teaching that takes place in Omani Basic Education English classrooms, while also contributing to learner anxiety and demotivation. If the syllabus examined incorporated the relevant contents of the GSL and AWL, not only would such negative outcomes be less likely to occur, but educators in Omani public schools would also be more aware of the vocabulary that their students would be exposed to and that they need to acquire to improve their comprehension and reading fluency. Fluency and specialised vocabulary acquisition are the next best step for learners only when the GSL and AWL words have been made known to them. For these reasons, the current study adds support for the need of an extensive and systematic evaluation of the English textbooks currently employed in Omani Basic Education schools.

References

Al-Jardani, K. S. (2012). English language curriculum evaluation in Oman. *International Journal of English Linguistics, 25*, 40.

Ansari, H. and Babaii, E. (2002). Universal characteristics of EFL/ESL textbooks: A step towards systematic textbook evaluation. *The Internet TESL Journal, 82*, 1-9.

Baker, P., Hardie, A. and McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.

Ban, H. (2002). Higashi ajia to beiei no eigo kyokasho no keiryo teki kaiseki hikaku [Metrical comparison of English textbooks in East Asia, the USA, and the UK]. *Journal of the Faculty of Humanities and Social Science, 2*, 75-82.

Bauman, J. (n.d.). The General Services List. Enterprise Training Group. Retrieved from http://jbauman.com

Berber-Sardinha, T. (2002). Comparing corpora with WordSmith Tools: How large must the reference corpus be? Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.13.7923

Chen, Q. and Ge, G. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles. *English for Specific Purposes, 26*, 502-514.

Chung, T. M. and Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language, 152*, 103-116. Retrieved from http://nflrc.hawaii.edu/rfl/October2003/chung/chung.pdf

Cowan, J. R. (1974). Lexical and syntactic research for the design of EFL reading materials. *TESOL Quarterly, 84*, 389-399.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 342*, 213-238.

Ellis, R. (1997). *SLA research and language teaching*.

Oxford: Oxford University Press.

Fox, G. (1998). Using corpora data in the classroom. In Tomlinson B. (Ed.), *Materials development in language teaching* (pp. 25-43). Cambridge: Cambridge University Press.

Gardner, D. and Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics, 35*, 1-24. doi:10.1093/applin/amt015

Heatley, A., Nation, I. S. P. and Coxhead, A. (2002). RANGE and FREQUENCY programs. Available online: http://www.victoria.ac.nz/lals/about/staff/paul-nation

Hsu, W. (2009). College English textbooks for general purposes: A corpus-based analysis of lexical coverage. *Electronic Journal of Foreign Language Teaching, 61*, 42-62.

Hutchinson, T. and Torres, E. (1994). The textbook as agent of change. *ELT Journal, 484*, 315-328.

Ishikawa, S. (2008). *Eigo corpus to gengo kyoiku* [English Corpus and Language Education]. Tokyo: Taishukanshoten.

Kennedy, C. (1987). Innovating for a change: Teacher development and innovation. *ELT Journal, 413*, 163-170.

McEnery, T., Xiao, R. and Tono, Y. (2006). Corpus-based language studies. An advanced resource book. London: Routledge.

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, Malaysia: Universiti Putra Malaysia.

Menon, S. and Mukundan, J. (2010). Analysing collocational patterns of semi-technical words in science textbooks. *Pertanika Journal of Social Sciences and Humanities, 18*, 241-258.

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes, 25*, 235-256. Retrieved from http://www.u.arizona.edu/~karaj/pages/Reviews/Mudraya2006.pdf

Mukundan, J. (2003). State-sponsored textbooks: Are there hidden costs in these "free" books? *The English teacher, 62*, 133-143.

Mukundan, J. (2007). Irregularities in vocabulary load and distribution in same level textbooks written by different writers. *Indonesian JELT, 31*, 99-118.

Mukundan, J. (2009). *ESL textbook evaluation: A composite framework*. Köln, Germany: Lambert Academic Publishing.

Nakata, T. (2008). English vocabulary learning with word lists, word cards, and computers: Implications from cognitive psychology research for optimal spaced learning. *ReCALL Journal, 20*, 3-20.

Nation, I.S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House Publishers.

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

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

Ng, Y. J. (2011). Evaluating white space of a Malaysian secondary ELT textbook. In Mukundan, J. & Nimehchisalem, V. (Eds.), *Reading on ELT materials V* (pp. 220-232). Serdang: Faculty of Educational Studies,Universiti Putra Malaysia.

Ng, Y. J., Chong, S. T., Mohamed Nor, M., Tarmizi, M.A. A. & Nor Mohamad, A. F. (2012). Corpus based analysis of the TOEFL course books: What are the words we should teach our students? *International Review of Social Sciences and Humanities IRSSH, 32*, 152-160.

Ng, Y. J., Lee, Y. L., Chong, S. T., Sahiddan, N., Philip, A., Azmi, N. H. N., & Tarmizi, M. A. A. (2013). Development of the Engineering Technology Word List for Vo-

cational Schools in Malaysia. *International Education Research*, *1*(1), 43-59. Retrieved from http://dx.doi.org/10.12735/ier.v1i1p43

Richards, J. C., Platt, J., Platt, H. and Candlin, C. N. (1992). *Longman dictionary of language teaching and applied linguistics.* London: Longman.

Schmitt, N. (2000). *Vocabulary in language teaching*. Stuttgart: Ernst Klett Sprachen.

Schmitt, N., Schmitt, D. & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels test. *Language Testing, 181*, 55-88.

Scott, M. (2008). *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software.

Sheldon, L. E. (1988). Evaluating ELT textbooks and materials. *ELT Journal, 424*, 237-246.

Stubbs, M. (1996). *Text and corpus analysis*. London: Blackwell.

Thornbury, S. (2002). *How to teach vocabulary*. Harlow: Longman.

Thornbury, S. (2004). *Natural grammar: The keywords of English and how they work*. Oxford: Oxford University Press.

Tomlinson, B. (1998). *Materials development in language teaching*. Cambridge: Cambridge University Press.

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

Ure, J. (1971). Lexical density and register differentiation. In Perren G. & Trim, J. L. M. (Eds.), *Applications of linguistics* (pp. 443-452). London: 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.

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes, 28*, 170-182.

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

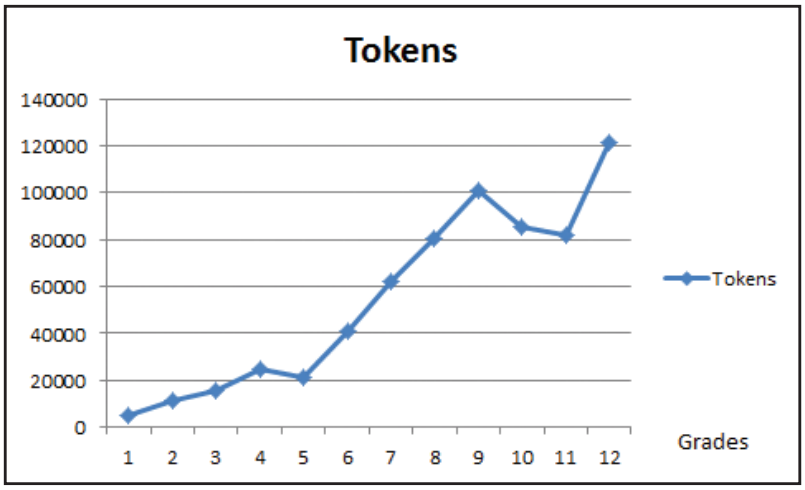


Figure 1:

Total Number of Tokens across Grades 1-12 Omani English Language Textbooks

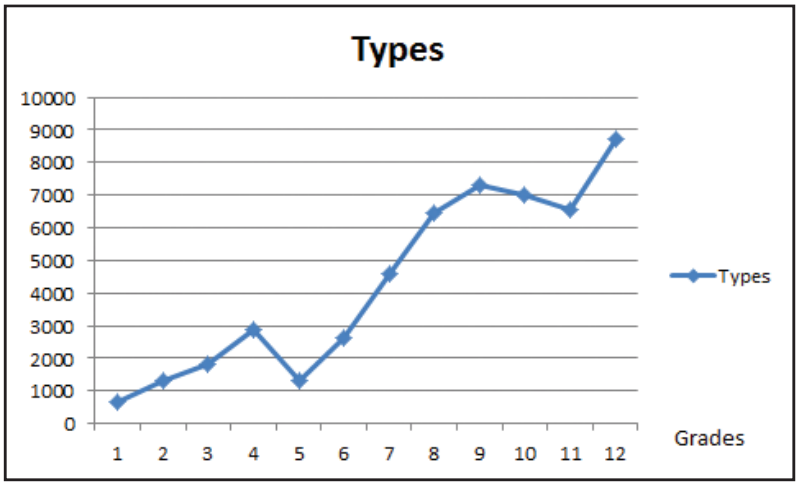


Figure 2:

Total Number of Types across Grades 1-12 of the Omani English Language Text Density Ratio

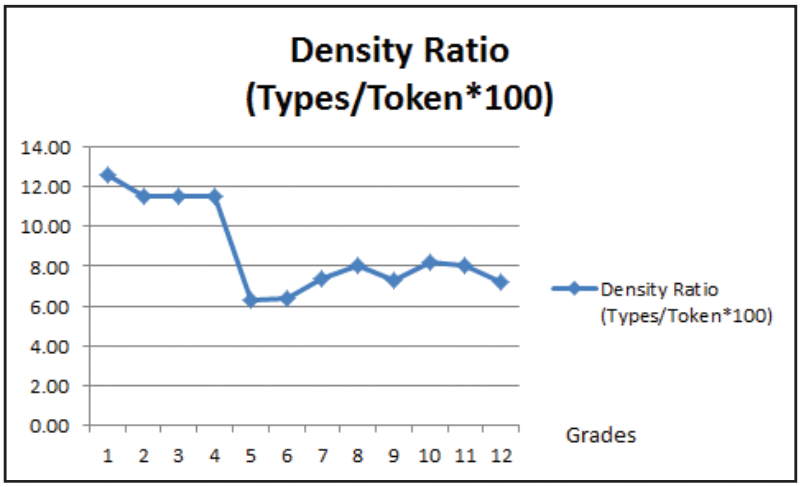


Figure 3:

Density Ratio (%) of Different Textbooks (From Grades 1-12)

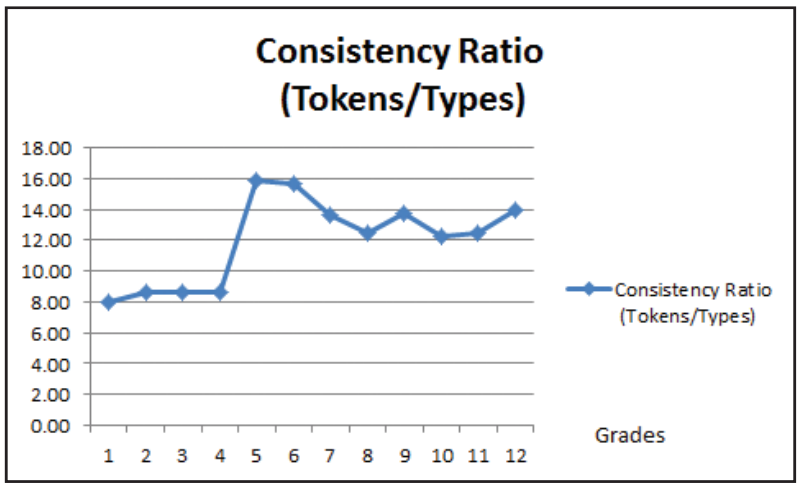
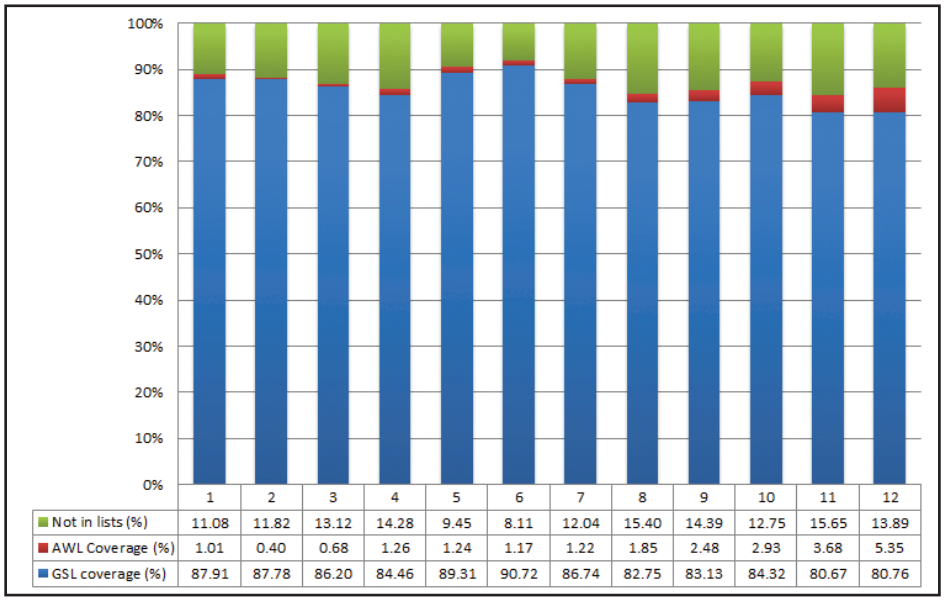


Figure 4:

Consistency Ratio of Different Textbooks (From Grades 1-12)



	1	2	3	4	5	6	7	8	9	10	11	12
Not in lists (%)	11.08	11.82	13.12	14.28	9.45	8.11	12.04	15.40	14.39	12.75	15.65	13.89
AWL Coverage (%)	1.01	0.40	0.68	1.26	1.24	1.17	1.22	1.85	2.48	2.93	3.68	5.35
GSL coverage (%)	87.91	87.78	86.20	84.46	89.31	90.72	86.74	82.75	83.13	84.32	80.67	80.76

Figure 5:

Coverage of the GSL and AWL (in terms of percentages) Found in Grades 1-12 of the Omani English Textbooks

Appendix A: Keywords Identified from the Analysis

ABBREVIATION
ABORIGINAL
ABORIGINE
ABORIGINES
ACKNOWLEDGMENTS
ACNE
ACTIVATE
ADJECTIVE
ADOLESCENCE
ADVERB
ADVERBIAL
ADVERBS
AFFIXES
AIRPORT
ALLITERATION
ALPHABET
ALPHABETICAL
AMAZING
APOSTROPHE
APOSTROPHES
ARTEFACTS
ASIA
ASMA
ASTEROID
ASTRONAUT
ASTRONAUTS
ATTRIBUTIVE
AVIARY
BABOON
BADGE
BADGES
BALLOON
BAMBOO
BANANA
BANANAS
BARBECUE
BARCODES
BASKETBALL
BEACH
BEACHES
BEE
BEES
BIKE
BIL
BILLBOARDS
BINSERT
BIO
BIOGRAPHIES
BIOGRAPHY
BIOMASS
BLANKS
BLOG
BLOGS
BOOMERANG
BORED
BORING
BRACKETS
BRAINSTORM
BUBBLES
BUDOOR
BUSTER
BUSTERS
BUTTERFLY
CAMEL
CAMELS
CARNIVAL
CARROT
CARROTS
CARTON
CARTOON
CATERPILLAR
CELEBRATE
CELEBRATIONS
CELEBRITIES
COCKY
COCONUT
COPYRIGHT
COULD

CHANT
CHARITIES
CHARITY
CHEETAH
CHEF
CHESS
CHIPS
CHOCOLATE
CHOPSTICKS
CINEMA
CITIZENSHIP
CLASSBOOK
CLASSIFIEDS
CLASSMATE
CLASSMATES
CLASSROOM
CLOAK
CLOAKS
CLONING
CLOWN
CLUELESS
CLUES
COCKROACHES
DISASTERS
DOWNLOADED
DRAGON
DUBAI
EARTH
COLLOCATE
COLLOCATIONS
COLUMN
COLUMNS
COMETS
COMIC
COMICS
COMMA
COMMAS
COMPANYING
COMPARATIVES
COMPREHENSION
CONDITIONER
CONGESTION
CONJUNCTIONS
CONS
CONSERVE
CONSONANT
CONSTELLATION
CONSTELLATIONS
CONTRACTIONS
COOKIE
COOKIES
FOOTPRINTS
FORT
FOSSIL
FOX
FOXY
COUNTERS
COURSEBOOK
COWPOX
CRAFTS
CREW
CROCODILE
CROCODILES
CROSSWORD
CRUISE
DELICIOUS
DENTIST
DEPART
DESERTIFICATION
DESTINATION
DIAGRAM
DIAGRAMS
DIALOGUE
DIALOGUES
DIARY
DINOSAUR
DINOSAURS
DIRECTORATE
DISASTER
GRASSLANDS
GREENHOUSE
GROOM
GROUNDWATER
GRUFF

ECOCITIES
EMAIL
EMBEDDED
ENDANGERED
ENDINGS
EQUATOR
ESSAY
EXAM
EXAMS
EXTINCT
FANTASTIC
FESTIVAL
FESTIVALS
FICTION
FICTIONAL
FIREFIGHTER
FIREWORKS
FLOWERY
FOLKTALES
FOOTPRINT
INFO
INSOLE
INTERNET
INTERPERSONAL
INTERRUPTIONS
INTERVIEW
INTERVIEWER
INTERVIEWS
FRIDGE
FROG
FUELS
FUNDRAISING
FUNFAIR
GENETIC
GENRE
GENRES
GEOTHERMAL
GERMS
GERUND
GERUNDS
GIRAFFE
GIST
GLACIERS
GLOSSARY
GLOVES
GOODWILL
GRANDDAD
GRAPH
LEATHERBACK
LITTER
LOWERS
MADAM
MAGAZINE
MAGAZINES
MALLS
MASKS
GUESTHOUSE
GUIDEWAY
GYM
HALLOWEEN
HEADINGS
HELICOPTER
HEN
HERITAGE
HERMIT
HOBBIES
HOMEWORK
HONEYCOMB
HOOD
HOSPITALITY
HURRICANE
HYDROPOWER
IDIOMS
IMPOLITE
INFINITIVE
INFINITIVES
ONLINE
OPERA
ORIGAMI
OUCH
PALM
PANDA
PANDAS
PAPARAZZI

INTRAPERSONAL
ITINERARY
JEANS
JELLY
JELLYFISH
JOURNALISM
JUMBLED
JUNGLE
JUNK
KANGAROO
KANGAROOS
LANDSLIDE
LANDSLIDES
LAPTOP
LARVA
LAYOUT
LEAFLET
PIZZA
PLANETS
PLASTIC
PLOT
POLLUTE
POLLUTED
POLLUTION
PORRIDGE
PORTFOLIO
POSSESSIVE
POSTC
MATHS
MEDALS
MIRAGE
MOBILE
MODAL
MODALS
MONSTER
MOSQUE
MOSQUITOES
MOVIES
MULTIWORD
MUSEUM
MUSEUMS
MUSLIMS
NATIONALITIES
NATIONALITY
NOMINEE
PRONOUNS
PROOFREAD
PROPHET
PROS
PULSE
PUMPKIN
PUNCTUATION
PUPPET
PUPPETS
PYRAMIDS
QUANTIFIERS
PARTICIPLE
PARTICIPLES
PARTNER
PASSPORT
PENDRAGON
PENGUIN
PENGUINS
PERSONALISE
PESTICIDE
PHONEMIC
PHONETIC
PHOTO
PHOTOS
PHRASE
PHRASES
PICNIC
PITCHER
REUSABLE
REWRITE
RIDDLE
RIDDLES
ROBOT
ROBOTS
ROCKET
ROOSTER
SANDALS
SATELLITES
SAUCEPAN

POSTCARD
POSTCARDS
POTATO
POLAR
POLITELY
PREDICATIVE
PREFIX
PREFIXES
PREPOSITION
PREPOSITIONS
PRESENTER
PRESENTERS
PREVIEW
PRONOUN
SHOWTIME
SHRIKES
SHUTTLE
SIGNPOSTS
SIHAM
SIMILE
SIMILES
SINAN
SINBAD
SINCETO
SINGULAR
SKATING
SKIING
SKIM
QUESTIONNAIRE
QUIZ
QUIZZES
RAINFOREST
RAINFORESTS
RECTANGLE
RECTANGLES
RECYCLE
RECYCLED
REEFS
REFERENCE
REFLEXIVE
REGGIE
RENEWABLE
SOUNDBITES
SPACECRAFT
SPACEPORT
SPACESHIP
SPACESUIT
SPELLINGS
SPICY
SPIDER
SPIDERGRAM
SPIDERS
SQU
SQUEERS
STAGECRAFT
STATIVE
SCARED
SCARF
SCARY
SCORPION
SCRAPBOOK
SCRIPT
SEAWATER
SEMESTER
SEQUENCERS
SHARK
SHARKS
SHOPAHOLIC
SHOPKEEPER
SHORTS
SWEATER
TASTEBUDS
SYLLABLE
SYLLABLES
SYNONYMS
TAPESCRIPT
TASTY
TEEN
TEENAGE
TEENAGER
TEENAGERS
TEENS
TELESCOPE
TELEVISION

SKIMMING
SKIP
SMALLPOX
SMALLSVILLE
SMART
SMOGTOWN
SNOWBALLS
SNOWMAN
SOFA
SOLAR
SOOT
TSUNAMI
TURBINES
TURNIP
TURTLE
TURTLES
TWISTER
TWISTERS
UNDERLINE
UNDERWATER
UNESCO
UNHEALTHY
UNREAL
UNSAFE
UNSCRAMBLE
URANUS
SUBORBITAL
SUBWAY
SUFFIX
SUFFIXES
SULTAN
SULTANATE
SUPER
SUPERLATIVE
SUPERLATIVES
SUPERMARKET
SURNAME
VEGETABLES
VOCABULARY
VOLCANO
VOLLEYBALL
WATERMELON
WHALE
WHALES
WEB
WEBSITE
WEBSITES
WHITEBOARD
WHITENING
WIKI
WIKIMEDIA
WIKIPEDIA
TENNIS
TENSES
TIMETABLE
TORNADO
TORNADOES
TORTOISE
TOURISM
TRAFFIC
TRANSRAPID
TRIANGLE
TROUSERS
WILDLIFE
WINDMILLS
WINKLE
WOLF
WORDLIST
WORDPRESS
WORKBOOK
WORLDNET
ZEBRA
ZEBRAS
ZOO
ZOOKEEPER
ZOOS