



**INSTITUTIONEN FÖR PEDAGOGIK
OCH SPECIALPEDAGOGIK**

A CORPUS-VOCABULARY ANALYSIS OF GRADE 9 ENGLISH TEXTBOOKS IN SWEDEN

Dianne Valize Garcia

Uppsats/Examensarbete:	15 hp
Program och/eller kurs:	LVXA1G
Nivå:	Grundnivå/Avancerad nivå
Termin/år:	Vt/2021
Handledare:	Vi Thanh Son
Examinator:	Giulia Messina Dahlberg

Abstract

Title: *A Corpus-Vocabulary Analysis of Grade 9 English Textbooks in Sweden*

Author: Dianne Valize Garcia

Supervisor: Vi Thanh Son

Abstract: The importance of vocabulary in language acquisition is undeniable. Since textbooks play a vital role in learning a language as a subject in school, this paper investigates the vocabulary content of grade nine English textbooks in Sweden. This corpus examination is concerned with what word types are found from the samples based on Nation's (2013) theory, the academic words by Coxhead (2000) and the pedagogical reflections of the vocabulary to school syllabus and learning goals. The results show that the books contain around 40,000 – 44,000 tokens and 2,000 - 2,700 word families which are mainly the high-frequency words. Concerning the academic words, less than 3% of words in the book belong to the Academic World List. With these results, the books cover the most common words in English or the high-frequency words, which are very helpful in communication and which equip the learner to be a B1 or B2 user based on CEFR level; but the low input of academic vocabulary implies insufficient preparation for reading academic texts.

Table of Contents

1. Introduction	5
2. Background	6
2.1 The myth of native speaker	6
2.2 English in Swedish schools.....	7
2.3 Textbooks in Swedish schools	8
2.4 The acquisition of vocabulary.....	9
2.4 Definition of terms	11
3. Theoretical framework	12
3.1 Types of words	12
3.1.1 British National Corpus base word lists	14
3.2 Academic words	15
3.2.1 Academic base word lists	15
3.3 Comprehension, text coverage, learning goals	16
3.3.1 The syllabus for English	18
4. Material and method	17
4.1 Data collection	20
4.1.1 Range software	21
4.2 Data analysis	21
4.3 Limitations	22
4.4 Ethical considerations	22
5. Results and discussions	23
5.1 Word types on British National Corpus.....	23
5.2 Academic words	28
5.3 Pedagogical reflections: learning goals and school syllabus	30
6. Conclusions	34
References	36
Appendices	38

Appendix A Range40
Appendix B CEFR Descriptions41
Appendix C Gleerups 9 headwords base word list 1.....42
Appendix D Good Stuff Gold D headwords base word list 145
Appendix E Academic Word list by Coxhead.....48

1. Introduction

In second language acquisition, there seems to be more focus on how students learn but less on what kind of textbooks students' read. Based on the author's personal experience, when students in the classroom do not reach a passing grade, the teacher's teaching skill is often criticised and held accountable often by parents. Although choosing appropriate materials such as textbooks is part of the teacher's skill, if the material available in the market is deficient or inadequate in content, teachers are left with minimal choice. Teachers solely rely on what the publishers offer, and the content of the textbooks is beyond teachers' jurisdiction. This being said, vocabulary content in textbooks has escaped many syllabus theorists over the last 50 years because attention was given primarily to communicative orientation, but the specifics of vocabulary content is lacking, if not missing (Milton, 2009).

Regarding the acquisition of vocabulary, Sinclair and Renouf (1988) suggest that learners should be taught vocabulary that native speakers frequently use and that not a very large quantity of vocabulary is needed to be taught, especially in initial levels. Milton (2009) also argues that we tend to use particular words in everyday language than others. This eventually leads to the question of how many words a native speaker knows as a point of reference. Goulden et al. (1990) suggest an educated English native speaker knows conservatively below 20,000 word families. Nation (2013) points out that the first 2,000 high-frequency words of English based on a corpus (e.g. British National Corpus) must be the primary goal in learning because these words cover a large part of texts. With this said, the idea of a native speaker must not be taken as a preconceived notion as other strands of research contradicts its existence; and regard it as faulty and mythical (Davies, 2010).

Furthermore, Coxhead (2000) also establishes the Academic Word List to help teachers and language learners to know the most words required to study at tertiary institutions. Because academic words are necessary for understanding texts in various disciplines, learners who know these words would be benefited. Specialised vocabulary as the academic words must be treated and handled like the high-frequency words (Nation, 2013). In other words, these should also be part of the learner's goal. Generally, the more a learner knows a word, the better. With this, researchers in second language acquisition have recognised the immense value of vocabulary in learning a language (Alhuditi, 2017).

From the insights stated, it is undeniable that the importance of textbook content must be considered with the utmost care. Textbook selection by the teacher must be dealt with by strict criteria; one of which is that a good textbook must contain a broad and rich vocabulary input. Although textbooks are not the only source of vocabulary in language learning especially in a foreign language setting, they still are the primary source of the words for many learners (Milton, 2009).

In particular, no study, to the author's knowledge, on vocabulary in textbooks in Sweden uses corpus approach and Nation's (2001, 2013) theory. This paper aims to fill this gap. The critical contribution of this work is to create an understanding of the vocabulary realm in two grade 9 English textbooks used in Swedish schools. More specifically, this study aims to investigate the following questions:

1. What types of vocabulary are presented in Gleerups 9 and Good Stuff Gold D?
2. What are the academic words based on the Academic World List found from the two books?
3. What are some pedagogical reflections of the book's vocabulary content towards learning goals and syllabus objectives?

2. Background

2.1 The myth of native speaker

The native and non-native distinction is not objective because there is nothing "native" about speaking a language which means languages are not biological entities but rather culturally constructed (The Hyperpolyglot Activist Dr. Yebra Lopez, 2021). The categorisation of native and non-native dates back to the 17th and 18th century. Under this period, the national and colonial governmentality starts to formally emerge, at which people with particular cultural and linguistic characteristics emblematic to the national and colonial governance are legitimised while those who do not comply are marginalised (Florens, 2013). In general, the European colonisation around the world and the formation of nation-states are deeply associated with language ideologies; that is, the colonial subjects are imposed with a Eurocentric language (Florens, 2013). One of these languages is English.

In an interview by ENGL Poly U (2013), Davies points out that when people talk about "the" native speaker, what people usually have in mind is the ideal model of learning a

language. Whereas, in reality, native speakers have varied normalcies: some are highly educated while some are not; some people read while some barely open a book; some live in the city while some are from the country. These different normalcies put into the question of the essence of a native speaker, consequently disputing the existence of one. Davies further argues:

Chomsky has always said, really there is no such thing as a native speaker. The only thing you are a native speaker of is your idiolect. So, my English is different from yours...It is even different from my brother's. We have our own idiolect -- everybody has a slightly different. So, who is a native speaker? (ENGL PolyU, 2013).

David Crystal also supports this view. In a more recent interview by Canguro English (2019), Crystal argues that even the RP or the received pronunciation is spoken by only 2% of the population and rarely turns up even in London. With the many emerging varieties of English worldwide, these varieties challenge the concept of what standard English is. As Davies suggests, standard English is only the written aspect of the English language because it is the area where the speakers of the different variations tend to agree upon (ENGL PolyU, 2013).

Therefore, a standard written language is possible, but the notion of a native speaker is a myth.

With the insights presented above, this paper uses the label “native speaker” only for the lack of a term but the author does not adhere to the objective realities of native speaker which could not be verified empirically.

2.2 English in Swedish schools

The English language has been taught as a foreign language in Swedish schools since 1946. In the 1960s, Sweden launched an impressive educational reform creating a bilingual nation where people are expected to learn Swedish and English (Norlund, 2016). In 1994 – 1995, English became a compulsory subject alongside Swedish and Mathematics (Andersson, 2013).

However, there is no fixed year or age when English is taught formally. Typically, it should not be later than year four or age 10 (Norlund, 2016). Thus, teaching English in Swedish elementary schools varies, with some children learning at an earlier age of 7 or in grade 1, while others begin at the age of 9 or in grade 3 (Norlund, 2016). The education curriculum also requires German, Spanish, and French to be introduced as foreign languages, but these languages are not taught until age 12 or grade 6. The early English introduction in Swedish schools reveals the importance of English and its strong position compared with other languages (Lainio, 2000).

Additionally, all students are required to take the English national tests in grades 6 and 9. The aim is to assess students' proficiency in four skills: speaking, listening, reading, and writing. Furthermore, a final passing grade in English when students leave grade 9 or the last school year in elementary school is also compulsory to enter senior high school. These school policies prove that English has a high position in Swedish education, and the demand to be fluent in this language is apparent compared to other foreign languages mentioned earlier.

2.3 Textbooks in Swedish schools

Most teachers use textbooks as the primary teaching material compared with other materials such as recorded conversations in tapes, films, and radios. The survey from Skolverket (2006) shows that most teachers use textbooks in virtuously every lesson than other teaching materials.

One primary reason for the teacher's reliance on textbooks for teaching English is that many teachers, especially those new in the field, perceive textbooks as dependable and trustworthy (Abello-contesse & Lopse-jimez, 2010). Another reason is that from the teacher's perspective, it is convenient. Using textbooks does not require as much planning and preparation as "someone else has already thought" about the lesson and put it into writing (Skolverket, 2006). Some teachers also prefer using textbooks because textbooks contain pre-planned tasks suitable for the students' different abilities. Corresponding workbooks often accompany textbooks for the students to practice, and textbooks come with a teacher's guide for the teachers to guide in lesson execution. According to Skolverket's (2006) interview with English teachers, they argue that using textbooks saves them much time in lesson planning and material preparation. It is a great time-saver for those who are often pressured with time.

Despite the wide usage of textbooks in the classroom, teachers have encountered substantial criticism. The choice of the vocabulary in textbooks is one area critics has prompted. Vocabulary components differ remarkably in various textbooks; writers of textbooks do not appear to regard a set of core vocabulary in creating textbooks (Carter & Mc Carthy, 1988). Aside from this, there is also no standard list of core vocabulary enforcing the writers to use. Rather, vocabulary in the textbooks largely relies on the writer's personal preference.

As early as 1917, Palmer (as cited by Milton, 2009) offered sensible suggestions on what he thought textbooks should contain and how they should be organised. He argues that as society develops and becomes more scientific in language teaching and learning, we must find a logical

sequence for presenting materials of these types. He further implies that school syllabuses and textbooks should resemble and complement each other in their vocabulary choices. The vocabulary choices should be of high importance; otherwise, if the selection of vocabulary is inappropriate or random, these may hinder the student's learning. Thus, Palmer (1917 as cited by Milton, 2009) believes that every word has a proper, correct place where it is introduced to the best advantage and not just anywhere else in a programme.

2.4 The acquisition of vocabulary

One probing question that is often asked is how we know that we know a specific word. This question can be analysed to understand that the knowledge of a word consists of several dimensions and not just one (Takac, 2008). For instance, word knowledge includes the way how words sound and how they are pronounced (phonological dimension) or how they are to be used (syntactic dimension), and/or what they mean (semantic dimension).

A standard convention in distinguishing word knowledge is receptive/passive and productive/active (Milton, 2009). The receptive knowledge of a word associates with recognising words in reading and listening, while productive knowledge of a word associates recognising words in speaking and writing. It is believed that language learners develop their knowledge in vocabulary starting from receptive progressing to productive (Alhuditi, 2017). Since it has been established that the learner's receptive word knowledge is more significant than his/her productive knowledge, academic ministries and educational material designers categorise their wordlist, which learners are expected to know into active or passive. In the Hungarian context, the curriculum developers propose that learners learn around 1,600 words by the 8th grade; 1,200 must be known actively while 400 must be known passively (Milton, 2009).

Another convention proposed by scholars Anderson and Freebody (1981) distinguishes the breadth and depth of knowledge of a word. The number of words a language learner knows is indicated as breadth, and depth refers to what the language learner knows about these words. Word breadth, for instance, involves recognising words in a foreign or second language passively; that is, one may know recognise or remember it when hearing or seeing it but may not necessarily know what it means. On the other hand, word depth involves knowledge of a word and its associates, collocations, and functions.

The binary conventions of receptive and productive breadth and depth may benefit language teachers and material designers, yet the complexity surrounding word knowledge and acquisition remain crucial. A more systematic classification of word knowledge in form, meaning and use; and involving receptive and productive scopes is proposed by Nation (2001, 2013). His proposition is better understood when presented in detail, as Table 1 illustrates.

Table 1. Aspects of knowing a word (Nation, 2013 p. 49).

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express the meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	constraints on use	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often we can use this word?

Note: R means receptive, P means productive

Table 1 presents the different aspects of word knowledge according to Nation (2013). The inclusion of the written and spoken aspects under the form category is apparent in the table. The word part is the aspect of word knowledge which includes recognising that word is made of other parts. For instance, in the word *underemployed*, this word is made up of *under*, *employ* and *-ed*.

The second aspect is form and meaning, which is subdivided into three. They are form and meaning, concepts and referents and associations. The first sub-part primarily encompasses the ability of the learner to connect the word form (spoken and written) into its meaning. The

second sub-part deals with the range of meanings of a single word. The ability to distinguish homonyms such as the word *bark* in a *dog's bark* or *bark of a tree* is a concept and referents. In connection with the preceding sub-part, association concerns the necessity in determining the parts of speech (e.g. noun, verb, adjective, adverbs) in a lexicon and that a word has other terms surrounding it. In the first example, the word *underemployed* is associated with *jobless* and *worklessness*.

The aspect of use is likewise subdivided into three. The first part is the grammatical function which concerns knowledge of parts of speech and the grammar patterns the word applies to. For instance, the word *increase* can be used both as a noun as the *increase of oil prices* and as a verb as *the manager increases the price*. The second part is the recognition of what the word typically occurs with. As for the word *increase* both as noun and verb, it is usual to see collocations *dramatical increase* and *dramatically increase*. Lastly, the knowledge of constraints and use deals with the frequency of occurrence and the context (e.g. cultural) of word usage. *Attain* is a common word found in academic texts, while the adjectives *fatty* and *shrimp* are typical adjectives for nicknames in Thai but may not be so typical in English (Nation, 2013).

Though scholars perceive knowing a word in different conventions, what can be inferred from their insights is that knowing a word has several dimensions (e.g. phonological, syntactic, and semantic). Ideally, a good knowledge of a word includes all these dimensions manifested in the learner's ability to react in the same way as an educated native speaker (Takac, 2008). However, Takac (2008) further suggests that this perfect knowledge of all dimensions is not an all-or-nothing argument; that is, even partial knowledge of a word represents a certain degree of knowing a word. A learner may know the form and meaning of a word but may lack knowledge in understanding the word's constraints. Thus, a learner's word knowledge is a should be perceived as continuum where the receptive and production is situated.

2.5 Definition of terms

This short section defines the key terms adapted in this paper. The definition of these terms is derived from Nation (2001, 2013), who is the main proponent of the theory used in this investigation. The set of terms includes:

1. *Tokens*: a word is counted every time it appears. In this sentence, “The book is covered blue, and the book is old”, there are ten tokens. The word "book" appeared twice, so it is counted as two tokens.
2. *Types*: a word is not recounted when it appears again. For instance, in this sentence, “The book is covered blue, and the book is old”. The word "book" is counted only as one in frequency.
3. *Word families*: this includes a headword, its inflected forms, which are closely derived from it as "selfish", "selfishly", and "unselfishly".
4. *Text coverage*: “refers to the percentage of running words in the text known by the readers” (Nation, 2006, p. 61).

3.Theoretical framework and literature review

This chapter presents the theories used in this study and its related literature. Section 3.1 discusses the type of words and section 3.2 discusses the academic vocabulary, and section 3.3 deals with the pedagogical reflections of the vocabulary towards learning goals and syllabus objectives.

3.1 Type of words

A complex question in learning vocabulary in a second or foreign language is how many words a learner must know to use the language they are learning. Researchers in vocabulary acquisition have created several word lists that allow writers and teachers to see what vocabulary types are the most beneficial for English language learners. In 2001, Nation developed the three types of vocabulary based on their occurrence: high-frequency, mid-frequency, and low-frequency words from the British National Corpus or BNC. Table 2 presents the coverage of these words.

Table 2. Coverage of high, mid, and low-frequency words of British National Corpus (Nation, 2013).

Vocabulary Types	Coverage by %
High-frequency / 2,000 word families, with proper nouns etc.	90%
Mid-frequency / 7,000 word families	9%
Low-frequency/ approximately 50,000 words	1%
Total	100%

Only a small group of words are essential in learning vocabulary, because they cover an enormous proportion of running words in both written and spoken texts, and these words appear

in different kinds of usages in the language (Nation, 2013). This group of words is also known as the high-frequency word and is relatively small, consisting of only 2,000 word families. The words included in this group are generated through a *General Service List* of English words by Michael West, which he created in 1953. In this word list, 165 word families comprise mainly function words such as *a*, *two*, *some*, and *to*; the remaining word families are content words comprising common nouns, verbs, adjectives and adverbs.

The next group of words is called mid-frequency. According to Nation (2013), a mid-frequency group is a large number of practical words that occur often but not often enough to make it to the high-frequency list. Nevertheless, these practical words occur enough to be the next learning goal after the high-frequency words are learned. In a separate study by the same theorist in 2006, he suggested that for an individual to have a 98% coverage of what s/he is reading, s/he must know about 6,000-9000 words families. To simply put it, though mid-frequency words do not occur as often as high-frequency words, it is still instrumental to learn the mid-frequency words because it leads to clear rewards. Schmitt and Schmitt (2014), argues that one essential reward is the ability of the learner to engage with English for authentic purposes such as watching television programs and movies. The primary purpose for watching television and movies usually is pleasure, and if a learner knows fewer mid-frequency words, it affects the ease of viewing and consequently enjoyment (Schmitt & Schmitt, 2014). Mid-frequency words consist of 7, 000 word families; sample words include *boast*, *evolution*, and *entrepreneur*.

The last group of words is low-frequency. This is a large group that appear rarely and cover only a tiny proportion of a text. This type of words is beyond the most frequent 9, 000 words in English (Nation, 2013). Most of these words are proper names; some are technical words that may be frequent in some texts but not in any other texts. Some of these words may also be old-fashioned, formal, vulgar, or belonging to a particular dialect, while others can be words from other languages. Some examples include *gibbous*, *plummet* and *bifurcate*.

Nation (2013) points out that the boundary between these classifications is arbitrary, especially between the high-frequency and mid-frequency words. Some mid-frequency words did not make it to the high-frequency list, and when another corpus is used, it leads to a different ranking of these words located in the boundary. This means that the high, mid, and low-frequency words in Corpus of Contemporary American English (COCA) may slightly differ from

British National Corpus (BNC). But for this study, BNC is the primary corpus that is used in the examination.

3.1.1 British National Corpus base word lists

The British National Corpus or BNC is a collection of 100 million samples gathered from a wide range of sources, designed to represent English's spoken and written language (University of Oxford, 2015). In a corpus as large as this, the base word list is divided by thousand. Each thousand has a corresponding coverage. The table below is presented to provide a clear understanding.

Table 3. British National Corpus coverage by base word list (Nation, 2013)

List	Token Coverage in %	Token Coverage in % with proper nouns, marginal words, and compounds
1 st 1,000	77.96	81.14
2 nd 1,000	8.10	89.24
3 rd 1000	4.36	93.60
4 th 1000	1.77	95.60
5 th 1000	1.04	96.41
6 th 1000	0.67	97.08
7 th 1000	0.45	97.53
8 th 1000	0.33	97.86
9 th 1000	0.22	98.08
10 th 1000	0.28	98.23

The most significant data is perhaps the sudden drop in coverage from the 1st 1000 down. What is clear in attaining 95% of coverage is for a learner to know around 4,000 word families, and to achieve 98% coverage, a learner needs to increase knowing word families up to 9,000 to 10,000 word families.

These base lists come ready with the software program Range, which is employed to analyse a corpus. These base lists consist of the basic form of the words and the derived forms. Another essential aspect of understanding the base lists is that word forms in the lists are organised into word families under Nation's (2006) and Coxhead (2000) headwords. For instance, the headword *cook* has the following family members *cooks*, *cooked* and *cooking*.

3.2 Academic words

When language learners reach the mastery of the high-frequency word and or the first 2,000 – 3,000 in English, it is smart to direct vocabulary training to specific areas in vocabulary learning (Nation, 2013). Some of these areas are the mid-frequency words and the academic words. The definition of academic vocabulary varies depending on the scholar; Barber et al. (1962) define it as useful scientific words; while others termed it as "sub-technical vocabulary" (Anderson, 1980) and "semi-technical vocabulary" (Farrell, 1990).

One of the most famous studies on academic vocabulary is investigated by Coxhead (2000), where she created an Academic Word List based on a compilation of 3.5 million running words of written academic texts. This list is a replacement of the previous University Word List created by Xue and Nation in 1984. Some samples of this vocabulary are *accumulate*, *achieve*, *adjusted*, *compound*, *proportion*, *policy*, and *sustained*, which generally constitute 9% of the running words in academic text and not elsewhere. Thus, learning the Academic World List is helpful for learners in coping with academic texts' vocabulary (Nation, 2013).

The importance of learning academic vocabulary is apparent, and it should be one of the goals in learning a language for several reasons (Nation, 2013). First, numerous academic words are widely used in academic texts but rarely present in general non-academic texts (Nation, 2013), so it is essential for language learners to be familiar with some of these words in preparation for future studies in higher levels. Second, academic words cover a significant amount of words in academic texts. Thus, knowing more academic words leads to a better understanding of the academic texts. Lastly, academic vocabulary refers to what the people in the academe often do, which relates to research. Hirsh (2004) employed the “review- method-result-discussion” separation of academic articles to show that specific words are associated with certain divisions of academic reading. Academic words are often employed to perform academic actions such as applying, reviewing, analysing, evaluating, critiquing, and describing. This is beneficial in recognising how these texts embody the nature of academic readings (Nation, 2013).

3.2.1 Academic base word list

As mentioned earlier, the Academic World List by Coxhead (2000), is a smaller corpus, particularly focusing on the words that occur mainly in academic texts. Regarding the word list, each sub-list contains 60 word families (c.f. 1000 word families in BNC). Like the base word lists

in BNC, these base word lists are ready-made with the program (Range), which will be described more in the method section. The academic base word list coverage is presented below.

Table 4. Academic World List coverage

Number of word families and sub-lists	Coverage in %
1 st 60 (words)	3.6%
2 nd 60	1.8%
3 rd 60	1.2%
4 th 60	0.9%
5 th 60	0.8%
6 th 60	0.6%
7 th 60	0.5%
8 th 60	0.3%
9 th 60	0.2%
10 th 60	0.1%
Total 570 words	10.0%

In Table 4, the total number of academic words is 570. As the sub-lists decline, the percentage of coverage also decreases. The words in sub-list 1 are the most frequent words; sub-list two as the next most frequent; and sub-list ten as the least frequent (Coxhead, 2000).

3.3 Comprehension, learning goals and the syllabus for English

The relationship between comprehension and coverage is strong. In other words, the more words that the learners know, the better understanding of the spoken and written language is (Milton, 2009). Therefore, it is fundamental to know the degree of lexical coverage needed to assist comprehension. In other words, it deals with how many words we need to know and what percentage of understanding it leads to.

Previous studies in this specific area examine the amount of vocabulary necessary for understanding different texts in various genres, both in spoken and written texts (Laufer, 1989; Waring, Schmitt, and McCarthy, 1997; Milton, 2009; Nation, 2001,2006). When it comes to spoken discourse, such as a movie or a lecture, a lexical coverage of 98% contributes to a meaningful understanding. The researchers suggest that a vocabulary scope of 5,000 to 9,000 word families is required for the learner to know to achieve a coverage of 98%.

In respect to written texts, Laufer (1989) suggests that vocabulary knowledge is the most significant area needed for comprehension. He argues that students need to be familiar with at

least 95% of the running words to ensure good reading comprehension. When computed, this means that one out of 20 words is not known to the reader. To attain 95% text coverage, English learners have to have a vocabulary size of 5,000 word families. Nation (2013) suggests a similar argument. There is a robust linear relationship between comprehension and coverage; the more words are known, the better comprehension. However, Nation (2013) demands a higher text coverage of 98% threshold, at which English learners can attain reasonable comprehension to be able to read for pleasure. As an illustration, this means that one out of 50 words is not known to the reader. Hence, a learner needs 8,000-9000 word families vocabulary to deal with written text if the ideal coverage is 98%.

Furthermore, not all texts compel the same vocabulary size to be understood. Some texts call for more word families, and some call for less. The text coverage dramatically relies on the genres of the text. Each genre (e.g. novel, newspaper, graded novel) demands different coverage. The table below explains this assertion further.

Table 5. Vocabulary size to achieve 95% and 98% coverage of various texts, including proper nouns (Nation, 2013).

Text	95%	98%	Proper nouns
Novels	4,000 word families	9,000 word families	1-2%
Newspapers	4,000 word families	8,000 word families	5-6%
Writing for children	4,000 word families	10,000 word families	3.5%
Children's movies	4,000 word families	6,000 word families	1.5%
Spoken English	3,000 word families	7,000 word families	1.3%

Table 5 shows that each text genre requires different word families to attain a certain level of coverage. As expected, the text coverage of 98% demands a significant degree of vocabulary knowledge across text genres. There is also a pretty giant leap of knowing word families, for instance, from 4,000 word families to 9,000 word families in reading novels. Yet, it only differs 2% in text coverage from 95% to 98%. This ensures the learner an adequate comprehension of what s/he is reading. According to Laufer (1989), a “reasonable” comprehension of texts requires a conservative 95% text coverage but Nation requires 98% text coverage. As we can see the notion of reasonable comprehension depends on the theorists and the corpus in use. However, the clear upshot is that learners must learn a good number of words in order to operate effectively

(Schmitt, 2010) and that if the student's primary goal is to extract information from texts, considerable comprehension is needed (Schmitt & Schmitt, 2014).

3.2.1 The syllabus for English

The main contents of the English syllabus at the elementary level aim at developing the students' knowledge of the language in areas and contexts where English is used; as well as to boost their confidence in their ability to use the language in different situations and for different purposes both in writing and speaking skills (Skolverket, 2011). Through the teaching of subject English, Skolverket (2011) states that students studying English shall be provided by the teacher in the classroom the conditions to:

- develop their ability to use English to communicate in speech and writing.
- deepen their understanding of spoken English in different situations and contexts.
- develop their ability to actively participate in conversations and written communication, express their thoughts in English and perceive the opinions and experiences of others.
- develop their ability to use English orally in different contexts to tell, describe, and motivate their opinions.
- develop their ability to read different kinds of texts for experiences, information and knowledge.
- develop their ability to express themselves varied and confidently in writing to tell, describe, and motivate their opinions.
- develop their ability to analyse, process and improve the language towards increasing variety and security.
- develop their ability to use aids and to critically examine sources of information.
- develop their ability to reflect on lifestyles and cultures in English-speaking countries and make comparisons with their own experiences.
- develop their ability to reflect on and take responsibility for their language learning and consciously use working methods that promote their learning.
- develop their ability to plan, implement and evaluate tasks independently and in collaboration with others (Skolverket, 2011, author's translation).

These guidelines on content are what teachers should be able to provide the students with inside their classroom.

As for the students' assessment, students finishing grade 9 have to satisfy the proficiency criteria to get a passing grade. The assessment criteria or knowledge requirements that students must achieve by the end of the 9th school year are listed in the next section. The student must be able to:

- clearly understand, even if something is regionally varied, speech in instructions, stories and descriptions concerning known conditions and own areas of interest.
- actively participate in conversations about known topics and contribute to the communication with the help of different strategies.
- express orally and describe something that he or she has seen, heard, experienced or read and express and argue for an opinion on something important to him or her.
- read and assimilate the content of simpler fiction and other narratives. descriptive and argumentative texts that deal with known subject areas.
- request and provide information in writing and tell and describe something.
- know everyday life, social and cultural traditions in some countries where English has a central position and can make some comparisons with their own cultural experiences.
- reflect on and conclude their way of learning English.
- choose and use aids for text reading, writing and other language activities.
- discuss on their own and collaborate with others in planning and carrying out tasks and thereby draw conclusions from their work (Skolverket, 2011, author's translation).

Students need to fulfil all these knowledge requirements so they can proceed to senior high school. Obviously, failure to satisfy the knowledge requirements leads to a failing grade.

Unfortunately, there have been very few studies dealing with the vocabulary size of English textbooks in Sweden and relating it to a syllabus or the knowledge requirements. However, what can be inferred from Skolverket's knowledge requirements is that it clearly calls for particular vocabulary size, but we do not know how large; and ideally, textbooks should provide enough vocabulary to meet the institution's prerequisites.

4. Material and method

This section discusses the materials created for study and the methods used in investigating the samples. Section 4.1 presents the data collection, section 4.2 deals with the data analysis, and section 4.3 explains the limitations of the study.

4.1 Data collection

The primary materials of this paper were two books used in grade 9 in teaching English as a Foreign Language. Based on their respective website's description, these two books were widely used in different schools across the country. Since there was no centralised data available that presented which English textbooks were most used in Sweden, any existing English textbooks in the market were assumed worthy as samples. The two textbooks were chosen because well-known publishers publish them. The descriptions of the sample textbooks are in the next paragraphs.

Good Stuff Gold D (hereafter, GSD) is the 3rd book of the Good Stuff series in grades 7 to 9. According to their website, the Good Stuff series is the most popular English book in Sweden throughout the years. This book is popular because of its fun texts, clear and straightforward structure. The books in the series are purposely created to achieve their goals in learning the English language. These books were written by Andy Coombs, Annika Bayard, Roland Hagvarn and Kjell Johansson in 2019, published by Liber AB.

Similarly, Glerups Engelska 9 (hereafter, GU9) is also the 3rd book of the Glerups series for grades 7-9. Glerups series is a comprehensive digital learning aid designed based on Lgr 11 by Skolverket. The books in this series are updated in terms of content and design following the syllabus. The series was written by Jeremy Taylor and published by Glerups Utbildning AB.

A license had to be purchased individually to access these books online. Each license allowed the user to have access to the material for 12 months. There were two kinds of licences available for purchase, the teacher's license and the student's license. The student license was selected for two primary reasons. Firstly, the student licence is significantly cheaper. Second, there is no apparent need to include the teacher's guide or extra material in the study. This study focused on the kind of vocabulary the students are exposed to when reading their textbooks. In practice, the students are not exposed to the teachers' guide. In this regard, the author fully acknowledges the actualities that there could be extra photocopyable materials that may be handed to the students.

The online version was preferred rather than the physical book for convenience. It is easier to convert the digital book into a text file for analysis rather than purchasing a hard copy

and scan it afterwards. For the procedure, the contents of the digital books were copy-pasted on a word document, but not all the contents could be accommodated; some components had to be excluded because of the limited time. This is further explained in the limitations section of this paper. Afterwards, the word documents were converted into text files. After the books were converted into text files, the files were processed by a software called Range.

4.1.1 Range program

Range software was a vocabulary analysis program developed by Nation and Heatley in 2002. This program, previously known as VocabProfile, was used in examining loads of vocabulary texts. Range was downloaded free of charge from the Victoria University of Wellington (New Zealand) website. This tool had built-in base word lists from different corpora, which compared the book text files when uploaded and processed. It automatically generated results regarding how many tokens, types, and families were found from the book text files corresponding to each base word list. Sample copies of the program and results are attached in the Appendix section of this paper so the reader can have a better perspective on how the software appears and operates.

4.2 Data analysis

The distinct advantage of quantitative research is that it permits us to compare and contrast several cases (e.g. tokens, word families) by employing easy numerical figure (Litosseliti, 2020). The quantitative method provides a concrete number substantial in mapping out cases in a corpus-based study such as this paper. The analysis of quantitative data generates a solid foundation for descriptions, and it enables interpretations and findings to be based on measured quantities rather than impressions; that is, it can be examined by others for authenticity (Denscombe, 2010). The interpretation and analysis were applied once Range did the calculations.

The primary analysis employed in this paper was corpus analysis. Applying corpus analysis was employed in broader research investigations in language teaching and language learning (McCarthy and O'Keeffe, 2010). The usage of this particular methodology generated perspectives that were way beyond the realm of grammar and lexis. In varied areas such as language acquisition, corpus linguistics was a plausible tool in finding answers to research questions. Scholars have used the corpus-analytic tool for the past decades (e.g. Sinclair 1991,

2004). The application of language corpora generated access to natural occurring language as corpus methods support exploratory and discovery learning (Bernardini 2004 as cited by Cheng 2010).

Cheng (2010) saw the corpus analysis application in two dimensions: capturing reality and providing realistic models for language learners. According to Braun (2005), corpus-based depiction of language contributes rich, illustrative, and updated data as a resource in developing materials used in teaching. Allan (2000) also suggested that corpus was essential in adjusting and staging input to address the learners' proficiency level. Based on all these assumptions, corpus analysis was believed the effective method to address inquiries in this paper.

For research question number 1, the corpus of British National Corpus was applied (s.v. type of words). As for research question number 2, the corpus of Academic Word List was used. For research question number 3, the pedagogical reflections of learning goals and syllabus objectives were interpreted based on the earlier outcomes of research question numbers 1 and 2.

4.3 Limitations

It should be maintained that only textbooks Good Stuff Gold D and Gleerups Engelska 9 were part of the examination; other books in the series are excluded. Furthermore, digital copies were the main source sample of this investigation. The primary samples were the written texts, but they did not include the online exercises. Unlike a regular book, online books were more complex to use. There are hyperlinks attached, images, videos, voice recordings which are often found in the exercises. These sections compelled a tremendous amount of time when copied to documents. These were excluded due to time constraints.

4.4 Ethical considerations

Since this corpus-based study did not involve people, there had been no ethical issues regarding protecting the participants' interests in general. However, one crucial ethical consideration of this paper was the copyright law of the books, which safeguarded the publishers' rights. The contents of the book were prohibited from being published without the consent of the publishers. This paper did not present nor introduce any texts implicating the substance of the content. The attached Appendix C and D were merely headwords of the vocabulary from the textbook based

on the corpus and were not presented in a manner where context is revealed. The texts were employed confidentially with consideration of the data protection principles.

5. Results and discussions

This section presents the results of the two books under investigation. Section 5.1 discusses the types of words; section 5.2 deals with the academic words. Lastly, section 5.3 explains the pedagogical reflections of the vocabulary content.

5.1 Word types based on British National Corpus

As for the types of words, each book is compared with the base word lists separately first, then the cumulative discussion follows. The results are presented in the table below.

Table 6. Glerups 9 types of words

Word list	Tokens/%	Types/%	Families
1 st 1000	31569/78.02	1760/40.44	896
2 nd 1000	3508/ 8.67	888/20.40	579
3 rd 1000	2034/ 5.03	508/11.67	386
4 th 1000	637/ 1.57	211/ 4.85	181
5 th 1000	269/ 0.66	116/ 2.67	102
6 th 1000	133/ 0.33	67/ 1.54	59
7 th 1000	146/ 0.36	57/ 1.31	49
8 th 1000	80/ 0.20	39/ 0.90	34
9 th 1000	56/ 0.14	26/ 0.60	21
10 th 1000	45/ 0.11	19/ 0.44	18
not in the lists	1987/ 4.91	661/15.19	?????
Total	40464	4352	2325

Table 6 presents the tokens, types, and families from the book Glerups 9. Key findings emerge that the 1st base word list has the highest tokens of 78 % coverage, 1,760 tokens with 40.44%, and 896 families. We can see a trend from this result. As the base word list increases to the 10th base word list, the tokens, types, and families also decrease. This finding hints that the mid-frequency and low-frequency words do not replete the book and that high-frequency words are utilised regularly. A similar pattern of results is obtained in Good Stuff Gold D, as shown in the table below.

Table 7. Good Stuff Gold D type of words

World List	Tokens/%	Types/%	Families
1 st 1000	35000/80.49	2099/37.30	948
2 nd 1000	3265/ 7.51	1181/20.99	708
3 rd 1000	1406/ 3.23	604/10.73	435
4 th 1000	534/ 1.23	288/ 5.12	233
5 th 1000	422/ 0.97	196/ 3.48	162
6 th 1000	185/ 0.43	112/ 1.99	91
7 th 1000	144/ 0.33	79/ 1.40	66
8 th 1000	80/ 0.18	46/ 0.82	38
9 th 1000	92/ 0.21	32/ 0.57	30
10 th 1000	40/ 0.09	27/ 0.48	23
not in the lists	2318/ 5.33	963/17.11	?????
Total	43486	5627	2734

The results in Table 7 lead to a similar conclusion in the previous table where the 1st base word list generates the highest tokens of 35,000 with 80% coverage, 20,99 types with 37.30% coverage and 948 families. Similar to Glerups 9, the percentage decreases as the word list increases up to the 10th. Both books also yield tokens that are not in the lists. These words are assumed to belong to 11th and above base word lists which are low-frequency words.

Table 8. Summation of vocabulary per book

	Glerups 9	Good Stuff Gold D	Total
Tokens	40, 464	43,486	83,950
Types	4,352	5,627	9,979*
Families	2,325	2,734	5,059*

*manually added on face value

As for Table 8, key findings reveal that Good Stuff Gold D has greater numbers in all categories. Good Stuff Gold D has 43,464 tokens and 5,627 types, while Gler Ups 9 has 40, 464 tokens and 4,352 types. However, the two books did not differ significantly in word families, Good Stuff Gold D has 2,734 families, and Gler Ups 9 has 2,325 word families. The difference is only 409 word families.

However, though there is a substantial gap between tokens, there is not much difference with word families. The assumption is that even though the words in the two books are similar, the words are used more frequently in Good Stuff Gold D. This trend also applies to the word

types. When we compute Good Stuff Gold D Gold types minus Gleerups 9 types, it results in a difference of 1,275 types. This is an essential finding in the understanding that extensive use of tokens feasibly generates larger types.

The two books are combined for analysis to gain a broader perspective of grade 9 English textbooks in Sweden. The table below illustrates the result.

Table 9. Cumulative result based on world list on British National Corpus

World List	Tokens/%	Types / %	Wf (2 books/Range)	Wf (2 books/face value)	Same wf	Different wf
1 st 1000	66,569/79.30	2,430/31.90	977	1844	867	110
2 nd 1000	6,773/8.07	1,590/20.87	836	1287	451	385
3 rd 1000	3,440/4.10	909/11.39	595	821	226	369
4 th 1000	1,171/1.39	450/5.91	352	414	62	290
5 th 1000	691/0.82	288/3.79	236	264	28	208
6 th 1000	318/0.38	164/2.15	136	150	14	122
7 th 1000	290/0.35	130/1.71	108	115	7	101
8 th 1000	160/0.19	82/1.08	68	72	4	64
9 th 1000	148/0.18	56/0.74	49	51	2	47
10 th 1000	85/0.10	43/0.56	37	41	4	33
not in the lists	4,305/5.13	1,475/19.39	(?????)	(?????)	(?????)	(?????)
Total	83,950	7,617	3,394	5,099	1,665	1,729

* wf: word families

In Table 9, the cumulative result is presented when the two books are merged, processed through the Range program as one file. The number of tokens is presented in the second column with their corresponding percentage in the text. For instance, the 66,569 tokens cover 79.30% of the running text. The types in the third column are also calculated the same way. The 2,430 types cover 31.90% of running text. The fourth column presents the base words found from the tokens and types.

As expected, the first 1,000 words generate the most tokens, the most types and most families. The second 1,000 words also provide the second number of tokens, the second number of types and families, and the trend continues. This pattern is similar as observed earlier: as the list moves towards the 10th, the rate decreases in all categories. A striking result from this table is that the total number of types and families does not equal Table 8's data, but only the tokens do. A brief explanation is that the two books contain similar types and word families, so they are counted as one by Range, but tokens, as previously defined, are counted per occurrence of the

word regardless if they similar or not. Thus, the tokens' value in Table 6, 7 and 8 remains consistent.

As for the fifth column in Table 9, these values represent the two books based on word families on Table 6 and Table 7 are added manually, not by Range. As for the sixth column, the values present the same word families found in the book. The word families by Range are subtracted by the word families added on face value which generates the same word families used by the two books to attain the result. For instance, 1844 is subtracted by 977 equals 867 word families. As anticipated, the remaining values result in word families' difference, as shown in column seven. The calculation for this is that the same word families is subtracted by the same word families by Range. For instance, 977 minus 876 equals 110 different word families.

Key findings emerge from the data above: the two books employed the most similar words in base word list one, while they significantly differ in the second and third base word lists. Although the first and second base word lists belong to the high-frequency words, this finding shows us that Gleerups 9 and Good Stuff Gold D textbook writers are almost parallel with using the first base word list but not in the second base word list. From this, we can infer that the first 1,000 word families are indispensable content. Regarding the second base word list, the significant difference of 385 word families suggests that though these word families are part of the high-frequency words, the two books treated this list disparagingly. The difference of 369 in the third base word list is also crucial. Based on Nation's (2013) proposition, this base word list belongs to the mid-frequency word. However, Schmitt and Schmitt (2014) argue that the 2,000 word families boundary is rather too low. They pointed out that although a learner can communicate fairly at 2,000 word families, the 3,000 word families represent a significant milestone in language learning. Therefore, the difference of the third base word families implies that one of the books prove to be very insufficient in high-frequency words if Schmitt and Schmitt's (2014) parameter is applied but maybe just insufficient (not very) in mid-frequency if Nation's (2013) parameter is used.

Furthermore, taking the data from Table 9 and relating it with Table 2 with the high, mid, and low-frequency words, the results are shown in Table 10.

Table 10. Cumulative result and vocabulary types

Vocabulary Types	Coverage by % (Nation)	Wf (2 books/face value)	Coverage by % (2 books)	GU9 %	GSD %
High-frequency / 2,000 wf	90%	1,844wf	87.37%	86.69%	88%
Mid-frequency / 7,000 wf	9%	1,544 wf	7.42%	8.29%	6.58%
Low-frequency/ 50,000 words	1%	37 wf	.10%	5.02%	5.42%

*wf: word families

*Gleerups 9: GU9

*Good Stuff Gold D: GSD

In table 10, the cumulative result from Gleerups 9 and Good Stuff Gold D is compared with the vocabulary types and their coverage. The 1st and 2nd word lists are added to get the high-frequency word families, the 3rd to 9th base word lists are added to the mid-frequency, and the 10th and not in the lists are added to get the low-frequency (see Tables 6 and 7 for the base word list results per book). This summation also applies in getting the coverage.

The important finding from this data is that the two books cover 87.37% out of ideal 90% coverage with 2,000 word families, which implies a very good coverage. Individually, the books also provide a very credible coverage with 86.69% and 88%. What is interesting in this finding is that even though the two books treated the second base word list differently, as pointed out in Table 9, it did not significantly affect so much in the high-frequency coverage because coverage depends on the frequency of words/tokens rather than the quality or the kind the headword. This might give us an idea that a good textbook must have rich word families with enough tokens of those rich words to increase coverage.

Good Stuff Gold D seems to significantly drop behind with only 6.58% in mid-frequency words. The lack of focus in mid-frequency is not something surprising. This result leads to a similar conclusion where Matsuka and Hirsh (2010) argues that upper-intermediate course books provide very few opportunities for learning vocabulary at mid-frequency level as well as Schmitt and Schmitt's (2014) investigation of New Head Way Upper-intermediate suggests that the book does not promote the acquisition of mid-frequency words.

On the other hand, Gleerups 9 seems to have given mid-frequency words an adequate space in their textbooks because the coverage is 8.29%, which is very high, almost reaching 9%. From this result, it is clear that textbooks writers vary in giving space to which vocabulary types

writers write in the textbooks. Some writers may prefer emphasising high-frequency words, while others may reduce high-frequency words to make space to the mid-frequency.

5.2 Academic words

The Academic World List draws from a broader range of texts from particular disciplines, requiring very specific vocabulary. The two books were run on Range with the Academic Word List. The results are presented below.

Table 9. Academic words per book

	Tokens /%	Types /%	Families
Gleerups 9	1082/ 2.67% (non-academic 97.33%)	174/ 4.00%	174
Good Stuff Gold D	756/ 1.74% (non-academic 98.26%)	194/ 3.45%	194

The data shown in Table 9 reveals that the two books contain less than 3% coverage of academic words. Gleerups 9 has 2.67% coverage, which is .93% higher than Good Stuff Gold D, with 1.74% coverage. Good Stuff Gold D has more families, more types, but fewer tokens than Gleerups 9. This finding supports the premise that though Good Stuff Gold D has more families and more types, it repeats less of its academic words because the percentage of the tokens are pretty low. This result is crucial in acquiring vocabulary because repetition of words has a vital role in learning.

Generally, the axiom in successful vocabulary learning is that a word must be recycled; and that coursebook designers need to repeat vocabulary in the textbooks if their books are to be effective (Milton, 2009). The more writers repeat a word in a textbook; the more learners are presumably exposed to it; thus, the more students are likely to learn it (Milton, 2009). According to Nation (1993), only a few textbooks designers systematically recycle vocabulary because of the volume of vocabulary presented in textbooks. Consequently, it proves impossible to repeat every item in strict and systematic order. Overall, the data suggest that both books score low in academic word tokens, types, and families, but there are no standard criteria that indicate how many academic words should be introduced in a course book.

The non-academic vocabulary in Table 9, with very high coverage of 97.33% and 98.28% frames a question if there are words from the Academic Word List which may already belong to

the high, mid, and low-frequency words. Coxhead (2000) explains that the words in the Academic Word List “does not include words in the most frequent 2000 words of English”, but this is based on the previous General Service List created by West in 1953 and not any recent corpus. So, it is worth processing the Academic Word List on the British National Corpus to find out the uncertainty. Then, relate the results with the data in Table 9 on the individual books to determine if there are academic words that may already belong to the high, mid, and low-frequency classifications based on BNC.

Table 10. Academic Word List in BNC

Vocabulary Type	Coverage %	GU9 in BNC	GSD in BNC	AWL in BNC
High-frequency / 2,000 wf	90%	86.69%	88%	28.8%
Mid-frequency / 7,000 wf	9%	8.29%	6.58%	70.37%
Low-frequency/ 50,000 words	1%	5.02%	5.42%	.82%

*wf: word families

*Gleer Ups 9: GU9

*Good Stuff Gold D: GSD

Perhaps the most striking result in Table 10 is that the words found in Academic Word List (Coxhead, 2000) are found in British National Corpus. Column 4 in Table 10 shows that 28.8% of academic words belong to the high-frequency word, an enormous number of 70.3% of academic words belongs to the mid-frequency, and a meagre rate of .82% belongs to the low-frequency. This result implies that the learner also touches the sphere of the academic words by learning the high-frequency words. That means, if a learner knows the first 2,000 word families, s/he also covers 28.8% academic words already (based on BNC nonetheless). However, learning mid-frequency words suggests that more academic words are covered. From this finding, we can infer that most academic words belong to mid-frequency words. So, if a textbook provides a tiny space for mid-frequency words, lesser academic words are incorporated consequently. Therefore, Nation’s (2013) 9% text coverage in BNC deficits primarily by 61.37% in Academic World List. Although Gleerups 9 provided excellent coverage of 8.29% (nearly 9% on BNC), it proves pretty inadequate when compared with the 70.37% academic words. On the contrary, if Schmitt and Schmitt’s (2014) parameter of 3,000 high-frequency word families is considered, more academic words are subsumed. As one can see, the notion of word frequency varies from theorist to theorist. In reality, this is crucial in determining which theorist textbook designers adhere to in writing a textbook. A writer may consider the 2,000 high-frequency as the main input, while

another writer might consider the 3,000 high-frequency vocabulary input. What is important for now is to realise to what degree these textbooks prepare the learner for reading academic texts.

Furthermore, the low-frequency words seem not to contain a considerable amount of academic words, as the data shows. Thus, this finding might suggest that academic low-frequency words are less meaningful and scarcely relevant. From this standpoint, however, it is necessary to indicate that some low-frequency words may belong to the keywords of an academic text which makes it very crucial in understanding the text (Milton, 2009). Not knowing keywords in a text may affect a learner's comprehension.

To give the data in Table 10 another perspective, we can say that not knowing academic words eventually reduces text coverage. For instance, in Gleerups 9 with a 9 86.69% high-frequency coverage, if a learner does not know any academic words, we can subtract 28.8% from 86.69%, which generates 57.89% coverage. This coverage is deficient and is far from the minimal coverage of 95% or the reasonable coverage of 98%. At which, this puts the learner in a very unfavourable situation in terms of comprehension. Hence, the words from the Academic Word List are not only crucial in reading academic texts but are essential in completing the high and mid-frequency words coverage.

5.3 Pedagogical reflections: learning goals and school syllabus

As illustrated in Table 5, different reading genres (e.g. novels, newspapers) demand various word families to attain a thorough coverage, whether 95% or 98%. Good textbooks must contain large quantities of vocabulary Milton (2009), and ideally, textbooks used in the classroom should equip learners to attain these substantial coverages.

The data present that the two books under investigation primarily offer 2,325 word families for Gleerups 9 and 2,734 word families for Good Stuff Gold D (see Table 6). This result demonstrates two things concerning learning opportunities and vocabulary acquisition. The books provide vocabulary based mainly on high-frequency words and less on others, as explained earlier. When word families in the two books contrast with the different genres, the book's word families seem little, as shown in Table 11.

Table 11. Vocabulary size to achieve 95% and 98% coverage of various texts, including proper nouns and textbooks results.

Text	95%	98%	GU9	GSD
Novels	4,000 word families	9,000 word families	2,325 wf	2,734wf
Newspapers	4,000 word families	8,000 word families	2,325 wf	2,734wf
Writing for children	4,000 word families	10,000 word families	2,325 wf	2,734wf
Children's movies	4,000 word families	6,000 word families	2,325 wf	2,734wf
Spoken English	3,000 word families	7,000 word families	2,325wf	2,734wf

*wf: word families

The table above presents the comparison of word families from the two books under examination and the different text genres with their equivalent word families needed for comprehension. Novels demand 9,000 word families. To achieve 98% coverage. For comparison, GU9 and GSD rate below 3,000 word families, implying a shortage of word families input. To illustrate, if we take 9,000 (98%) word families in a novel subtracted by GU9 word families 2,325, this equals a lack of 6,675 word families. From this result, it is clear that both books display a shortage of word families in achieving both 95% coverage or 98% coverage. If these two books are the only source of vocabulary input, the learner is not exposed to enough words to equip them for authentic texts. However, in general, textbooks in the classroom are not the only source of vocabulary input by the learner. Other media such as movies, television shows and songs are also ways of learning vocabulary both inside and outside the classrooms. Since, textbooks are still the principal source of vocabulary input in the classroom (Milton, 2009) the kind of vocabulary the learners are exposed to from the textbooks needs to be maximised.

Regarding learning goals, the two textbooks provide an adequate amount of vocabulary to equip the learner to use the language, at least on this grade level. Laufer and Ravenhorst-Kalovski (2010) point out the significance of setting vocabulary goals based on learners' comprehension or proficiency level. Nation (2013) points out that if a native speaker knows about 20,000 word families or the high-frequency words, to which a second language learner seems to be very ambitious, the vocabulary threshold of 2,000 word families is sufficient to enable the learner to use the language. If Nation's (2013) proposition is assumed, the two books satisfy the expectations. This being said, it is important to note again that the academic words

were found mainly in mid-frequency words. Schmitt (2005) further argues that native-like competence is rather demanding to achieve for second language learners. Therefore, a second language learner does not have to attain a native-like vocabulary size to use English (Milton, 2009), but he points out a learner's vocabulary size is highly relevant to one's language overall performance. So, if a learner wants to achieve extraordinary language skills, s/he needs to know more words than the threshold of 2,000 word families.

One of the criteria upon finishing grade 9 is that a student must satisfy Skolverket's (2011) criteria, to repeat here, to "be able to read and assimilate the content of simpler fiction and other narratives, descriptive and argumentative texts that deal with known subject areas". This criterium suggests that a learner might not need so many words as he is expected to read only "simpler fiction", yet "other narratives" suggests variation and perhaps complexities of texts, especially in literature which might include more word families. In this regard, since authentic texts such as novels require 9,000 word families, which is too pressing, it is assumed by the author that Skolverket's requirement is directed towards reading graded readers or simplified literature texts.

Concerning the assessment criteria by Skolverket's (2011) of the expected proficiency of a learner upon finishing 9th grade, there is no concrete data on how much vocabulary size a student needs to get a specific grade (e.g. A, B, C, D, E or F). There are also no standard vocabulary tests given to students on this grade level except the English national tests but knowing the number of word families in national tests is not accessible. In other words, we do not know how many word families a 9th-grade learner needs to get a passing grade in national tests or the final grade. However, a small study conducted by Kakaee (2018) in upper-secondary or English 5 and English 6 shows that the vocabulary size of the Swedish English learners ranges from 6,000 word families in English 5 to 6,550 word families in English 6. We can roughly assume that for 9th graders to get a passing grade or pass national tests, s/he needs around 5,000-6,000 word families. This number remarkably corresponds to 95% coverage of the different genres mentioned in Table 12. However, Kakaee (2018) suggests that students perform unsatisfactorily in English national tests.

Another poignant reflection could be that if textbooks vocabulary content provides approximately 2,000- 2,700 word families and that students in English 5 and 6 have an estimated 5,000-6,000 word families learned, this contributes to the idea of extramural language learning.

Learners acquire a considerable amount of vocabulary outside school, as shown in gaming studies by Sylven and Sundqvist in 2012. When calculated, the 6,000 word families learned in English 5 and 6 subtracted by the 3,000 (max) word families from the textbook vocabulary input; this equals 3,000 word families. These 3,000 word families may be attributed to extramural learning. This may raise an assumption that a student learns an equal amount of word families outside school as s/he is inside. Although several studies prove the impact of extramural English in language acquisition, the exact number of word families learned by a learner outside the classroom remains unknown. Hence, the author’s assumption that learners learn a proportional amount of vocabulary outside the classroom as inside remains preliminary as the boundaries of what is learned inside and outside is hazy. On the other hand, Puimège and Peters (2019) points out that learners have more extramural English as they grow older and their vocabulary sizes increase as they age. With this finding, it may be possible that at age 16 (English 5) or 17 (English 6), the extramural learning is as competitive as what they learn in school.

Since studies of vocabulary size linked to specific examination or grades in a Swedish setting, the Common European Framework of Reference (hereafter, CEFR) developed by the Council of Europe in 2001 to encourage reflection of what the word families found from the two books implies. Meara and Milton (2003) built vocabulary size scores using the X-Lex test in correspondence to CEFR levels. The table below presents their findings.

Table 12. CEFR Levels and EFL Vocabulary size (Milton, 2009)

CEFR Level	- user	Words
A1	(basic)	<1,500
A2	(basic)	1,500-2,500
B1	(independent)	2,500- 3250
B2	(independent)	3,250- 3,750
C1	(proficient)	3,750- 4,500
C2	(proficient)	4,500- 5,500

Table 12 illustrates the link between vocabulary size to CEFR examination levels. As a basic user, level A1 is the lowest proficiency level, requiring learners to know at least 1,500 words and the highest level, as a proficient user, requiring learners 4,500-5,500 words. By comparing the two books’ word families (2,325 GU9 and 2,734 GSD) to the CEFR levels, these books afford learning vocabulary that coincides with B1 or B2 independent users. A description

of B users is provided below, while the full descriptions of all the levels are attached in the Appendix section of this paper if there is a need for reference. In level CEFR (2001) global description, B1 is described as a learner who:

- can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure;
(c.f. clearly understand, even if something regionally coloured, speech in instructions, stories and descriptions concerning known conditions and own areas of interest, Skolverket 2011).
- can deal with most situations likely to arise whilst travelling in an area where the language is spoken;
- can produce simple connected text on familiar topics or of personal interest;
- can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
(c.f. be able to orally tell and describe something that he or she has seen, heard, experienced or read and express and argue for an opinion on something that is important to him or her, Skolverket, 2011).

B2 users as a learner who:

- can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation;
- can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party;
(c.f. be able to actively participate in conversations about known topics and with the help of different strategies contribute to the communication, Skolverket, 2011).
- can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.

Although there are fundamental similarities between Skolverket's criteria in passing grade 9 and the B1 and B2 level descriptions, no direct assumption can be made if B1 and B2 are the levels required to pass grade 9 in Sweden. What can be inferred from the data is that two sample books vocabulary is equivalent to B1 and B2 proficiency, nothing more; and that to become a C1 or C2 proficient or advanced user, a book with larger vocabulary size is required. According to Milton (2009), the more vocabulary is presented in textbooks; the more vocabulary learners seem to

acquire. To repeat, the importance of vocabulary in a textbook is evident. A textbook containing a considerable amount of words is ideal for learners if higher proficiency is the primary goal.

6. Conclusions

The importance of vocabulary learning is an integral part of language acquisition. A learner with a more extensive vocabulary performs better than a learner with a smaller vocabulary. In this regard, the weight of textbook selection in the classroom needs careful evaluation by the teacher. The teacher must choose a textbook with rich vocabulary input that considers the different vocabulary types and the academic words to maximise the student's language acquisition.

As for research question number one, Glerups 9 and Good Stuff Gold D differ in tokens, types, and word families, but they do not contrast greatly. Gler Ups 9 has 40,464 tokens, 4,352 types and 2,325 word families, while Good Stuff Gold D has slightly higher with 43,464 tokens, 5,527 tokens and 2,734 word families. Both books rate high in high-frequency words while Glerups 9 rate higher than Gold Stuff D in mid-frequency and both rate fairly in low-frequency.

Concerning research question number two and three, the sample books Glerups 9 and Good Stuff Gold D present between 2,325 to 2,734 word families per book based and less than 3% tokens or less than 200 word families text are academic words. On this basis, it is concluded that the books cover high-frequency words, which is enough for the learner to use English. However, it does not suggest a good preparation for academic readings since most academic words belong to the mid-frequency, which was revealed when the Academic Word List was compared with British National Corpus. In addition, the word families in the textbooks equip the learners to be B1 and B2 users in the CEFR model. Nevertheless, no conclusion can be drawn if B1 and B1 levels are the exact levels equivalent to get a passing grade set by Skolverket. Overall, the results demonstrate that the two books encompass the most frequent words in the corpus.

Future studies could fruitfully explore the issue further by including other school materials such as workbooks; include online interactive exercises; online audio materials would be broader in scope. The author believes that future research should look for word families and word types in national tests apart from looking for word types in school textbooks. In addition, an examination of the relationship between the numbers of word families and grades (e.g. the number of word families assumed known to get A) as a learning outcome in English is also a strong recommendation. This is a pivotal component to overcome learning difficulties. A

longitudinal study in the extramural acquisition of vocabulary (e.g. the amount of vocabulary learned outside school) with age as a critical variable is also a strong recommendation. This is beneficial for teachers because it allows them to include extramural learning in designing their lessons inside the classroom; that is, if they know their students are learning certain types of vocabulary outside their class, their lesson could focus on another aspect of vocabulary learning. Second, teachers can also propose certain games to play or movies to watch as a form of relaxed homework without the pressure of being tested afterwards. Lastly, investigating how the different variations of English are presented in textbooks, whether they reinforce the myth of “native speakerness” or not, might prove a vital area for future examination.

References

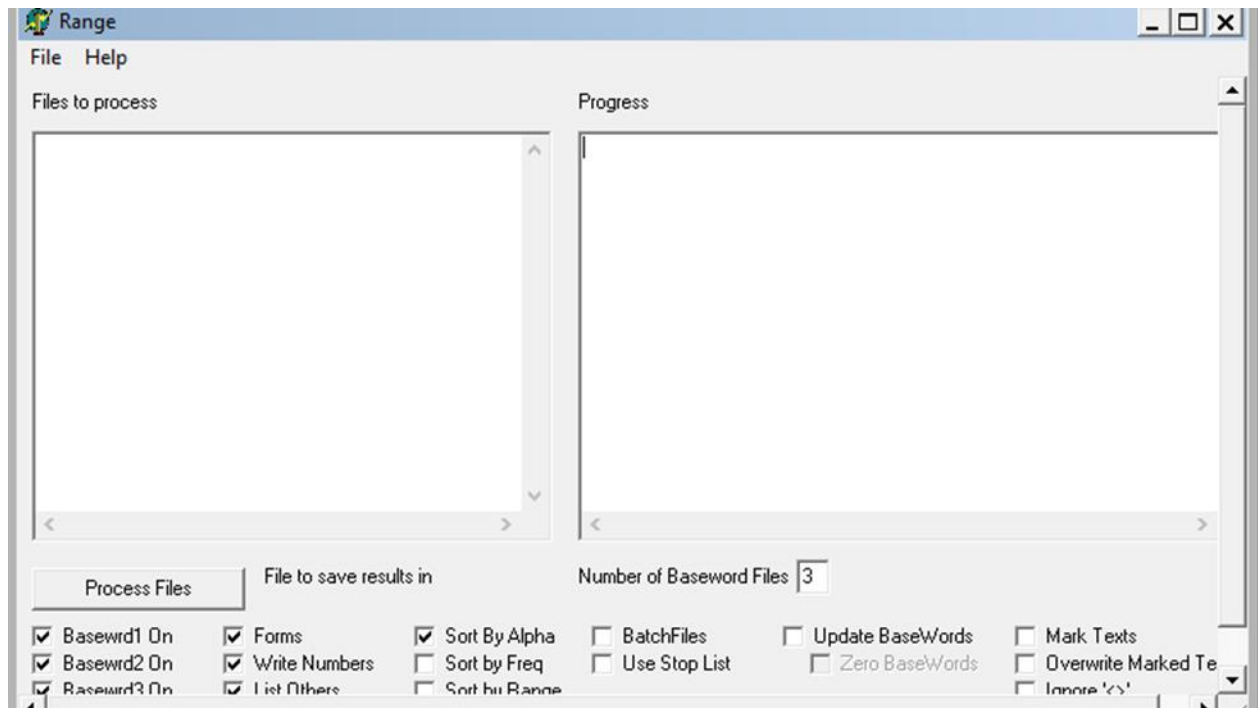
- Abello-Contesse, C., & López-Jiménez, M. D. (2010). The treatment of lexical collocations in EFL textbooks. *Exploring new paths in language pedagogy: Lexis and corpus-based language teaching*, 95-109.
- Alhudithi, E. (2017). *Corpus-based analysis of English vocabulary input provided in K-12 textbooks used in Saudi Arabia*, A (Doctoral dissertation, Colorado State University).
- Allan, R. (2009). Can a graded reader corpus provide 'authentic input'?. *ELT Journal*, 63(1), 23-32.
- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. *Comprehension and teaching: Research reviews*, 77-117.
- Andersson, Ma. (2013). English in Sweden: English as a Second Language in Sweden in a Theoretical Perspective (Dissertation). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-22378>
- Barber, C. L. (1962). Some measurable characteristics of modern scientific prose. *Contributions to English Syntax and Phonology*, 1-23.
- Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17, 47-64.
- Canguro English. (2019, August 21). *The myth of native speaker (with David Crystal)*. [Video] Retrieved from <https://www.youtube.com/watch?v=p-kZLP2FWUI>
- Carter, R., & McCarthy, M. (2014). *Vocabulary and language teaching*. Routledge.
- Cheng, W. (2010). What can a corpus tell us about language teaching. *The Routledge handbook of corpus linguistics*, 319-332.
- Coombs, A., Bayard, A., Johansson, K. & Hagvörn, R. (2019). *Good Stuff Gold D Digital (elevlicens) 12 mån*. Liber.
- Council of Europe. (2001). The CEFR Levels.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213
- Denscombe, M. (2010). *The good research guide [Elektronisk resurs] for small-scale social research projects*. (4th ed.) Maidenhead: Open University Press.
- ENGL PolyU. (2013, March 19). *In conversation with Prof. Alan Davies*. [Video]. YouTube. Retrieved from <https://www.youtube.com/watch?v=gPr254HbXHs>
- Farrell, P. (1990). Vocabulary in ESP: A Lexical Analysis of the English of Electronics and a Study of Semi-Technical Vocabulary. *CLCS Occasional Paper No. 25*.
- Flores N. (2013). Silencing the Subaltern: Nation-State/Colonial Governmentality and Bilingual Education in the United States. *Critical Inquiry in Language Studies*, 10:4, 263-287, DOI: 10.1080/15427587.2013.846210
- Hirsh, D. J. (2004). *A Functional Representation of Academic Vocabulary: A Thesis Presented to the Victoria University of Wellington in Fulfilment of the Requirements for the Degree of Doctor of Philosophy in Applied Linguistics* (Doctoral dissertation, Victoria University of Wellington).
- Kakae, Y. (2018) The comprehensibility of reading texts in National Tests in English. Retrieved from *GUPEA*: <http://hdl.handle.net/2077/58053>
- Lainio, J. (2000). Protection and rejection of minority majority languages in the Swedish system. *Current Issues in Language & Society*, 7(1), 32-50.

- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension. *Special language: From humans thinking to thinking machines*, 316323.
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension.
- Litosseliti, L. (Ed.). (2020). *Research methods in linguistics*. Bloomsbury Publishing.
- Meara, P. M., & Milton, J. (2003). *X-lex: the Swansea levels test*. Express Publishing.
- McCarthy, M., & O'Keeffe, A. (2010). What are corpora and how have they evolved. *The Routledge handbook of corpus linguistics*, 3-13.
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.
- Nation, I. S. P. (1993). Vocabulary size, growth, and use. *The bilingual lexicon*, 6.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P., Waring, R., Schmitt, N., & McCarthy, M. (1997). Vocabulary: Description, acquisition and pedagogy. *Japan: Notre Dame Seishin University*, 6.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening?. *Canadian modern language review*, 63(1), 59-82.
- Nation, I.S.P. (2013). *Learning vocabulary in another language*. (Second edition). Cambridge: Cambridge University Press.
- Nordlund, M. (2016). EFL textbooks for young learners: A comparative analysis of vocabulary. *Education Inquiry*, 7(1), 27764.
- Puimège, E, & Peters, E. (2019). Learners' English Vocabulary Knowledge Prior to Formal Instruction: The Role of Learner-Related and Word-Related Variables. *Language Learning*, 69(4), 943-977
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Springer.
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4).
- Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. *Vocabulary and language teaching*, 140, 60.
- Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *Recall*, 2012, Vol. 24, Iss. 3, Pp. 302-.321, 24(3), 302-321.
- Skolverket. (2006). *Läromedlens roll i undervisningen. Grundskollärares val, användning och bedömning av läromedel i bild, engelska och samhällskunskap*. Stockholm: Skolverket
- Skolverket. (2011). *Läroplan och kursplaner för grundskolan, engelska 2011*. Retrieved 2021-4-22 from <https://www.skolverket.se/undervisning/grundskolan>
- Takač, Pavičić V. (2008). *Vocabulary learning strategies and foreign language acquisition*. Clevedon: Multilingual Matters.
- Taylor, J. (2018). *Gleerups nya engelska 7-9, digital, elevlicens 12 mån*. Gleerups Utbildning AB.
- The Hyperpolyglot Activist_Dr.Yebra Lopez. (2021). *The "native speaker" myth*. [Video]. Retrieved from <https://www.youtube.com/watch?v=aGJkmICot0>
- University of Oxford. (2015). *British National Corpus*. Retrieved 2021-4-22 from <http://www.natcorp.ox.ac.uk/corpus/>

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language learning and communication*, 3(2), 215-229.

Appendices

Appendix A Range program



Sample result

The screenshot shows the output window of the Range application. It displays a table with four columns: 'WORD LIST', 'TOKENS/%', 'TYPES/%', and 'FAMILIES'. Below the table is a list of baseword files and their statistics.

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
one	31569/78.02	1760/40.44	896
two	3508/ 8.67	888/20.40	579
three	2034/ 5.03	508/11.67	386
four	637/ 1.57	211/ 4.85	181
five	269/ 0.66	116/ 2.67	102
six	133/ 0.33	67/ 1.54	59
seven	146/ 0.36	57/ 1.31	49
eight	80/ 0.20	39/ 0.90	34
nine	56/ 0.14	26/ 0.60	21
ten	45/ 0.11	19/ 0.44	18
not in the lists	1987/ 4.91	661/15.19	?????
Total	40464	4352	2325

Number of BASEWRD1.txt types: 6857 Number of BASEWRD1.txt families: 1000
Number of BASEWRD2.txt types: 6370 Number of BASEWRD2.txt families: 1000
Number of BASEWRD3.txt types: 5880 Number of BASEWRD3.txt families: 1000
Number of BASEWRD4.txt types: 4865 Number of BASEWRD4.txt families: 1000
Number of BASEWRD5.txt types: 4294 Number of BASEWRD5.txt families: 1000
Number of BASEWRD6.txt types: 4102 Number of BASEWRD6.txt families: 1000
Number of BASEWRD7.txt types: 3679 Number of BASEWRD7.txt families: 1000
Number of BASEWRD8.txt types: 3419 Number of BASEWRD8.txt families: 1000
Number of BASEWRD9.txt types: 3196 Number of BASEWRD9.txt families: 1000
Number of BASEWRD10.txt types: 2982 Number of BASEWRD10.txt families: 1000

Table of Ranges: Types

Appendix B CEFR Descriptions (Council of Europe, 2001)

PROFICIENT USER	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
INDEPENDENT USER	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
BASIC USER	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Appendix C Glerups 9 headwords base list 1

A	AWAY	BRIGHT	COMPLETE	DREAM	FELLOW	GOOD
ABLE	AWFUL	BRING	COMPUTER	DRESS	FEW	GOODBYE
ABOUT	BABY	BROTHER	CONCERN	DRINK	FIELD	GOVERN
ABOVE	BACK	BROWN	CONSIDER	DRIVE	FIGHT	GRANDFATHER
ABSOLUTE	BAD	BUILD	CONTINUE	DROP	FIGURE	R
ACCEPT	BAG	BURN	CONTROL	DRUG	FILL	GRASS
ACROSS	BALL	BUS	CONVERSATION	DRY	FILM	GREAT
ACT	BANK	BUSH	N	DURING	FINAL	GREEN
ACTUAL	BAR	BUSINESS	COOK	EACH	FIND	GREY
ADD	BASE	BUSY	COOL	EAR	FINE	GROUND
ADDRESS	BASIC	BUT	CORNER	EARLY	FINGER	GROUP
ADMIT	BATH	BUY	COST	EARTH	FINISH	GROW
ADVERTISE	BE	BY	COULD	EAST	FIRE	GUESS
AFFORD	BEACH	CAKE	COUNT	EASY	FIRST	GUN
AFRAID	BEAR	CALL	COUNTRY	EAT	FISH	GUY
AFTER	BEAT	CAMP	COUPLE	EDGE	FIT	HAIR
AFTERNOON	BEAUTY	CAN	COURSE	EDUCATE	FIVE	HALF
AGAIN	BECAUSE	CAR	COURT	EGG	FIX	HALL
AGAINST	BECOME	CARD	COVER	EIGHT	FLAT	HAND
AGE	BED	CARE	CRAZY	EITHER	FLOOR	HANDLE
AGO	BEFORE	CARRY	CRIME	ELEVEN	FLOWER	HANG
AGREE	BEGIN	CASE	CROSS	ELSE	FLY	HAPPEN
AHEAD	BEHIND	CAT	CRY	EMPLOY	FOLLOW	HAPPY
AIR	BELIEVE	CATCH	CUP	EMPTY	FOOD	HARD
ALL	BELOW	CAUSE	CUT	END	FOOT	HARDLY
ALLOW	BENEATH	CENTRE	DAD	ENGINE	FOOTBALL	HAT
ALMOST	BESIDE	CERTAIN	DANCE	ENJOY	FOR	HATE
ALONE	BEST	CHAIR	DANGER	ENOUGH	FORCE	HAVE
ALONG	BET	CHANCE	DARK	ENTER	FOREST	HE
ALREADY	BETTER	CHANGE	DARLING	ESPECIALLY	FORGET	HEAD
ALRIGHT	BETWEEN	CHARGE	DATE	EVEN	FORM	HEALTH
ALSO	BEYOND	CHEAP	DAUGHTER	EVENING	FORTUNATE	HEAR
ALTHOUGH	BIG	CHECK	DAY	EVER	FORWARD	HEART
ALWAYS	BILL	CHICKEN	DEAD	EVERY	FOUR	HEAT
AMAZE	BILLION	CHILD	DEAL	EXACT	FREE	HEAVY
AMONG	BIRD	CHIP	DEAR	EXCEPT	FREEZE	HELL
AMOUNT	BIRTH	CHOICE	DEATH	EXCITE	FRESH	HELLO
AND	BIT	CHOOSE	DECIDE	EXCUSE	FRIDAY	HELP
ANGRY	BLACK	CHRISTMAS	DEEP	EXPECT	FRIEND	HERE
ANIMAL	BLOOD	CHURCH	DEFINITE	EXPENSIVE	FRIGHT	HIDE
ANOTHER	BLOW	CITY	DEGREE	EXPERIENCE	FROM	HIGH
ANSWER	BLUE	CLASS	DELICIOUS	EXPLAIN	FRONT	HILL
ANY	BOARD	CLEAN	DEPEND	EXPRESS	FULL	HISTORY
APART	BOAT	CLEAR	DIE	EXTRA	FUN	HIT
APPARENT	BODY	CLIMB	DIFFERENCE	EYE	FURTHER	HOLD
APPEAR	BONE	CLOCK	DIFFERENT	FACE	GAME	HOLE
AREA	BOOK	CLOSE	DIFFICULT	FACT	GARDEN	HOLIDAY
ARM	BORING	CLOSED	DIG	FAIR	GAS	HOME
AROUND	BORN	CLOTHES	DINNER	FALL	GENERAL	HONEST
ARRANGE	BOTH	CLUB	DIRTY	FAMILY	GENTLE	HONOUR
ARRIVE	BOTHER	COAT	DISCOVER	FAR	GET	HOPE
ART	BOTTLE	COFFEE	DO	FARM	GIRL	HORRIBLE
AS	BOTTOM	COLD	DOCTOR	FAST	GIVE	HORSE
ASHAMED	BOX	COLLECT	DOG	FAT	GLAD	HOSPITAL
ASK	BOY	COLLEGE	DOOR	FATHER	GLANCE	HOT
AT	BREAD	COLOUR	DOUBLE	FAVOURITE	GLASS	HOUR
AUNT	BREAK	COME	DOUBT	FEAR	GO	HOUSE
AUTUMN	BREAKFAST	COMFORT	DOWN	FEED	GOD	HOW
AWARE	BREATH	COMPANY	DRAW	FEEL	GOLD	HOWEVER

HUGE	LET	MONEY	OPEN	PROMISE	SAY	SMELL
HUMAN	LETTER	MONTH	OR	PROPER	SCARE	SMILE
HUNDRED	LEVEL	MORE	ORANGE	PROTECT	SCHOOL	SMOKE
HUNGER	LIE	MORNING	ORDER	PUBLIC	SCIENCE	SNOW
HUNT	LIFE	MOST	OTHER	PULL	SEA	SO
HURRY	LIFT	MOTHER	OUGHT	PUSH	SEAT	SOFT
HURT	LIGHT	MOUNTAIN	OUT	PUT	SECOND	SOME
HUSBAND	LIKE	MOUTH	OVER	QUARTER	SECURE	SON
I	LINE	MOVE	OWN	QUEEN	SEE	SONG
ICE	LIP	MOVIE	OWNED	QUESTION	SEEM	SOON
IDEA	LIST	MRS	PACK	QUICK	SELF	SORRY
IF	LISTEN	MUCH	PAGE	QUIET	SELL	SORT
IMAGINE	LITTLE	MUM	PAIN	QUITE	SEND	SOUND
IMPORTANT	LIVE	MUSIC	PAINT	RABBIT	SENSE	SOUTH
IN	LOAD	MUST	PAIR	RACE	SERIOUS	SPACE
INDEED	LOCAL	NAME	PAPER	RADIO	SERVE	SPEAK
INFORM	LOCK	NATION	PARDON	RAIN	SERVICE	SPECIAL
INSIDE	LONG	NATURE	PARENT	RAISE	SET	SPEND
INSTEAD	LOOK	NAUGHTY	PARK	RATE	SETTLE	SPORT
INSURE	LORD	NEAR	PART	RATHER	SEVEN	SPOT
INTEREST	LOSE	NEAT	PARTICULAR	REACH	SEVERAL	SPRING
INTERNET	LOT	NECESSARY	PARTY	READ	SEX	SQUARE
INTO	LOUD	NECK	PASS	READY	SHAKE	STAGE
INVOLVE	LOVE	NEED	PAST	REAL	SHALL	STAND
ISLAND	LOVELY	NEIGHBOUR	PAY	REALISE	SHAPE	STAR
ISSUE	LOW	NEVER	PENNY	REALLY	SHARE	STARE
IT	LUCK	NEW	PEOPLE	REASON	SHE	START
JOB	LUNCH	NEWS	PERFECT	RECENT	SHIP	STATE
JOIN	MACHINE	NEXT	PERHAPS	RECORD	SHIRT	STATION
JOKE	MAD	NICE	PERSON	RED	SHOE	STAY
JUDGE	MAIN	NIGHT	PHOTOGRAPH	RELATE	SHOOT	STEAL
JUMP	MAJOR	NINE	PICK	REMEMBER	SHOP	STEP
JUST	MAKE	NO	PICTURE	RENT	SHORT	STICK
KEEP	MAN	NOBODY	PIECE	REPLY	SHOULD	STILL
KEY	MANAGE	NOISE	PLACE	REPORT	SHOULDER	STONE
KICK	MANY	NONE	PLAN	RESPONSIBLE	SHOUT	STOP
KID	MARK	NORMAL	PLANT	REST	SHOW	STORE
KILL	MARKET	NORTH	PLAY	RETURN	SHUT	STORY
KIND	MARRY	NOSE	PLEASE	RICH	SHY	STRAIGHT
KING	MASTER	NOT	PLENTY	RID	SICK	STRANGE
KISS	MATTER	NOTE	PLUS	RIDE	SIDE	STREET
KITCHEN	MAY	NOTHING	POINT	RIGHT	SIGHT	STRIKE
KNOCK	MAYBE	NOTICE	POLICE	RIGHTS	SIGN	STRONG
KNOW	MEAL	NOW	POOR	RING	SILLY	STUDENT
LADY	MEAN	NUMBER	POP	RISE	SIMPLE	STUDY
LAKE	MEET	NURSE	POSITION	RIVER	SINCE	STUFF
LAND	MEMBER	OBVIOUS	POSSIBLE	ROAD	SING	STUPID
LARGE	MENTION	ODD	POST	ROCK	SINGLE	SUBJECT
LAST	MESS	OF	POT	ROLL	SIR	SUCH
LATE	MIDDLE	OFF	POUND	ROOM	SISTER	SUDDEN
LAUGH	MIGHT	OFFER	POWER	ROUGH	SIT	SUGGEST
LAW	MILE	OFFICE	PREPARE	ROUND	SITUATION	SUIT
LAY	MILK	OFFICER	PRESENT	RUBBISH	SIX	SUMMER
LAZY	MILLION	OFTEN	PRESS	RULE	SIZE	SUN
LEAD	MIND	OIL	PRETTY	RUN	SKIN	SUNDAY
LEARN	MINUTE	OK	PRICE	SAD	SKY	SUPPORT
LEAST	MISS	OLD	PRINCE	SAFE	SLEEP	SUPPOSE
LEAVE	MISTAKE	ON	PRISON	SAIL	SLIGHT	SURE
LEFT	MISTER	ONCE	PROBABLY	SAME	SLIP	SURPRISE
LEG	MOMENT	ONE	PROBLEM	SATURDAY	SLOW	SWEET
LESS	MONDAY	ONLY	PROGRAMME	SAVE	SMALL	SWIM

SYSTEM	THEN	TOMORROW	TYPE	WASH	WHO	WOULD
TABLE	THERE	TONIGHT	UGLY	WASTE	WHOLE	WRITE
TAIL	THEY	TOO	UNCLE	WATCH	WHY	WRONG
TAKE	THICK	TOOTH	UNDER	WATER	WIDE	YARD
TALK	THING	TOP	UNDERNEATH	WAVE	WIFE	YEAR
TALL	THINK	TOTAL	UNDERSTAND	WAY	WILD	YELLOW
TAPE	THIRST	TOUCH	UNLESS	WE	WILL	YES
TASTE	THIRTEEN	TOWARD	UNTIL	WEAR	WIN	YESTERDAY
TAX	THIRTY	TOWN	UP	WEATHER	WIND	YET
TEA	THIS	TRACK	UPON	WEB	WINDOW	YOU
TEACH	THOUGH	TRAIN	USE	WED	WINE	YOUNG
TEAM	THOUSAND	TRAVEL	USUAL	WEDNESDAY	WINTER	ZERO
TEAR	THREE	TREAT	VAN	WEEK	WISH	
TELEPHONE	THROAT	TREE	VERY	WEIGHT	WITH	
TELEVISION	THROUGH	TRIP	VIDEO	WELL	WITHIN	
TELL	THROW	TROUBLE	VIEW	WEST	WITHOUT	
TEN	THURSDAY	TRUE	VISIT	WET	WOMAN	
TEND	TIE	TRUST	VOICE	WHAT	WONDER	
TERM	TIGHT	TRUTH	WAIT	WHEEL	WOOD	
TERRIBLE	TILL	TRY	WAKE	WHEN	WORD	
TEST	TIME	TUESDAY	WALK	WHERE	WORK	
THAN	TIRE	TURN	WALL	WHETHER	WORLD	
THANK	TO	TWELVE	WANT	WHICH	WORRY	
THAT	TODAY	TWENTY	WAR	WHILE	WORSE	
THE	TOGETHER	TWO	WARM	WHITE	WORTH	

Appendix D Good Stuff Gold D Headwords base list 1

A	AWARE	BREAKFAST	COME	DOUBLE	FATHER	GLAD
ABLE	AWAY	BREATH	COMFORT	DOUBT	FAVOURITE	GLANCE
ABOUT	AWFUL	BRIGHT	COMPANY	DOWN	FEAR	GLASS
ABOVE	BABY	BRING	COMPLETE	DRAW	FEED	GO
ABSOLUTE	BACK	BROTHER	COMPUTER	DREAM	FEEL	GOD
ACCEPT	BAD	BROWN	CONCERN	DRESS	FELLOW	GOLD
ACROSS	BAG	BUILD	CONSIDER	DRINK	FEW	GOOD
ACT	BALL	BURN	CONTINUE	DRIVE	FIELD	GOODBYE
ACTUAL	BANK	BUS	CONTROL	DROP	FIGHT	GOVERN
ADD	BAR	BUSH	CONVERSATI	DRUG	FIGURE	GRANDFATH
ADDRESS	BASE	BUSINESS	ON	DRY	FILL	ER
ADMIT	BASIC	BUSY	COOK	DURING	FILM	GRASS
ADVERTISE	BATH	BUT	COOL	EACH	FINAL	GREAT
AFFORD	BE	BUY	CORNER	EAR	FIND	GREEN
AFRAID	BEACH	BY	COST	EARLY	FINE	GREY
AFTER	BEAR	CAKE	COULD	EARTH	FINGER	GROUND
AFTERNOON	BEAT	CALL	COUNT	EAST	FINISH	GROUP
AGAIN	BEAUTY	CAMP	COUNTRY	EASY	FIRE	GROW
AGAINST	BECAUSE	CAN	COUPLE	EAT	FIRST	GUESS
AGE	BECOME	CAR	COURSE	EDGE	FISH	GUN
AGO	BED	CARD	COURT	EDUCATE	FIT	GUY
AGREE	BEFORE	CARE	COVER	EGG	FIVE	HAIR
AHEAD	BEGIN	CARRY	CRAZY	EIGHT	FIX	HALF
AIR	BEHIND	CASE	CRIME	EITHER	FLAT	HALL
ALL	BELIEVE	CAT	CROSS	ELEVEN	FLOOR	HAND
ALLOW	BELOW	CATCH	CRY	ELSE	FLOWER	HANDLE
ALMOST	BENEATH	CAUSE	CUP	EMPLOY	FLY	HANG
ALONE	BESIDE	CENTRE	CUT	EMPTY	FOLLOW	HAPPEN
ALONG	BEST	CERTAIN	DAD	END	FOOD	HAPPY
ALREADY	BET	CHAIR	DANCE	ENGINE	FOOT	HARD
ALRIGHT	BETTER	CHANCE	DANGER	ENJOY	FOOTBALL	HARDLY
ALSO	BETWEEN	CHANGE	DARK	ENOUGH	FOR	HAT
ALTHOUGH	BEYOND	CHARGE	DARLING	ENTER	FORCE	HATE
ALWAYS	BIG	CHEAP	DATE	ESPECIALLY	FOREST	HAVE
AMAZE	BILL	CHECK	DAUGHTER	EVEN	FORGET	HE
AMONG	BILLION	CHICKEN	DAY	EVENING	FORG	HEAD
AMOUNT	BIRD	CHILD	DEAD	EVER	FORTUNATE	HEALTH
AND	BIRTH	CHIP	DEAL	EVERY	FORWARD	HEAR
ANGRY	BIT	CHOICE	DEAR	EXACT	FOUR	HEART
ANIMAL	BLACK	CHOOSE	DEATH	EXCEPT	FREE	HEAT
ANOTHER	BLOOD	CHRISTMAS	DECIDE	EXCITE	FREEZE	HEAVY
ANSWER	BLOW	CHURCH	DEEP	EXCUSE	FRESH	HELL
ANY	BLUE	CITY	DEFINITE	EXPECT	FRIDAY	HELLO
APART	BOARD	CLASS	DEGREE	EXPENSIVE	FRIEND	HELP
APPARENT	BOAT	CLEAN	DELICIOUS	EXPERIENCE	FRIGHT	HERE
APPEAR	BODY	CLEAR	DEPEND	EXPLAIN	FROM	HIDE
AREA	BONE	CLIMB	DIE	EXPRESS	FRONT	HIGH
ARM	BOOK	CLOCK	DIFFERENCE	EXTRA	FULL	HILL
AROUND	BORING	CLOSE	DIFFERENT	EYE	FUN	HISTORY
ARRANGE	BORN	CLOSED	DIFFICULT	FACE	FURTHER	HIT
ARRIVE	BOTH	CLOTHES	DIG	FACT	GAME	HOLD
ART	BOTHER	CLUB	DINNER	FAIR	GARDEN	HOLE
AS	BOTTLE	COAT	DIRTY	FALL	GAS	HOLIDAY
ASHAMED	BOTTOM	COFFEE	DISCOVER	FAMILY	GENERAL	HOME
ASK	BOX	COLD	DO	FAR	GENTLE	HONEST
AT	BOY	COLLECT	DOCTOR	FARM	GET	HONOUR
AUNT	BREAD	COLLEGE	DOG	FAST	GIRL	HOPE
AUTUMN	BREAK	COLOUR	DOOR	FAT	GIVE	HORRIBLE

HORSE	LEAD	MIND	OIL	PRESS	ROUND	SITUATION
HOSPITAL	LEARN	MINUTE	OK	PRETTY	RUBBISH	SIX
HOT	LEAST	MISS	OLD	PRICE	RULE	SIZE
HOUR	LEAVE	MISTAKE	ON	PRINCE	RUN	SKIN
HOUSE	LEFT	MISTER	ONCE	PRISON	SAD	SKY
HOW	LEG	MOMENT	ONE	PROBABLY	SAFE	SLEEP
HOWEVER	LESS	MONDAY	ONLY	PROBLEM	SAIL	SLIGHT
HUGE	LET	MONEY	OPEN	PROGRAMM	SAME	SLIP
HUMAN	LETTER	MONTH	OR	E	SATURDAY	SLOW
HUNDRED	LEVEL	MORE	ORANGE	PROMISE	SAVE	SMALL
HUNGER	LIE	MORNING	ORDER	PROPER	SAY	SMELL
HUNT	LIFE	MOST	OTHER	PROTECT	SCARE	SMILE
HURRY	LIFT	MOTHER	OUGHT	PUBLIC	SCHOOL	SMOKE
HURT	LIGHT	MOUNTAIN	OUT	PULL	SCIENCE	SNOW
HUSBAND	LIKE	MOUTH	OVER	PUSH	SEA	SO
I	LINE	MOVE	OWN	PUT	SEAT	SOFT
ICE	LIP	MOVIE	OWNED	QUARTER	SECOND	SOME
IDEA	LIST	MRS	PACK	QUEEN	SECURE	SON
IF	LISTEN	MUCH	PAGE	QUESTION	SEE	SONG
IMAGINE	LITTLE	MUM	PAIN	QUICK	SEEM	SOON
IMPORTANT	LIVE	MUSIC	PAINT	QUIET	SELF	SORRY
IN	LOAD	MUST	PAIR	QUITE	SELL	SORT
INDEED	LOCAL	NAME	PAPER	RABBIT	SEND	SOUND
INFORM	LOCK	NATION	PARDON	RACE	SENSE	SOUTH
INSIDE	LONG	NATURE	PARENT	RADIO	SERIOUS	SPACE
INSTEAD	LOOK	NAUGHTY	PARK	RAIN	SERVE	SPEAK
INSURE	LORD	NEAR	PART	RAISE	SERVICE	SPECIAL
INTEREST	LOSE	NEAT	PARTICULAR	RATE	SET	SPEND
INTERNET	LOT	NECESSARY	PARTY	RATHER	SETTLE	SPORT
INTO	LOUD	NECK	PASS	REACH	SEVEN	SPOT
INVOLVE	LOVE	NEED	PAST	READ	SEVERAL	SPRING
ISLAND	LOVELY	NEIGHBOUR	PAY	READY	SEX	SQUARE
ISSUE	LOW	NEVER	PENNY	REAL	SHAKE	STAGE
IT	LUCK	NEW	PEOPLE	REALISE	SHALL	STAND
JOB	LUNCH	NEWS	PERFECT	REALLY	SHAPE	STAR
JOIN	MACHINE	NEXT	PERHAPS	REASON	SHARE	STARE
JOKE	MAD	NICE	PERSON	RECENT	SHE	START
JUDGE	MAIN	NIGHT	PHOTOGRAP	RECORD	SHIP	STATE
JUMP	MAJOR	NINE	H	RED	SHIRT	STATION
JUST	MAKE	NO	PICK	RELATE	SHOE	STAY
KEEP	MAN	NOBODY	PICTURE	REMEMBER	SHOOT	STEAL
KEY	MANAGE	NOISE	PIECE	RENT	SHOP	STEP
KICK	MANY	NONE	PLACE	REPLY	SHORT	STICK
KID	MARK	NORMAL	PLAN	REPORT	SHOULD	STILL
KILL	MARKET	NORTH	PLANT	RESPONSIBL	SHOULDER	STONE
KIND	MARRY	NOSE	PLAY	E	SHOUT	STOP
KING	MASTER	NOT	PLEASE	REST	SHOW	STORE
KISS	MATTER	NOTE	PLENTY	RETURN	SHUT	STORY
KITCHEN	MAY	NOTHING	PLUS	RICH	SHY	STRAIGHT
KNOCK	MAYBE	NOTICE	POINT	RID	SICK	STRANGE
KNOW	MEAL	NOW	POLICE	RIDE	SIDE	STREET
LADY	MEAN	NUMBER	POOR	RIGHT	SIGHT	STRIKE
LAKE	MEET	NURSE	POP	RIGHTS	SIGN	STRONG
LAND	MEMBER	OBVIOUS	POSITION	RING	SILLY	STUDENT
LARGE	MENTION	ODD	POSSIBLE	RISE	SIMPLE	STUDY
LAST	MESS	OF	POST	RIVER	SINCE	STUFF
LATE	MIDDLE	OFF	POT	ROAD	SING	STUPID
LAUGH	MIGHT	OFFER	POUND	ROCK	SINGLE	SUBJECT
LAW	MILE	OFFICE	POWER	ROLL	SIR	SUCH
LAY	MILK	OFFICER	PREPARE	ROOM	SISTER	SUDDEN
LAZY	MILLION	OFTEN	PRESENT	ROUGH	SIT	SUGGEST

SUIT	TONIGHT	WED
SUMMER	TOO	WEDNESDAY
SUN	TOOTH	WEEK
SUNDAY	TOP	WEIGHT
SUPPORT	TOTAL	WELL
SUPPOSE	TOUCH	WEST
SURE	TOWARD	WET
SURPRISE	TOWN	WHAT
SWEET	TRACK	WHEEL
SWIM	TRAIN	WHEN
SYSTEM	TRAVEL	WHERE
TABLE	TREAT	WHETHER
TAIL	TREE	WHICH
TAKE	TRIP	WHILE
TALK	TROUBLE	WHITE
TALL	TRUE	WHO
TAPE	TRUST	WHOLE
TASTE	TRUTH	WHY
TAX	TRY	WIDE
TEA	TUESDAY	WIFE
TEACH	TURN	WILD
TEAM	TWELVE	WILL
TEAR	TWENTY	WIN
TELEPHONE	TWO	WIND
TELEVISION	TYPE	WINDOW
TELL	UGLY	WINE
TEN	UNCLE	WINTER
TEND	UNDER	WISH
TERM	UNDERNEAT	WITH
TERRIBLE	H	WITHIN
TEST	UNDERSTAN	WITHOUT
THAN	D	WOMAN
THANK	UNLESS	WONDER
THAT	UNTIL	WOOD
THE	UP	WORD
THEN	UPON	WORK
THERE	USE	WORLD
THEY	USUAL	WORRY
THICK	VAN	WORSE
THING	VERY	WORTH
THINK	VIDEO	WOULD
THIRST	VIEW	WRITE
THIRTEEN	VISIT	WRONG
THIRTY	VOICE	YARD
THIS	WAIT	YEAR
THOUGH	WAKE	YELLOW
THOUSAND	WALK	YES
THREE	WALL	YESTERDAY
THROAT	WANT	YET
THROUGH	WAR	YOU
THROW	WARM	YOUNG
THURSDAY	WASH	ZERO
TIE	WASTE	
TIGHT	WATCH	
TILL	WATER	
TIME	WAVE	
TIRE	WAY	
TO	WE	
TODAY	WEAR	
TOGETHER	WEATHER	
TOMORROW	WEB	

Appendix E Academic Word List by Coxhead (2000)

(Averil Coxhead's) High-Incidence Academic Word List (AWL) – Alphabetical Order

Words of highest frequency are followed by the number 1

abandon	8	bias	8	constitute	1	distinct	2	fee	6
abstract	6	bond	6	constrain	3	distort	9	file	7
academy	5	brief	6	construct	2	distribute	1	final	2
access	4	bulk	9	consult	5	diverse	6	finance	1
accommodate	9	capable	6	consume	2	document	3	finite	7
accompany	8	capacity	5	contact	5	domain	6	flexible	6
accumulate	8	category	2	contemporary	8	domestic	4	fluctuate	8
accurate	6	cease	9	context	1	dominate	3	focus	2
achieve	2	challenge	5	contract	1	draft	5	format	9
acknowledge	6	channel	7	contradict	8	drama	8	formula	1
acquire	2	chapter	2	contrary	7	duration	9	forthcoming	10
adapt	7	chart	8	contrast	4	dynamic	7	foundation	7
adequate	4	chemical	7	contribute	3	economy	1	found	9
adjacent	10	circumstance	3	controversy	9	edit	6	framework	3
adjust	5	cite	6	convene	3	element	2	function	1
administrate	2	civil	4	converse	9	eliminate	7	fund	3
adult	7	clarify	8	convert	7	emerge	4	fundamental	5
advocate	7	classic	7	convince	10	emphasis	3	furthermore	6
affect	2	clause	5	cooperate	6	empirical	7	gender	6
aggregate	6	code	4	coordinate	3	enable	5	generate	5
aid	7	coherent	9	core	3	encounter	10	generation	5
albeit	10	coincide	9	corporate	3	energy	5	globe	7
allocate	6	collapse	10	correspond	3	enforce	5	goal	4
alter	5	colleague	10	couple	7	enhance	6	grade	7
alternative	3	commence	9	create	1	enormous	10	grant	4
ambiguous	8	comment	3	credit	2	ensure	3	guarantee	7
amend	5	commission	2	criteria	3	entity	5	guideline	8
analogy	9	commit	4	crucial	8	environment	1	hence	4
analyse	1	commodity	8	culture	2	equate	2	hierarchy	7
annual	4	communicate	4	currency	8	equip	7	highlight	8
anticipate	9	community	2	cycle	4	equivalent	5	hypothesis	4
apparent	4	compatible	9	data	1	erode	9	identical	7
append	8	compensate	3	debate	4	error	4	identify	1
appreciate	8	compile	10	decade	7	establish	1	ideology	7
approach	1	complement	8	decline	5	estate	6	ignorance	6
appropriate	2	complex	2	deduce	3	estimate	1	illustrate	3
approximate	4	component	3	define	1	ethic	9	image	5
arbitrary	8	compound	5	definite	7	ethnic	4	immigrate	3
area	1	comprehensive	7	demonstrate	3	evaluate	2	impact	2
aspect	2	comprise	7	denote	8	eventual	8	implement	4
assemble	10	compute	2	deny	7	evident	1	implicate	4
assess	1	conceive	10	depress	10	evolve	5	implicit	8
assign	6	concentrate	4	derive	1	exceed	6	imply	3
assist	2	concept	1	design	2	exclude	3	impose	4
assume	1	conclude	2	despite	4	exhibit	8	incentive	6
assure	9	concurrent	9	detect	8	expand	5	incidence	6
attach	6	conduct	2	deviate	8	expert	6	incline	10
attain	9	confer	4	device	9	explicit	6	income	1
attitude	4	confine	9	devote	9	exploit	8	incorporate	6
attribute	4	confirm	7	differentiate	7	export	1	index	6
author	6	conflict	5	dimension	4	expose	5	indicate	1
authority	1	conform	8	diminish	9	external	5	individual	1
automate	8	consent	3	discrete	5	extract	7	induce	8
available	1	consequent	2	discriminate	6	facilitate	5	inevitable	8
aware	5	considerable	3	displace	8	factor	1	infer	7
behalf	9	consist	1	display	6	feature	2	infrastructure	8
benefit	1	constant	3	dispose	7	federal	6	inherent	9

Academic Word List – Alphabetical

inhibit	6	migrate	6	precede	6	reverse	7	tense	8
initial	3	military	9	precise	5	revise	8	terminate	8
initiate	6	minimal	9	predict	4	revolution	9	text	2
injure	2	minimise	8	predominant	8	rigid	9	theme	8
innovate	7	minimum	6	preliminary	9	role	1	theory	1
input	6	ministry	6	presume	6	route	9	thereby	8
insert	7	minor	3	previous	2	scenario	9	thesis	7
insight	9	mode	7	primary	2	schedule	8	topic	7
inspect	8	modify	5	prime	5	scheme	3	trace	6
instance	3	monitor	5	principal	4	scope	6	tradition	2
institute	2	motive	6	principle	1	section	1	transfer	2
instruct	6	mutual	9	prior	4	sector	1	transform	6
integral	9	negate	3	priority	7	secure	2	transit	5
integrate	4	network	5	proceed	1	seek	2	transmit	7
integrity	10	neutral	6	process	1	select	2	transport	6
intelligence	6	nevertheless	6	professional	4	sequence	3	trend	5
intense	8	nonetheless	10	prohibit	7	series	4	trigger	9
interact	3	norm	9	project	4	sex	3	ultimate	7
intermediate	9	normal	2	promote	4	shift	3	undergo	10
internal	4	notion	5	proportion	3	significant	1	underlie	6
interpret	1	notwithstanding	10	prospect	8	similar	1	undertake	4
interval	6	nuclear	8	protocol	9	simulate	7	uniform	8
intervene	7	objective	5	psychology	5	site	2	unify	9
intrinsic	10	obtain	2	publication	7	so-called	10	unique	7
invest	2	obvious	4	publish	3	sole	7	utilise	6
investigate	4	occupy	4	purchase	2	somewhat	7	valid	3
invoke	10	occur	1	pursue	5	source	1	vary	1
involve	1	odd	10	qualitative	9	specific	1	vehicle	8
isolate	7	offset	8	quote	7	specify	3	version	5
issue	1	ongoing	10	radical	8	sphere	9	via	8
item	2	option	4	random	8	stable	5	violate	9
job	4	orient	5	range	2	statistic	4	virtual	8
journal	2	outcome	3	ratio	5	status	4	visible	7
justify	3	output	4	rational	6	straightforward	10	vision	9
label	4	overall	4	react	3	strategy	2	visual	8
labour	1	overlap	9	recover	6	stress	4	volume	3
layer	3	overseas	6	refine	9	structure	1	voluntary	7
lecture	6	panel	10	regime	4	style	5	welfare	5
legal	1	paradigm	7	region	2	submit	7	whereas	5
legislate	1	paragraph	8	register	3	subordinate	9	whereby	10
levy	10	parallel	4	regulate	2	subsequent	4	widespread	8
liberal	5	parameter	4	reinforce	8	subsidy	6		
licence	5	participate	2	reject	5	substitute	5		
likewise	10	partner	3	relax	9	successor	7		
link	3	passive	9	release	7	sufficient	3		
locate	3	perceive	2	relevant	2	sum	4		
logic	5	percent	1	reluctance	10	summary	4		
maintain	2	period	1	rely	3	supplement	9		
major	1	persist	10	remove	3	survey	2		
manipulate	8	perspective	5	require	1	survive	7		
manual	9	phase	4	research	1	suspend	9		
margin	5	phenomenon	7	reside	2	sustain	5		
mature	9	philosophy	3	resolve	4	symbol	5		
maximise	3	physical	3	resource	2	tape	6		
mechanism	4	plus	8	respond	1	target	5		
media	7	policy	1	restore	8	task	3		
mediate	9	portion	9	restrain	9	team	9		
medical	5	pose	10	restrict	2	technical	3		
medium	9	positive	2	retain	4	technique	3		
mental	5	potential	2	reveal	6	technology	3		
method	1	practitioner	8	revenue	5	temporary	9		