Vocabulary in textbooks for young learners

Marie Nordlund

An extensive vocabulary is necessary for successful communication in a foreign language. According to the Swedish National Agency for Education, the teaching of English in Sweden is often based on one textbook and on that textbook only. The purpose of this study is therefore to describe what vocabulary is presented in one teaching material aimed at school years 4–6 (age group 10–12 years) in Sweden. More precisely, the following questions have been considered: (i) what are the most frequent words in the textbooks? and (ii) to what extent do words recur in the textbooks? To answer these questions, a corpus of the texts in the three books was compiled. The analyses show that although the texts appear to have been chosen to appeal to the age groups aimed for, there does not seem to be any consciously considered rationale behind the choice of what vocabulary to include, many words occur only once and the necessary repetition of words is thus lacking.

Keywords: English, frequency, repetition, vocabulary acquisition, word frequencies

Marie Nordlund, senior lecturer, Department of Arts, Communication and Education, Luleå University of Technology
Marie.Nordlund@ltu.se

Introduction

Knowledge of vocabulary is imperative to communication in a foreign language. Research on vocabulary learning has demonstrated, among other things, the importance of vocabulary in making oneself understood, the size of vocabulary needed and how vocabulary is best taught and learned (see, e.g., Milton, 2009; Nation, 1990, 2001, 2008; Read, 2000; Schmitt, 2000). In many foreign-language classrooms, the textbook is often the main, and sometimes the only, teaching material used; despite this centrality, few studies have focussed on the correlation between the structure of teaching materials and vocabulary content (see, however, Meara & Suárez García, 2010; Rixon, 1999). In a study of seven international textbooks in English for beginners, Rixon (1999) found that the books were lacking in several respects, one of them being inadequate reuse of words. Recurrent meetings with words enable the automatization of word knowledge and recognition (Tyler, 2012, p. 53) and the development of a sight vocabulary (Coady, 1997, p. 235; Laufer, 1997, p. 23)—the direct and unconscious decoding of the meaning of words read is as important as learning to directly decode words heard. Hence, vocabulary learning should be in focus in the beginners' classroom (Coady, 1997, p. 229). To build up a vocabulary in a new language is not an easy task, and learners need all the help they can get—from teachers as well as from the structure and content of the teaching material used. Since Rixon's (1999) results indicate that textbooks often conform badly to advances within our understanding of the role of vocabulary acquisition in second language learning, they beg the question, Do other textbooks in English have the same flaws? In the Swedish context, studies have focussed on the high school level (Ljung, 1990, 1991), but so far, no comprehensive study has focussed on how vocabulary is presented in textbooks for young learners. The results of a survey carried out by the Swedish National Agency for Education on the use of teaching materials (Skolverket, 2006) indicate that it is far from evident that the contents of textbooks used actually correspond to the learning objectives set out in the national curriculum. There is thus more than one reason to study this area further.

Learning vocabulary

The relation between the development of a sight vocabulary and the automatic decoding of words may seem self-evident, but determining what constitutes 'knowing' a word is not as straightforward. Nation (1990, p. 31), for instance, lists eight different aspects of knowing a word:

- its meaning(s)
- its written form
- its spoken form

- its grammatical behaviour
- its collocations
- its register
- its associations
- its frequency

Clearly, not all words in one's vocabulary are 'known' to the same extent; some are recognised and understood in context only (receptive vocabulary), while others are readily available for productive use. However, there are also degrees in between, as Melka (1997) puts it:

> Word recognition is possible even when the word is stored incompletely or when word production is still impossible. Similarly, even if the word has not reached a totally productive stage, many of its features may be productively known by the subject. (p. 88)

Furthermore, building up a vocabulary in a new language, Nation's (1990) list indicates, is not just a matter of adding new words to those already known (i.e., acquiring breadth of vocabulary); it also requires expanding the knowledge of words already acquired and stored in the mental lexicon (i.e., acquiring depth of vocabulary). This expansion includes, but is not restricted to, further knowledge of collocations, nuances of meaning, synonyms and antonyms and can be achieved when words are used and encountered in many different contexts. Within cognitive linguistics the construction of the mental lexicon is described as a network (Aitchison, 2012) where words are joined together by more or less strong relations, which depend on the learner's previous experience of and encounters with the words. As Schoonen and Verhallen (2008) explain, "learning new words is more than the acquisition of isolated lexical units: new words are embedded in a lexical network which means that all kinds of connections with related words have to be established" (p. 213). Transferring new words from short-time working memory to long-time memory is therefore facilitated if the information can be "hooked" onto information that is already stored:

> Vocabulary is best learned when the meaning of the word(s) is illustrated, for example by a picture, an action, or a real object. The children should then meet and use the word(s) in relevant contexts, in order to 'fix' them in their minds. This helps establish their relationship to other words, so that a vocabulary network is built up. (Phillips, 1993, p. 68)

In this vocabulary network, then, breadth is signalled by the number of nodes present, whereas the number of links between different nodes corresponds to depth.

Frequency—that is, how often learners meet new words and how often words recur in different contexts—is central in this process. Ellis (2002) maintains that "the recognition and production of words is a function of their frequency of occurrence in the language" and that "each time we process [a word] there is a reduction in processing time" (p. 152). Even though most researchers agree that learners need to meet new words several times before they can be integrated into the mental lexicon, the minimum number of times necessary vary, but ten times is a number that is mentioned more often than others (Matsuoka, 2012, p. 158). A high number of words occurring only once in a textbook must then be considered a waste of time: "If recycling is neglected, many partially known words will be forgotten, wasting all the effort already put into learning them" (Schmitt, 2000, p. 137). Findings from a study of word frequencies in textbooks (Kachroo, 1962, as cited in Nation, 1990, p. 43) point in the same direction: the majority of the informants had no knowledge of more than half of those words that occurred only once or twice in the books, whereas words occurring at least seven times were understood by most informants. Repetition is, thus, a necessary ingredient in successful vocabulary learning. Results from a comprehensive American study on memory indicate that a time interval of 1–4–10 (repetition one, four and ten days after the first learning opportunity), for example, is optimal and results in better retrieval than, for instance, 5–5–5 (Klingberg, 2011, p. 65; see also Nation, 2003, p. 14). Generally speaking, the more frequently a learner uses or is exposed to a word, the easier it will be to remember and use appropriately. Ideally, words should also appear in different contexts, as this promotes the formation of networks and the development of vocabulary depth.

Core vocabulary

When children with English as their first language start school around the age of five, they generally have a productive vocabulary of approximately 3,200 words (Biemiller, 2003, p. 327)—a vocabulary that grows with around 500 words each year (Nation & Waring, 1997, p. 7; Schmitt, 2000, p. 3). Since a second language learner seldom has access to daily input of authentic language, the same development of vocabulary cannot be expected. The question is then *what* words to learn.

There have been attempts to define a core vocabulary for English. One is Ogden's *Basic English,* the development of which began in the 1930s. *Basic English* contains 850 words and, it is claimed (Ogden, 1944, p. 10), is sufficient for everyday conversations. The problem with *Basic English* is that it is not as basic as its name would suggest. The words included are highly

polysemous and estimations indicate that, added together, its 850 words actually have 12,425 different meanings that the learner has to acquire (Carter & McCarthy, 1988, p. 3).

A later attempt is West's (1953) *A general service list of English words* (*GSL*), which comprises 2,000 words. The list was compiled from corpus data and contains information about the different meanings of the words as well as how frequent these meanings are. Just as *Basic English*, the words in the *GSL* are said to make participation in everyday conversations possible;[1] additionally, they also enable the learner to understand approximately 80% of a text on a general topic (Carter & McCarthy, 1988, p. 7). In more specialized texts, much of the semantic burden lies on low-frequency words. For a learner to able to guess the meaning of unfamiliar words with the help of context, that kind of text is more likely to require a coverage of 95–98% (Laufer, 1997, p. 127; Nation & Waring, 1997, p. 10; Schmitt, 2008, p. 330), which would call for a vocabulary containing the 3,000 most frequent words (Nation & Waring, 1997, p. 11).

Within the European context, work on establishing lexical thresholds for different usage levels is longstanding and has resulted in the development of the *Common European framework of reference for languages* (*CEFR*; Council of Europe, 2001). The original work preceding the framework (e.g., van Ek & Trim, 1998a, b) contained word lists indicating a vocabulary of approximately 1,000 words for basic users (A2 level) and 2,000 words for independent users (B1 level). Primarily, beginning learners should acquire highly frequent but neutral words since these can be used in more than one context (Schmitt, 2000, p. 37). A realistic goal for basic language use could thus be the 2,000–3,000 most frequent words (ibid., p. 84; see also Stæhr, 2008). An aspect worth mentioning in this context is that there is not much discussion in the literature about *when* this goal should be achieved, even though Cameron (2001, p. 75) maintains that acquiring 500 new words per year is a reasonable goal for young language learners. From an educational point of view, it is relevant to consider how much time learners have at their disposal to acquire a vocabulary of this size since this will have implications for the creation of national curricula and the planning of teaching in both a short-term and a long-term perspective as well as for progression in the design of teaching materials. In the Swedish national curriculum for the compulsory school (Skolverket, 2011), no vocabulary size is mentioned among the goals to achieve, but "it is reasonable to expect that students should reach at least the B1 level in English at the end of compulsory school [school year 9] in order to obtain a pass" (Skolverket, 2012, p. 28, my translation). In theory, then, after 480 hours of English teaching (Skolverket, 2011), students

[1] Read (2004, p. 150) is, however, of the opinion that a more realistic figure for everyday communication is a vocabulary of 3,000 words.

should have knowledge of the most frequent words in English. However, as shown by Stæhr (2008) in his study of the vocabulary of Danish 15–16-year-olds, this might not conform to reality.

Textbooks in language teaching

Views on whether or not textbooks should be used in the classroom vary, but many student teachers in Sweden today have a bias against working with textbooks and almost regard it as inappropriate (personal observations). They consider textbook contents to be out of date and dialogues to be contrived; textbooks are static rather than dynamic, and they are not appealing to learners. Student teachers are not the only ones to have these opinions, researchers agree. In particular, it is the structure of dialogues and the language used in them that have been criticised for being unnatural (Cameron, 2001, p. 69; Kirk & Carter, 2010, p. 37), but also more general criticism has been voiced: "There is some concern … that ELT course books contradict rather than reflect contemporary developments in applied linguistics" (Matsuoka & Hirsh, 2010, p. 59).

Despite the critique, teachers, and in particular new teachers without the experience that comes from many years of teaching, often rely heavily on textbooks in their planning and teaching (Skolverket, 2006; see also Matsuoka & Hirsh, 2010, p. 59). Quite often, it is the textbook—and, consequently, the textbook writer(s)—that decides what should be taught in the classroom and is seen as the authority as regards what counts as a well-planned class (Abello-Contesse & López-Jiménez, 2010, p. 96) because teachers do not have the time and do not always have the competence needed to analyse teaching materials and evaluate their suitability (Council of Europe, 2001, p. 141). Textbooks are products that "greatly influence the learning/teaching process" (ibid.), and it is important that the methods used in language teaching are those considered most effective (ibid., p. 142). A conclusion to be drawn from this is that it is equally important that the structure and contents of textbooks are designed in accordance with research within the field of second language acquisition. The question is whether this is always the case.

The present study

The purpose of the present study is to describe what vocabulary is presented in textbooks used in Swedish intermediate schools. The following questions will be addressed:

- What words are most frequent in the textbooks?
- To what extent do words recur in the textbooks?

Material

The material analysed comprises three textbooks aimed at school years 4–6 in Sweden (pupils aged 10–12 years) from the series *Good Stuff*, published by Almqvist & Wiksell/Liber. Together with two other series, it is one of the most commonly used textbooks in Sweden. The books analysed—*Good Stuff 4* (school year 4), *Good Stuff 5* (school year 5) and *Good Stuff A*[2] (school year 6)—will henceforth be referred to as *GS4*, *GS5* and *GSA*. The texts in the books are arranged in thematic units (see Table 1). Both *GS4* and *GS5* contain eight units, each covering more than one theme (for example, *Sports/School/Friends*) and containing a varied number of different texts. In *GSA*, each of its thirteen units is devoted to one theme only (for instance, *Sports*).

Table 1. Thematic units in *GS4, GS5* and *GSA*.

Good Stuff 4	***Good Stuff 5***	***Good Stuff A***
Numbers / Home	Home / City	Animals
Colours / Family	Shops / Clothes	Sports
House / Time / Pets	Nature / Crime	Food
Food / The table	In the country / Hobbies	A folk tale
Body / Clothes	Friends / Fun	Pirates
Hobbies / Weather	School / Stories	Fighters
Sports / School / Friends	Film / Music	Countries
Fun / Holidays	Australia	Clothes
		Magic
		Explorers
		Dragons
		Horror
		December

Analysis

To analyse and compare the vocabulary presented in the three books, a corpus of all the texts was compiled. The corpus consists of three sub-corpora, one for each book. The size of the corpus in number of tokens (running words) and types (individual words) is shown in Table 2.

[2] *Good Stuff A* is followed by *Good Stuff B–D* aimed at school years 7–9.

Table 2. Number of tokens and types in the corpus.

Sub-corpus	**Number of tokens**	**Number of types**
GS4	9,500	1,293
GS5	13,916	1,658
GSA	17,617	2,126
Total	41,033	3,186

To facilitate analyses and comparison, the corpus has been tagged with information about word class, number, tense, etc. Tagging has been conducted according to the principles used for the construction of the British National Corpus (BNC). To search the corpus the concordancing software tool MonoConc Pro has been used.

MonoConc Pro counts singular and plural forms of nouns, comparative and superlative forms of adjectives, and verb inflections as different types. This has been adjusted manually in Table 2 above. In other words, in this study, for example, *cat-cats*, *old-older-oldest* and *ask-asks-asking-asked* are treated as one type each. An additional adjustment has also been made for the category of nouns. According to the tag key used, names for days, months, holidays and countries are tagged as proper nouns and, as such, would not show in the results generated by the software program for a search of nouns. In this study, they have been treated as common nouns and have been added to the data analysed.

Even though the whole corpus has been tagged, the present study is limited to an analysis of the open word classes adjectives, nouns and verbs. These constitute the three largest categories of content words as well as being the most important word classes. They also account for a major part of the material analysed as a whole. The books vary slightly in this respect: in *GS4* the three word classes comprise 46% of the total number of tokens, and in *GS5* and *GSA* the corresponding figures are 44% and 45%, respectively. Expressed in types, the dominance of these three word classes is even more pronounced, with figures ranging between 72 and 78%. The data, after manual adjustments, are presented in Table 3.

Table 3. Number of types per word class and % of total number of types per book.

Word class	***GS4***	***GS5***	***GSA***
Adjectives	155 / 12.0	219 / 13.2	288 / 13.5
Nouns	575 / 44.5	778 / 46.9	932 / 43.8
Verbs	201 / 15.5	292 / 17.6	372 / 17.5
Total	931 / 72.0	1,289 / 77.7	1,592 / 74.8

Findings and discussion

Word frequencies

The most frequent words in each word class are presented in Tables 4–6 below. The cut-off point for high frequency differs between word classes and is, to a certain extent, arbitrary. The decision is based on where it is possible to discern a clear difference in frequency between types—that is, where there is a sharp decline in frequency and where remaining words are spread out relatively evenly with tokens on every frequency level. A tendency for all three books is that the lower the frequency, the more types. In other words, there are proportionally more low- and mid-frequency word types than high-frequency ones. This applies to all three word classes.

Table 4. Adjectives with a frequency of at least 20 tokens.

GS4	***GS5***	***GSA***
good (32)	big (37)	good (65)
little, red (28)	good (32)	big (58)
big (20)	little (28)	old (28)
	old (26)	small (25)
	happy, new (23)	bad (22)

The exact number of tokens is given in brackets. Words in the same cell have the same number of tokens.

High-frequency adjectives account for approximately one-fifth of the total number of adjective tokens in the books. As can be seen from Table 4, there is little variation among the most common items. All the frequent adjectives are words which can be used in many different contexts, and they are also among the 1,000 most frequent words in larger general corpora such as the BNC and the Corpus of Contemporary American English. The most pregnant difference is seen in *GSA*, where *good* and *big* stand out with 65 and 58 tokens, respectively. The analysis also shows that other, semantically related, adjectives exist in the data (e.g., *big: great*, *large, enormous*) but not at all to the same extent. Both *good* and *big* have a wide range of usages and do not generally affect the interpretation of the nouns they modify either positively or negatively. Thus, it might prove difficult (or be less appropriate) to substitute them for other adjectives which might give more emphasis or add connotations to the utterance. However, since *good* and *big* recur in similar constructions in all the books, their persistent use does not add much to vocabulary depth. Therefore, in general, a more focussed use of other adjectives, giving the learners access to more nu-

ances and ways of expressing themselves, would be desirable for the development of both breadth and depth of vocabulary.

Another difference between the books is the presence of colour terms in *GS4*. Only *red* entered the top list, but added together, colours account for almost one-fifth of the total number of adjective tokens. The corresponding figures for colour terms in *GS5* and *GSA* are merely four and seven per cent, respectively. In English textbooks written for young learners in Sweden, colour terms are generally of high frequency. A quick and by no means scientific search on Google (5 April 2013) for the terms 'young learners' and 'colours' also resulted in more than 1.5 million hits, many of them containing exercises to use in the classroom. *GS4* could, thus, be said to follow what seems to be a set pattern: at an early stage, young learners should learn how to describe the colour of an object.

If there is little variation among high-frequency adjectives, the opposite holds for nouns (as shown in Table 5).

As is clear from Table 5, the number of nouns with a frequency of at least 15 tokens is quite similar in *GS4* and *GS5*: 14 and 15 types, respectively. There is a marked increase in *GSA*, however, with 40 different types. The variation between the books as regards high-frequency nouns is considerable: only six types are common to all the books (in alphabetical order: *dad, day, friend, mum/mom, school* and *time*).

A closer analysis reveals that the themes, and hence the texts, chosen for inclusion in the books have a major impact on what words are highest in frequency. Words like *dad* and *mum/mom* are a natural part of texts about children and their families, but they are also frequent because they are explicitly named as participants in dialogues. The high number of tokens for *man* (*GS5* and *GSA*)*, boy* (*GSA*) and *assistant* (*GSA*) can also be explained in the same way. This will sometimes skew the vocabulary count. In the 100 million word corpus BNC, *assistant* has a total frequency of 27 tokens, almost the same as in the 41,000 word corpus compiled for the present study. When the concordances for the *GS4* words *bed, doctor* and *monkey* are analysed in more detail, it is revealed that most of the tokens stem from the song *Five little monkeys*. If these tokens were not counted, none of the words would reach the 15-token level. Indeed, *monkeys* would not have any tokens at all. In similar ways, it is possible to explain the high frequency of other nouns in *GS4* as well as in *GS5* and *GSA*.

Table 5. Nouns with a frequency of at least 15 tokens.

GS4	***GS5***	***GSA***
dad (48)	mum/mom (44)	people (63)
mum/mom (38)	people (43)	man (57)
day (36)	day (40)	dragon (42)
dog (29)	dad (38)	day (41)
friend (25)	man (32)	food, snake (36)
bed, name (23)	friend (30)	friend (35)
time (21)	time, wolf (26)	mum (33)
doctor (20)	honey (24)	time (31)
monkey (19)	film (19)	witch (29)
summer, week (18)	Australia, hand, zoo (17)	dad, water, year (28)
school (16)	life, school (15)	animal (26)
fish (15)		pirate, story (25)
		woman (24)
		assistant (23)
		boy, home (22)
		human (21)
		clothes, thing (20)
		cane toad, child, school, world (19)
		country, door, house, name (18)
		fire, narrator, soldier, way (17)
		hand, ship (16)
		Christmas, head, night (15)

The exact number of tokens is given in brackets. Words in the same cell have the same number of tokens.

In *GSA*, many of the high-frequency nouns stand out and do not have an equal position in general corpora. Many of them would not reach a frequency of 15 tokens had it not been for the themes chosen. This is the case for words such as *dragon, snake, pirate, witch, cane toad, narrator* and *soldier*. In the material accessed for the study, it is not stated what rationale is behind the choice of themes. The decidedly skewed results, favouring words such as *cane toad* which have little use in common contexts, lead to the conclusion that vocabulary was not incorporated into the planning of the text-

books. Sadly, this is not uncommon in language-learning textbooks. Nation (1993) and, more recently, Meara and Suárez García (2010) and Burton (2012) point out that the inclusion of vocabulary seems to be random and, to a large extent, depends on the individual preferences of textbook writers. Even though the themes chosen for the *Good Stuff*-series may be considered appealing to learners in the ages aimed for and, consequently, favourable for vocabulary learning, what is needed are textbooks which to a much larger extent than now make use of the insights gained through corpus linguistic studies as regards frequency and usefulness of words (Matsuoka, 2012). Texts that are written exclusively for textbooks—most texts in *Good Stuff* are produced by the textbook writers themselves—could easily be checked against frequency lists to avoid unnecessary use of too many low-frequency words.

Verbs in the three books generally have much higher frequencies than the adjectives or nouns, and the cut-off point for high frequency has been set to at least 40 tokens (as illustrated in Table 6). Only main verbs have been considered. Modal auxiliaries have been left out of the analysis, and so have *be, do* and *have* in their auxiliary functions as well as instances of *go* used in the future-time construction *be going to*.

Table 6. Main verbs with a frequency of at least 40 tokens.

GS4	***GS5***	***GSA***
be (438)	be (609)	be (811)
like (83)	say (70)	have (108)
go (68)	get (64)	go, say (70)
want (55)	go (63)	look (68)
come (54)	see (60)	know (58)
get (52)	look (55)	want (56)
say (47)	have, know (50)	come, think (55)
love (41)	do (47)	see (50)
have (40)	want (46)	take (48)
	take (41)	eat, get (45)
		do (40)

The exact number of tokens is given within brackets.
Words in the same cell have the same number of tokens.

The number of verb types reaching the 40-token level is relatively similar when the three books are compared: nine, eleven and fourteen types, respectively. As can be clearly seen, *be* is the most frequent verb in all three books and represents a large number of tokens: 21% in *GS4*, 20% in *GS5* and 22%

in *GSA*. Usages of *be* are mainly focused on descriptions (*be a/an N* and *be Adj*) and presentations (*I am X, This is X*), knowledge that must be considered relevant for young learners and is commonly included in beginners' textbooks.

Many of the types presented in Table 6 are also common to all the books with no more than four types exclusive for one book only: *like* and *love* in *GS4* and *think* and *eat* in *GSA*. Further, the number of tokens for many of the verbs shared is similar regardless of the book. *Be* is an exception, but apart from that, *have* is really the only verb falling out of the pattern with its more than doubled frequency in *GSA* as compared to *GS5*. This increase has several explanations: (i) The construction *have to do something* is more than three times as common in *GSA*. (ii) Third person singular *has* increases markedly in *GSA*. (iii) The past tense form *had* is more frequent in *GSA*. (iv) The largest difference lies, however, in the increased use of the construction *have a/an N*, which can be linked to the higher number of noun types in *GSA*. The supply of objects or attributes that someone can own or be subscribed is simply larger in *GSA*.

The table of contents in *GS4* and *GS5* specify what words and expressions students are supposed to learn in each chapter, and suggestions for vocabulary homework in the teacher's books often consist of complete phrases. Thus, the intentions behind *Good Stuff* seem to be directed towards functional language, in line with objectives in the *CEFR* and within communicative language teaching, where the notion of what language users can do with language is highlighted. More than in the other books, the content in *GS4* is related to the domains laid down in the *CEFR* with many dialogues in which functional language is encountered, but as the series advances, so does the amount of narrative text at the expense of functionality.

Recurrence of words

A higher frequency for a word means that learners are more exposed to it, something which contributes to learning. For the exposure to have any effect, however, it should also occur over time. Learners must be given the opportunity to meet and process the word on several separate occasions in order to acquire a thorough knowledge of it (Nation, 1993; Webb, 2007), with the first repetition relatively shortly after the first encounter and then with prolonged intervals (Klingberg, 2011; Nation, 2003).

The high-frequency words discussed in the previous section constitute only part of the total number of words in the books. Hence, there are many words to which the learners are not exposed to the same degree. The cut-off point for low frequency has been set to a maximum of four tokens because although most researchers agree that at least ten encounters are needed for a word to be learned (Matsuoka, 2012, p. 158), a minimum of five encounters

is sometimes mentioned (Cameron, 2001, p. 84). Anything below that number must definitely be regarded as low frequency. The results are presented in Figures 1–3 below.

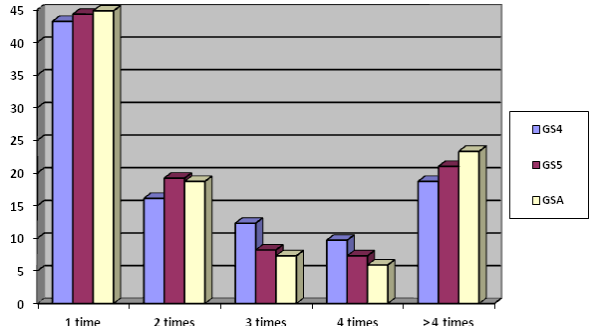

Figure 1. Frequency of adjective types (% of the total).

As shown in Figure 1, close to 45% of the adjectives in all three textbooks occur only once in the texts. When all the low-frequency adjectives are combined, they comprise between 77% and 81% of the total number of adjective types.

The pattern displayed by adjectives is the same also for nouns, as seen in Figure 2.

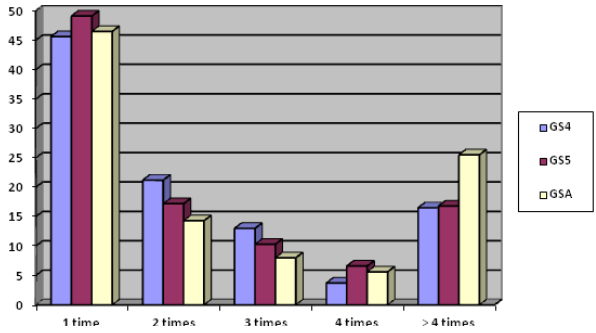

Figure 2. Frequency of noun types (% of the total).

The distribution of one-time nouns differs slightly between the books, with a range of 46%–50%. In *GS4* and *GS5*, low-frequency nouns account for approximately 83% of the total number of noun types; the corresponding figure in *GSA* is about 75%.

Even though the verbs conform to the pattern set by adjectives and nouns, the low-frequency figures are generally lower. This is illustrated in Figure 3. About 70% of all verb types in the three books have a frequency of maximum four tokens.

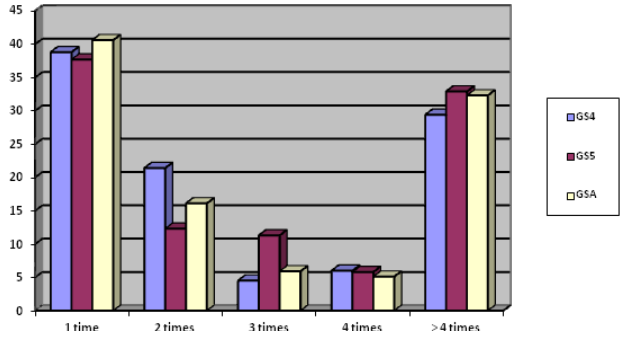

Figure 3. Frequency of verb types (% of the total).

However, the figures presented in Figures 1–3 above are somewhat misleading. Many of the words with a frequency of two, three or four tokens—and this applies to all three word classes—occur in the same text or song. Consequently, they have no exposure over time, which makes the learning opportunities provided by the books even more infrequent. Modifying the results accordingly yields new figures for one-time words: on average, 54% of adjectives, 59% of nouns and 47% of verbs occur only once.

When the three books are compared to see whether vocabulary is reused over the series as a whole, the result is similarly discouraging (as shown in Table 7).

Table 7. Words common to the three books in %.

Word class	Occurrence in one book	Occurrence in two books	Occurrence in three books
Adjectives	62.2	19.5	18.3
Nouns	64.9	19.8	15.3
Verbs	51.7	20.2	28.0
Total mean	59.6	19.8	20.5

An analysis of the words occurring in two books (one-fifth of the total) reveals that the combination *GS5+GSA* is the most common one. It is also shown that *GS4* has more adjectives and nouns in common with *GSA* than with *GS5*. Whether this is pure coincidence or a conscious form of repetition on the part of the authors—that is, that words are introduced in school year 4 in order to be reinforced in school year 6—cannot be concluded from the material at hand. An additional 20% of the words occur in all the three books, and many of these are also low-frequency words: 27% of common adjectives, 30% of the nouns and 22% of the shared verbs. With as much as 60% of the words occurring in only one book, the writers seem to have favoured (consciously or not) to opt for vocabulary breadth rather than depth. However, considering the high number of one-time words, the likelihood of learners achieving either breadth or depth must be questioned.

Since researchers on vocabulary acquisition have emphasised the importance of frequency and repetition more than once (Ellis, 2002; Nation, 1990; Schmitt, 2000), it is alarming to see how many of the words in the data analysed are of such low frequency. Without proper reuse, efforts spent on learning new vocabulary will be wasted. Unfortunately, the results of the present study are not unique. Nation (1990) reports similar figures: "In some courses, half of the different words occur only once and most of the words occur less than five times" (p. 44; cf. Matsuoka & Hirsh, 2010, p. 65). Of course, many of the words can be seen as 'fillers', words which only function as a background, to fill out the text and make it more interesting to read. As such, they can be seen as less important for classroom treatment. Yet, many of the low-frequency words are listed in the text glossaries, which, at least in Sweden, quite often means that they will be part of the vocabulary homework assigned to learners. Admittedly, this means that learners have to practice the low-frequency words, which should make them easier to learn and remember. On the other hand, despite this effort, they are unlikely to reach the level of exposure needed, thus wasting the time spent on learning them (Schmitt, 2000, p. 137). In practice, it would of course be impossible to reuse every word in a textbook to the degree necessary to achieve long-term retention. Moreover, students are not left on their own with the textbook as

the only source of language input. Vocabulary is certainly repeated in different classroom activities initiated by the teacher, and it is also possible that words are reused in exercises in the workbooks. This remains to be investigated for the *Good Stuff*-series, but it is very likely that textbook and workbook chapters mirror each other as regards vocabulary focus. What is more, many students develop their vocabulary through spare time activities performed in English, such as on-line computer games (Sundqvist, 2009), even though it might not be the same vocabulary presented in the textbooks. All things considered, however, it could be discussed whether the number of low-frequency words in the books analysed is not unnecessarily high and actually works more like a distractor, making learning more difficult.

Conclusion

There are many teachers in Sweden who, for different reasons, follow the textbook rigidly and plan all of their teaching in accordance with its structure and content (Skolverket, 2006). In the teaching material analysed here, aimed for young learners of English, variation in vocabulary is considerable. Many low-frequency words means that learners are not sufficiently exposed to new vocabulary, which in turn is not beneficial for learning, thereby making it more difficult for learners to reach the objectives set in the national curriculum and in the course plans. If such teaching materials are used in the classroom, teachers will need to go outside the textbook in order to provide learners with opportunities to engage with new vocabulary. In Swedish schools today, there are many excellent teachers who enable their pupils to develop their linguistic knowledge, but their less skilled colleagues need more support. The importance of having adequate teaching materials, whose structure and content are based on results from research within the field of second language acquisition, cannot be emphasised enough.

References

Abello-Contesse, Christián & López-Jiménez, M. Dolores (2010). The treatment of lexical collocations in EFL textbooks. In María Moreno Jaén, Fernando Serrano Valverde & María Calzada Pérez (Eds.), *Exploring new paths in language pedagogy: lexis and corpus-based language teaching* (pp. 95–109). London: Equinox.

Aitchison, Jean **(**2012). *Words in the mind: an introduction to the mental lexicon* (4th ed.). Oxford: Wiley-Blackwell.

Biemiller, Andrew (2003). Vocabulary: needed if more children are to read well. *Reading Psychology, 24*(3–4), 323–335.

Burton, Graham (2012). Corpora and coursebooks: destined to be strangers forever? *Corpora, 7*(1), 91–108.

Cameron, Lynne (2001). *Teaching languages to young learners*. Cambridge: Cambridge University Press.

Carter, Ronald & McCarthy, Michael (1988). *Vocabulary and language teaching*. London: Longman.

Coady, James **(**1997). L2 vocabulary acquisition through extensive reading. In James Coady & Thomas Huckin (Eds.), *Second language vocabulary acquisition: a rationale for pedagogy* (pp. 225–237). Cambridge: Cambridge University Press.

Council of Europe (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Strasbourg: Council of Europe, Modern Languages Division.

Ellis, Nick (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2), 143–188.

Kirk, Steven & Carter, Ronald (2010). Fluency and spoken English. In María Moreno Jaén, Fernando Serrano Valverde & María Calzada Pérez (Eds.), *Exploring new paths in language pedagogy: lexis and corpus-based language teaching* (pp. 25–38). London: Equinox.

Klingberg, Torkel (2011). *Den lärande hjärnan. Om barns minne och utveckling* [The learning brain: on children's memory and development]. Stockholm: Natur & Kultur.

Laufer, Batia (1997). The lexical plight in second language reading: words you don't know, words you think you know, and words you can't guess. In James Coady & Thomas Huckin (Eds.), *Second language vocabulary acquisition: a rationale for pedagogy* (pp. 20–34). Cambridge: Cambridge University Press.

Ljung, Magnus (1990). *A study of TEFL vocabulary*. Stockholm: Almqvist & Wiksell.

Ljung, Magnus (1991). Swedish TEFL meets reality. In Stig Johansson & Anna-Brita Stenström (Eds.), *English computer corpora: topics in English linguistics* (pp. 245–256). Berlin/New York: Mouton de Gruyter.

Matsuoka, Werner (2012). Searching for the right words: creating word lists to inform EFL learning. In David Hirsh (Ed.), *Current perspectives in second language vocabulary research* (pp. 151–177). Bern: Peter Lang.

Matsuoka, Werner & Hirsh, David (2010). Vocabulary learning through reading: does an ELT course book provide good opportunities? *Reading in a Foreign Language, 22*(1), 56–70.

Meara, Paul & Suárez García, Jesús (2010). Missing words: the vocabulary of BBC Spanish courses for adults. In María Moreno Jaén, Fernando Serrano Valverde & María Calzada Pérez (Eds.), *Exploring new paths in language pedagogy: lexis and corpus-based language teaching* (pp. 77–84). London: Equinox.

Melka, Francine (1997). Receptive vs. productive aspects of vocabulary. In Norbert Schmitt & Michael McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy* (pp. 84–102). Cambridge: Cambridge University Press.

Milton, James (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Nation, Paul **(**1990). *Teaching and learning vocabulary*. New York: Newbury House.

Nation, Paul (1993). Sixteen principles of language teaching. In Laurie Bauer & Christine Franzen (Eds.), *Of Pavlova, poetry and paradigms: essays in honour of Harry Orsman* (pp. 209–224). Wellington: Victoria University Press.

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

Nation, Paul (2003). Effective ways of building vocabulary knowledge. *ESL Magazine*, *6*(4), 14–15.

Nation, Paul (2008). *Teaching vocabulary: strategies and techniques*. Boston: Heinle Cengage Learning.

Nation, Paul & Waring, Robert (1997). Vocabulary size, text coverage and word lists. In Norbert Schmitt & Michael McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy* (pp. 6–19). Cambridge: Cambridge University Press.

Ogden, Charles K. (1944). *Basic English: a general introduction with rules and grammar* (9th ed.). London: Kegan Paul, Trench, Trubner & Co.

Phillips, Sarah (1993). *Young learners*. Oxford: Oxford University Press.

Read, John (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, John (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24*, 146–161.

Rixon, Shelagh (1999). Where do the words in EYL textbooks come from? In Shelagh Rixon (Ed.), *Young learners of English: some research perspectives* (pp. 55–71). Harlow: Longman.

Schmitt, Norbert (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schmitt, Norbert. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363.

Schoonen, Rob & Verhallen, Marianne (2008). The assessment of deep word knowledge in young first and second language learners. *Language Testing, 25*(2), 211–236.

Skolverket (2006). *Läromedlens roll i undervisningen. Grundskollärares val, användning och bedömning av läromedel i bild, engelska och samhällskunskap* [The role of teaching materials in teaching: secondary school teachers' choice, use and evaluation of teaching materials in Arts, English and Social Sciences]. Stockholm: Skolverket.

Skolverket (2011). *Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011* [Curriculum for the compulsory school, the preschool class and the leisure-time centre 2011]. Stockholm: Skolverket.

Skolverket (2012). *Internationella språkstudien 2011* [The 2011 international language study]. Stockholm: Skolverket.

Stæhr, Lars Stenius (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal, 36*(2), 139–152.

Sundqvist, Pia (2009). *Extramural English matters: out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary* (Doctoral dissertation). Karlstad University, Karlstad.

Tyler, Andrea (2012). *Cognitive linguistics and second language learning: theoretical basics and experimental evidence*. London: Routledge.

van Ek, Jan A. & Trim, J.M.L. (1998a). *Threshold 1990* (Rev. ed.). Cambridge: Cambridge University Press.

van Ek, Jan A. & Trim, J.M.L. (1998b). *Waystage 1990* (Rev. ed.). Cambridge: Cambridge University Press.

Webb, Stuart (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46–65.

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