A Corpus-based Study of Lexis in L2 English Textbooks

Cathrine Norberg
Luleå University of Technology, Sweden

Marie Nordlund
Luleå University of Technology, Sweden

Abstract—**Despite the fact that textbooks are central in foreign language learning, only limited research has explored to what extent L2 textbooks support language learning and whether the content in them is relevant from a vocabulary perspective. This study investigates the vocabulary in seven English textbooks used in Swedish primary schools. A corpus has been constructed based on the words in the textbooks. By means of a concordancing software tool, the material has been analyzed by comparing the vocabulary between the books and to words on the *New General Service List* and in the *VP-Kids corpus*. The analysis shows that many words in the textbooks occur only occasionally in common everyday language use. It also demonstrates that there is great variation in the number and selection of words across the books indicating that there does not seem to be a common thought behind word selection in textbooks used in Swedish schools.**

Index Terms—**vocabulary, course books, frequency lists, young learners of English, representativeness, corpus**

I. INTRODUCTION

Since the 1990s an increasing number of researchers have used corpus-based approaches in many areas of linguistic enquiry, among them, in the development of methods for language teaching (Biber & Reppen, 2002; Granger, 2002). Corpus linguistics has almost been established as a norm in the creation of dictionaries (McCarthy, 2008), and more recently in the construction of reference grammars (Burton, 2012). Corpora have been used as a powerful tool to identify non-native language features in foreign language learners' language production, such as the over-representation of certain phrases or linguistic features (Granger, 2002). As a pedagogical tool it has challenged many traditional approaches to language studies in a number of different ways (e.g. Hunston & Francis, 1999), and has shown particularly useful in studies of collocational patterns and word frequency (Schmitt, 2000).

As yet, however, only limited corpus linguistic research has focused on the construction of textbooks and the vocabulary content in them from a pedagogical point of view. The few studies that have employed corpus-based approaches with a focus on teaching materials have mainly focused on academic written materials (e.g. Hoey, 2000; Thompson, 2000). Studies addressing the needs of young learners have mainly been left unattended (Foster & Mackie, 2013; Keck, 2004). Considering the centrality of the textbook in foreign language learning and teaching, particularly for the lower ages of learning (Konstantakis & Alexiou, 2012; Skolverket, 2006), this is surprising.

The aim of this study is to explore i) whether textbooks used in Swedish primary schools are constructed in ways that support vocabulary learning, ii) whether there is consensus as to how many words, and the distribution of them, books used for the same school level should contain, iii) to what degree the words in the books correspond to common everyday language use, and iv) whether there appears to be a common thought behind the selection of words in the books. To do this a corpus consisting of seven English textbooks used in Swedish primary schools has been created. Following Foster and Mackie's (2013) construction of a literary corpus to analyze Dr. Seuss' works from a vocabulary standpoint, the appropriateness of using the books is discussed in the context of learning English as a foreign language (EFL), raising such questions as whether they facilitate or even hinder successful vocabulary acquisition. In analyzing to what extent the vocabulary in the textbooks corresponds to common everyday language the occurrence of lexical verbs, adjectives and nouns in the books has been compared to the representation of the same categories on the New General Service List (NGSL) (Browne, 2013) and in the *VP-Kids corpus* (Roessingh & Cobb, n.d.). Since the Swedish national curriculum states that all teaching should find its base in scientific research (Skolverket, 2011), an investigation of the construction and content of EFL textbooks from a pedagogical standpoint therefore appears both worthwhile and important. No similar study has been conducted in a Swedish context.

II. VOCABULARY KNOWLEDGE AND SIZE

To build up a substantial vocabulary is a cornerstone in foreign language learning. Vocabulary knowledge is essential in the development of reading (Cameron, 2001; Laufer & Ravenhorst-Kalovski, 2010), and has been identified as closely related to overall success in learning a foreign language. To study words is, however, not a quick fix. To fully

know a word means knowledge of meaning and form, register, collocation and association, among other things (Nation, 2013). To gain this knowledge and to enable automatization of word knowledge and recognition (Tyler, 2012) a word must be encountered many times in a variety of contexts. Words do not appear to be stored in isolation in our brains, but are connected to each other in semantic networks (Aitchison, 2012) and only when a learner encounters the same word often and in different contexts are those connections likely to be internalized and thus remembered (Nation, 2008). Language learning materials should therefore be constructed in ways that make it possible for learners to meet new words repeatedly and in different contexts (Cameron, 2001).

How often a word should recur in a textbook to enable learning is difficult to say. The figures vary among researchers with estimates ranging between 5–6 (Cameron, 2001; Nation, 1990) and twenty (Waring & Takaki, 2003) occurrences. Repeated exposure has been shown as 3–4 times more important for beginners than for advanced learners (Zahar, Cobb & Spada, 2001). Another important issue in foreign language learning is vocabulary size, both in terms of how many words a textbook for a particular level should comprise and how many words a student for a particular level of learning should know. Such figures are, for instance, presented in curricula in Greece, South Korea and Spain (Jiménez Catalán & Mancebo Francisco, 2008; Konstantakis & Alexiou, 2012; Shin & Chon, 2011), but are not visible in the learning objectives in Swedish curricula. In the works preceding the *Common European framework of reference for languages* (CEFR; Council of Europe, 2001), on which the Swedish syllabi are based, 2,000 words are, however, set as a target for B1 level – the level students should have reached for a passing grade at the end of compulsory school (school year 9).

It is also difficult to establish how many words a learner of a foreign language should know to achieve adequate comprehension of a text of average difficulty. But there are studies showing that knowledge of high-frequency words is closely related to text coverage and successful language learning (Nation, 2006; Nation & Beglar, 2007). Nation's (2006) study of what vocabulary is needed for unassisted comprehension of English, based on fourteen 1,000 word-family lists from the *British National Corpus* (BNC), shows that the first two lists cover 88% of the word tokens in a novel like *Lady Chatterley's Lover*. In a similar fashion, the 250 most frequent words in children's speech have been calculated to make up 75–80% of children's total language production (Roessingh & Cobb, n.d.). These figures undoubtedly demonstrate that a small number of words "do the bulk of the work" (Schmitt, 2000, p. 73), whereas others occur comparatively infrequently. Learning the first 1,000–2,000 words, which occur repeatedly in any discourse, should thus facilitate vocabulary learning and comprehension considerably, and provide a solid basis for further learning.

III. Previous Corpus-based Studies on Learning Materials for Young Learners

As mentioned, research focusing on the structures of textbooks for young learners is scarce (Keck, 2004), in particular with a focus on vocabulary learning and teaching (Criado & Sánchez, 2012). There are, however, some important examples. Shin and Chon (2011) studied the vocabulary profile of elementary and secondary English textbooks used in South Korea by comparing the vocabulary in them to the words on West's (1953) 2,000 *General Service List* (GSL), a list preceding NGSL, the words in Coxhead's (2000) *Academic Word List* and the vocabulary in three general corpora of English. They found that 68% of the words in the textbooks were not on the GSL, and a high number of words were academic words. The comparison of the vocabulary with the words in the three corpora showed that the textbooks contain a large number of words used infrequently in everyday language production.

Similar studies have been conducted with a focus on vocabulary in children's literature. Foster and Mackie (2013) analyzed the vocabulary frequency coverage of the words in Dr. Suess' books to the lexical coverage in the *VP-Kids corpus* and the BNC to determine the appropriateness of using the books in an EFL context. Their conclusion was that the books are fairly representative of both children's language and general language production. The frequency coverage compared to the *VP-Kids corpus* and the BNC showed almost identical figures: 86 and 84% respectively of the words in the Suess corpus were found among the 1,000 most frequent words in the two corpora. Their analysis also included a comparison of the most frequent lexical verbs, adjectives and nouns in Dr. Suess' writing to the same word classes in a children's literature corpus (the *CLLIP* corpus). Also this comparison showed a reasonably good correspondence, in particular for verbs and adjectives.

A corpus-based study to analyze the vocabulary in children's literature has also been conducted by Thompson and Sealey (2007). They compared the vocabulary profile of the *CLLIP corpus* to a corpus of adult fiction and newspaper texts to find out whether the language in writing for children demonstrates different linguistic properties compared to texts aimed at adults. They concluded that the vocabulary in children's fiction shares much of the characteristics of the language in adult fiction, but to a lesser degree the vocabulary profiles of news text. The top ten most frequent lexical verbs, adjectives and nouns in the two corpora showed a very high degree of overlap.

Another study focusing on vocabulary in a school context is Konstantakis and Alexiou's (2012) investigation of the vocabulary in five EFL textbooks used in Greece in the first two years of primary school. Quite contrary to researchers like Nation (2006, 2013) and Nation and Beglar (2007), they argue that the Greek books are insufficiently loaded with mid- and low-frequency words. Their analysis, based on a comparison to the BNC 2,000 word list, shows that the books include between 74% and 85% of the most frequent words on the BNC list. According to them, a vocabulary of this size is "insufficient for anything but the most basic form of communication" (Konstantakis & Alexiou, 2012, p. 40). A

similar standpoint is taken by Milton and Vassiliu (2000) who emphasize that frequency lists are not organized according to themes or topics. Many words describing the animal world for example, are not likely to appear among the 2,000 most frequently occurring words in general English, but are important in the world of children. Konstantakis and Alexiou's study (2012) also shows that the vocabulary across the Greek books is extremely varied in word selection and length.

Vocabulary content in teaching materials has also been addressed by Rixon (1999). She examined seven textbooks aimed at first-year beginner students of English with the purpose of uncovering evidence of consensus or divergence as to vocabulary content and vocabulary size. Like Konstantakis and Alexiou (2012), she discovered that there are wide differences in what appears as a suitable number of words to present to children in their first year of learning English. Her study also revealed that there is little consensus as to vocabulary selection across the books. Close to half of the total inventory of the words turned out to be unique to one or other of the books. Similar results are presented by Jiménez Catalán and Mancebo Francisco (2008) who conducted a corpus-based investigation of the vocabulary input in four EFL textbooks used in primary and secondary education in Spain. The present study intends to investigate whether similar phenomena are visible in English textbooks used in Swedish schools.

IV. EFL IN SWEDISH SCHOOLS

The start of EFL instruction in Swedish schools is flexible; schools may decide to start anywhere between school years 1 and 4, but the majority of students start their English education in school year 3. When EFL instruction is introduced teaching is usually limited to 20–30 minutes per week and consists mainly of songs and rhymes, and vocabulary most commonly comprises words from familiar domains (Skolverket, 2011), such as colors, clothes, family members, animals and body parts. As students advance to higher school levels, the number of classes each week increases and classes become longer. Regardless of starting age, the total number of hours for EFL instruction is 480 hours throughout the nine-year long compulsory school. National tests for English take place in school years 6 and 9 to assess students' listening, reading, speaking and writing skills. Textbooks are very infrequent for the youngest students and usually not introduced until school year 3 or 4, but from thereon they constitute an important feature of EFL teaching with approximately 75% of teachers using a textbook in almost every class (Skolverket, 2006).

V. MATERIAL AND METHODOLOGY

This study analyzes a corpus consisting of seven textbooks commonly used in Swedish primary school years 3 and 4 (students aged 9–10 years) published by three well-known Swedish publishers. The books *Happy* (Hansson, 2010)*, Lift off, juniors* (Bowen & Söderlund, 2007)*, Lighthouse* (Bowen & Söderlund, 2003) and *Right on!* (Nihlén, Gardenkrans & Robinson Ahlgren, 2006) are used for school year 3, and *Good stuff* (Keay, Coombs & Hoas, 2005)*, New champion* (Bermheden, Sandström & Wahlgren, 2005) and *What's up?* (Göransson, Hjälm, Widlund & Cowle, 2006) for school year 4. The corpus, hereafter referred to as SWYLC (the *Swedish Young Learner Corpus*), consists of a total of 34,380 running words. Although some of the books have accompanying workbooks, our analysis in this article is limited to the material in the textbooks. Since content words are the most important carriers of meaning in communication, the focus of attention is on lexical verbs, nouns and adjectives. They make up slightly more than a third of the corpus (12,714 words). SWYLC was constructed by scanning all the textbooks and transforming the texts into .txt format. The electronic documents generated were carefully checked to eliminate possible scanning errors due to font and background problems. Words that had not been scanned (e.g. words in pictures) were entered manually. All the texts were then tagged according to, among other things, word class, number and tense in accordance to the BNC Basic C5 Tagset[1]. To search the corpus the concordance software MonoConc Pro was used.

A type in this study includes the base form of a word and all inflected regular forms. Irregular word forms, both plural noun forms and verb forms, which are often learned as separate words, are listed as separate types. To analyze the proportion of "raw vocabulary input" in relation to new vocabulary introduced in the textbooks the type/token ratio (TTR) was used (Criado & Sánchez, 2012). It is often used to compare the lexical diversity of materials, and is calculated by using the formula type/token x 100 (Seong-sig, 2002). The higher this ratio the more varied and specialized a particular material is in terms of its vocabulary. This ratio is thus an important indication of to what extent words are recycled in a material. The TTR score can, however, only be used to compare material of equal size, as the type/token relationship typically follows a parabolic curve: the larger a text is, proportionally the fewer the types (Sánchez & Cantos, 1997). Since the books analyzed here differ in length, the standardized type/token ratio was used to homogenize the figures and make them comparable (see Table 1). It calculates the TTR based on every 1,000 word and calculates an average of it (Seong-sig, 2002). The *WordSmith Tools* (Scott, 2008) software program was used for the calculation of the STTR.

To establish to what extent the vocabulary input in the books reflect common language production, and to measure the proportion of high- and low-frequency words in them, the words in the books were compared to the occurrence of lexical verbs, adjectives and nouns in NGSL and *the VP-Kids corpus*. NGSL is derived from a 273 million-word sample of the *Cambridge English Corpus* (a 2-billion word corpus) and covers the 2,800 most frequently used words by native speakers of English. The *VP-Kids corpus* covers the 2,500 most frequently produced words by over 500 native English-

speaking children (5–7 years of age). It is derived from samples of spontaneous speech including more than 700,000 words. These two reference materials have been widely used in comparative analyses of lexical distribution and frequency coverage (e.g. Cobb, 2007; Foster & Mackie, 2013; Nation, 2004; Thompson & Sealey, 2007). In this study, comparison was made by using the Web-VP BNC-20 tool found on the Compleat Lexical Tutor v. 8 website.[2] The tool sorted the textbook material into groups matching three groups of NGSL and ten groups of the *VP-Kids corpus* (see Tables 3 and 4 below).

Even if NGSL covers general language, and our study has a focus on the language intended for young learners, a comparison to it, combined with the *VP-Kids corpus*, appeared relevant. It has been shown that there seems to be little difference in the distribution of the most frequently used words produced by children and general language production (Foster & Mackie, 2013). The study will thus not provide a comprehensive result, but is nevertheless likely to give telling indications of to what extent the books analyzed reflect general language production. It should be noted that in analyzing the vocabulary profile of SWYLC (see 6.3) compared to native language production, the comparison was only made to NGSL, as the *VP-Kids corpus* does not provide frequency information of separate word classes.

VI. ANALYSES

A. Types and Tokens in the Textbooks

To explore whether there is a systematic approach to vocabulary size in the construction of textbooks for young EFL learners in Sweden, and whether the books facilitate vocabulary learning, the number of types and tokens and the type/token ratio in the entire corpus and its sub-corpora were analyzed. The figures are presented in Table 1 below.

TABLE 1.
NUMBER OF TOKENS AND TYPES IN SWYLC AND ITS SUB-CORPORA.

	Tokens	Types[3]	STTR
SWYLC	34,380	2,899	30.31
Corpus Year 3	10,721	1,272	27.03
Happy	1,186	254	24.20
Lift off, Juniors	2,492	440	26.55
Lighthouse	4,464	520	23.15
Right on!	2,570	673	34.60
Corpus Year 4	23,668	2,538	31.65
Good Stuff	9,424	1,311	31.48
New Champion	5,758	926	33.34
What's up?	8,486	1,305	30.42

As can be seen, there is significant variation between the books both as to the number of word types (different words) and tokens (running words) in them. There are 419 more types in the book with the largest number (*Right on!*: 673) than the one with the fewest (*Happy*: 254) for school year 3. The difference in terms of the number of types for school year 4 is 385 with 1,311 types in *Good Stuff* compared to 926 in *New Champion*. The difference between the books with the lowest and highest number of tokens is 1,384 for year 3 and 3,666 for year 4. Also the type/token relationship displays notable differences across the books, in particular for school year 3 with a STTR of 23.15 for *Lighthouse,* to be compared to 34.60 for *Right on!*. The STTR for school year 4 is 30.42 for *What's up?* and 33.34 for *New Champion.* These differences in the base number of words in books intended for the same level of learning indicate that there is a lack of an explicit target of how many words students for a particular level of learning should know, and in line with this how many words books intended for the same school level should contain.

Differences in lexical diversity imply, as mentioned, different opportunities for repetition of new words introduced in materials. There is no established figure as to what is an optimal type/token relationship for a language learning material. But to bring some understanding of the figures presented here, the STTRs reported from similar studies on learning materials could be used. Criado and Sánchez (2012), for example, studied students' opportunities for repetition in two EFL textbooks used in Spanish elementary schools and reported almost identical STTRs as for SWYLC, and concluded that although the proportion of tokens differs between the books, both offer adequate opportunities for repetition with figures not deviating much from the expected pattern of normal distribution of words in texts where about 50% of the words in any non-manipulated text are likely to occur only once (Criado & Sánchez, 2012, pp. 90–91). It should, however, be mentioned that the figures presented in their study are not entirely comparable with those for SWYLC. Criado and Sánchez (2012, p. 82) regard different word forms, such as singular and plural forms of regular nouns as two separate types. In our study they are defined as two forms of the same word type. Following their definition of type, the number of types in SWYLC would be much higher, and thereby also the STTR.

Another point of reference for the type/token ratio presented for SWYLC is the figures presented in Seong-sig's (2002) study of four elementary school English textbooks used in South Korea which, among other things, analyzes the lexical diversity of the books, and compares the figures to the lexical variation of authentic spoken language. The STTR presented for the Korean books is on average 16.09, to be compared to 33.12 for general English. Seong-sig explains that the difference mainly depends on the fact that many expressions are repeated continually across the books to ensure

learning. In this light, the STTR for SWYLC may again appear high, and may also be taken as a sign that recycling of words has not been considered in constructing the texts.

To better understand the figures behind the type/token relationship presented for SWYLC from a vocabulary learning perspective, types and tokens of the lexical verbs, adjectives and nouns together with the type/token ratio for these categories in the respective books were analyzed (see Table 2 below). Since some of the categories presented include figures lower than 1,000, and the STTR is the average TTR based on every 1,000 words, TTRs are presented instead of STTRs. As mentioned, the TTRs of material with large differences in size are not one hundred percent comparable. Therefore the presented TTRs should be interpreted with some caution.

TABLE 2.
LEXICAL VERBS, ADJECTIVES AND NOUNS IN SWYLC AND ITS SUB-CORPORA.

	Lexical verbs			Adjectives			Nouns		
	Tokens	Types	TTR	Tokens	Types	TTR	Tokens	Types	TTR
Year 3	1,073	144	12.71	868	111	13.11	2,234	472	20.66
Happy	89	29	34.57	76	21	26.67	328	127	39.27
Lift off, Juniors	223	65	26.05	219	35	16.44	547	176	31.55
Lighthouse	435	64	14.68	336	58	17.26	791	193	24.32
Right on!	328	94	27.58	237	71	31.25	568	293	47.97
Year 4	2,478	271	10.56	1,292	266	20.41	4,156	1,083	25.76
Good Stuff	1,151	198	16.74	532	157	28.36	1,636	554	32.56
New Champion	557	124	22.40	387	122	32.03	890	357	39.75
What's up?	770	140	17.14	372	145	38.58	1,630	611	37.27

A comparison of the figures for *Happy*, the book with the highest TTR for lexical verbs (34.57) with the figures of *Lighthouse*, the book with the lowest TTR (14.68) for the same category, as an example, illustrates the correlation between the degree of recycling and a high or low TTR score. Closer analysis of the figures for the two books showed that 16 of the 29 verb types in *Happy* (55%) occur only one time in the book, and five of them (17%) more than five times. In *Lighthouse*, 16 of 64 of the verbs (30%) occur one time and eighteen (28%) more than five times. A low TTR is thus an indication of a higher degree of recycling. In a similar fashion, a closer look at the figures of the TTR for nouns in *Right on!*, the book with the highest TTR (47.97), compared to *Lighthouse*, the book with the lowest (24.32) revealed that 131 of 293 nouns (45%) in *Right on!* occur once in the book, and fourteen (5%) five or more times. In *Lighthouse*, 56 of 193 (29%) nouns occur one time, and forty of them (21%) five or more times. Again, a lower TTR indicates a higher degree of recycling. Considering that five exposures to a language item have been estimated as a minimum for learning to take place (Cameron, 2001; Nation, 1990) and that young learners need more than older learners, the figures presented for SWYLC show, although with some variation, that recycling has received little attention in constructing the books.

B. The Frequency Distribution of Lexical Verbs, Adjectives and Nouns in SWYLC.

The previous section discussed the number of types and tokens across English books used in Swedish primary schools. This section presents an analysis of the proportion of high- and low-frequency words in SWYLC, and how it corresponds to the vocabulary in NGSL. Using the Web-VP BNC-20 tool, the word classes analyzed were compared to the distribution of these words in NGSL (Table 3) and the *VP-Kids corpus* (Table 4). The tool sorted the material into three groups matching the groups of NGSL with K1 representing the 1,000 most frequently occurring words in English, K2 the next 1,000 most frequent words and so on (K1–K3). The words matching the words in the *VP-Kids corpus* followed the same principle, but with words sorted into ten groups of 250 words in descending frequency order. Words appearing beyond the list and corpus compared were categorized as off-list words.

TABLE 3.
FREQUENCY DISTRIBUTION OF LEXICAL VERBS, ADJECTIVES, NOUNS IN SWYLC COMPARED TO NGSL.

Frequency level	Lexical verbs		Adjectives		Nouns	
K1	53.63		32.04		30.35	
K2	22.15	(75.78)	25.57	(57.61)	15.62	(45.97)
K3	7.96	(83.74)	8.74	(66.35)	10.26	(56.23)
Off-list	16.26	(100)	33.65	(100)	43.77	(100)

The figures are percentage of total number of words with accumulated coverage in parenthesis.

As shown in Table 3, 53.63% of the lexical verbs, 32.04% of the adjectives and 30.35% of the nouns in SWYLC are found among K1 level words in NGSL. These figures make an average of 38.68% and show that a high percentage of the content words in the books, nouns and adjectives in particular, are words occurring rather seldom in common everyday language. On average a third of the words are not found among the 2,800 most frequent words in English. The figures for nouns in this respect are as high as 43.77%. It has been estimated that 80% of the words in English are found among K1 words (Nation, 2006). Even though this percentage includes function words as well, the figures presented nevertheless suggest that a high proportion of the lexical words in the books occurs relatively infrequently in everyday language use.

The comparison of SWYLC to the *VP Kids-corpus* (Table 4) shows a better correspondence compared to the figures presented in relation to NGSL with 52.48% (62.84% of the verbs, 45.30% of the adjectives and 49.29% of the nouns) of the words found among the 1,000 most frequently occurring words (see the fourth frequency level).

TABLE 4.
FREQUENCY DISTRIBUTION OF VERBS, ADJECTIVES, NOUNS IN SWYLC COMPARED TO THE *VIP KIDS-CORPUS*.

Frequency level	Lexical verbs		Adjectives		Nouns	
1–250	18.40		10.03		13.49	
251–500	20.83	(39.23)	15.53	(25.56)	13.42	(26.91)
501–750	13.89	(53.12)	9.71	(35.27)	12.08	(38.99)
751–1,000	9.72	(62.84)	10.03	(45.30)	10.30	(49.29)
1,001–1,250	6.60	(69.44)	8.74	(54.04)	6.08	(55.37)
1,251–1,500	5.56	(75.00)	5.18	(59.22)	6.52	(61.89)
1,501–1,750	3.82	(78.82)	4.85	(64.07)	5.19	(67.08)
1,751–2,000	2.08	(80.90)	4.85	(68.92)	4.08	(71.16)
2,001–2,250	2.43	(83.33)	3.24	(72.16)	3.71	(74.87)
2,251–2,500	1.74	(85.07)	1.62	(73.78)	3.56	(78.43)
Off-list	14.93	(100)	26.21	(100)	21.57	(100)

The figures are percentage of total number of words with accumulated coverage in parenthesis.

Seen in relation to estimations that the 250 most frequently occurring words used by children, of which 44% cover lexical verbs, adjectives and nouns, make up 75–80% of their total language production (Roessingh & Cobb, n.d.), the figures presented for SWYLC are again low and show that the Swedish books contain a high proportion of lexical words seldom used by native-speaking children. Only an average of 14% of the word classes investigated are found among the words listed in the first frequency band (1–250) of the *VP-Kids corpus*. On average a fifth of the words in SWYLC are not found in the *VP-Kids corpus*.

C. The Vocabulary Profile of SWYLC Compared to NGSL

To analyze the vocabulary profile of the books and compare it to general language production, the top ten most frequently occurring lexical verbs, adjectives and nouns in SWYLC and NGSL were compared. The words and their proportions are presented in Table 5. It should be noted that irregular word forms are listed as separate types in SWYLC, but not in NGSL. Since the *VP-Kid corpus* does not provide frequency information, no words from that corpus are presented. To enable comparison between the books, Tables 6–9 (Appendix) present the frequency of the top ten most frequently occurring words in the individual textbooks. Following Rixon (1999), and to answer the question about a possible core vocabulary in books intended for a particular school level, the words were divided into four groups illustrating in how many of the books they appear: one book, two–four, five–six, and all seven books.

TABLE 5.
THE TOP-TEN MOST FREQUENT LEXICAL VERBS, ADJECTIVES AND NOUNS IN SWYLC AND NGSL

Lexical verbs				Adjectives				Nouns			
SWYLC		NGSL		SWYLC		NGSL		SWYLC		NGSL	
like	0.74	**say**	0.38	red	0.31	good	0.22	mum[b]	0.38	time	0.23
go	0.63	**go**	0.36	big	0.28	new	0.14	dad	0.32	people	0.22
got	0.58	**know**	0.30	**little**	0.20	first	0.13	friend	0.23	year	0.20
see	0.42	**get**	0.29	white	0.20	right	0.12	name	0.27	way	0.13
come	0.35	think	0.28	black	0.20	last	0.10	dog	0.25	thing	0.13
want	0.29	**like**	0.27	blue	0.18	long	0.08	house	0.24	**day**	0.12
look	0.29	make	0.24	**old**	0.15	**great**	0.07	**day**	0.16	child	0.10
play	0.29	**see**	0.23	green	0.13	**little**	0.07	cat	0.15	company	0.10
say	0.20	take	0.19	**great**	0.12	**old**	0.07	beep	0.13	life	0.09
love	0.20	**come**	0.18	yellow	0.12	each[a]	0.07	animal	0.13	man	0.09

[a] *different, high, late, next* also 0.07
[b] includes instances of *mom*
Figures are percentage of running words.
Bold font illustrates words appearing in both SWYLC and NGSL.

As shown, six of the most frequent verbs (*like, go, got, see, come,* and *say*) are found among the most frequently occurring words in general language. Although this may be understood as a sign of correspondence, there are also notable differences. Closer analysis of the frequency patterns of the verbs identified as occurring with high frequency in both SWYLC and NGSL, for example, shows that *like, go, got* and *see* appear considerably more frequently in the books compared to the NGSL. *Like* and *go* account for 0.74% and 0.63% respectively of the total tokens in SWYLC compared to 0.27% and 0.36% in NGSL. Considering that SWYLC is a much smaller corpus compared to the material NGSL is based on, the focus on these two verbs in the textbooks stands out even more and may be seen as a distinctive feature of SWYLC. In *Happy* (see Appendix, Table 6) the percentage of *like* is close to 2% of all the words in the book. Since irregular forms are separate types in SWYLC, but not in NGSL, the proportional difference is even higher for *see* and *go*. It is also noticeable that the informal verb form *got* is extremely common in the textbooks. This figure is not shown explicitly in NGSL, as it is included among the tokens of *get*. But, considering that *get* accounts for 0.29% of the

verbs in NGSL, and *got* 0.58% of the words in SWYLC, the figures clearly show the importance placed on this verb form in the textbooks. It is the most common verb in *What's up?* appearing as often as 93 times.

On the whole, the verbs in the textbooks indicate a pattern where great emphasis is placed on students' ability to express likes and wants, their coming and going in the world, and their physical observations of it. Verbs expressing abstract reasoning and opinions such as *think* and *know* are not among the most frequent verbs in the books, but are listed as common words in NGSL. Closer analysis of *see* in the books shows that the majority of the tokens involve visual seeing and not understanding. Despite such signs of a core vocabulary as to the representation of verbs, it should be noted that only *like* and *see* are represented as high-frequency verbs in all the seven books (see Appendix, Table 6). *Play* and *go* occur in five and six books respectively, the other verbs in between two and four of the books.

If the verbs in the textbooks show a certain degree of overlap with the verbs in NGSL, the lexical profile of the adjectives and nouns shows a more diverging result: only three of the adjectives (*great, little* and *old*) and one of the nouns (*day*) in SWYLC are found among the ten most frequent adjectives and nouns in NGSL. The most eye-catching observation of the adjectives is the extreme focus on colors in the textbooks. Six color terms are found among the top ten most frequently occurring adjectives in SWYLC. *Red* occurs proportionally more frequently in the books than the most frequent adjective *good* in NGSL. Closer examination of the individual textbooks (Appendix, Table 7) moreover revealed that between three and seven colors are among the most common adjectives in the books (cf. Rixon, 1999). The dominance of colors is particularly noticeable in *Lift off, juniors* (year 3) with *red*, *black* and *yellow* as the three most frequent adjectives in the book covering close to 3% of all tokens. The same book also shows a high percentage of *green, blue* and *white*. No colors are listed among the most frequently occurring words in NGSL. Apart from color terms, there seems to be little consensus as to what adjectives students should learn. Of a total of 32 different adjectives listed as the top ten most common adjectives in the books (Appendix, Table 7), fifteen are unique to one book.

As for nouns, the textbooks indicate a focus on words for family members and the animal world with *mum/mom* and *dad*, and for the latter *dog* and *cat* as particularly common. Like color terms, these words are not listed as the most frequently occurring items in NGSL. It is also noticeable that general terms for humans (*people, man, child, woman,* etc.) are not among the most frequent words in SWYLC. These are common terms in general language production (*woman* is missing). Other similar superordinate terms frequent in NGSL, but not equally common in the teaching materials with the exception of *day* are, for instance, *time* and *year*. *Time* occurs with high frequency in two of the books, but not in the others. *Year* does not appear as a high-frequency word in any of the books.

It can thus be concluded, that except for the representation of verbs, the lexical profile of SWYLC diverges in many significant ways from everyday language. As for a common thought behind the selection of words in the books, there seems to be a general consensus among publishers that color terms, words for pets and family members, and ways of expressing likes are important words for young learners to know. Frequency data of this kind should, however, as pointed out by Thompson and Sealey (2007), be studied with caution, and should always be complemented with contextual analysis. Closer analysis of the words listed as high-frequency words in the individual books (Appendix, Tables 6–8) showed that although there seems to be a certain degree of consensus as to vocabulary selection, there is considerable variation across the books. Only *like* and *see* occur among the top ten most frequently occurring words in all the books*,* and 71 of a total of 118 words are unique to one book.

VII. DISCUSSION AND CONCLUSION

In analyzing the lexical content of English textbooks used in Swedish primary schools at two different school levels from a vocabulary acquisition standpoint, this study has shown that books used in Swedish schools for the purpose of learning and teaching English as a foreign language vary to a considerable extent both in vocabulary size, the type/token relationship and the selection of words. A comparison to reference materials (*New General Service List* and VP-Kids corpus) covering general language production and the language produced by native English-speaking children moreover shows that the books include a large proportion of low-frequency words, that is, words not frequently used in common everyday language.

Wide differences across the books in terms of content and vocabulary size do not only indicate an absence of a common pedagogical idea as to the construction and content of vocabulary intended for a specific school level in the Swedish school system, it also implies that students following different books will be provided with different opportunities for vocabulary exposure, and as a consequence possible disparities in vocabulary size (cf. Jiménez Catalán & Mancebo Francisco, 2008). Compared to the type/token ratio in books presented in similar studies (e.g. Seong-sig, 2002), the figures for the Swedish textbooks moreover appears high, which means that students using the books are provided with limited opportunities to consolidate new words introduced to them. Preliminary results from accompanying workbooks indicate similar results (Nordlund & Norberg, forthcoming).

One may perhaps believe that books intended for young learners should focus on the language used by children, and therefore deviate a bit from the frequency coverage of general language. The material analyzed does indeed show a better correspondence with the frequency of the words in the *VP Kids corpus,* but the figures for the books are again low in relation to those of native-speaking children's language production (Roessingh & Cobb, n.d.).

In establishing whether there is consensus as to what words students should learn for a specific school level and whether these words correspond to the vocabulary profile of everyday language, this study reports both diverging and

converging results. Similarities with general language use are mainly found among verbs with six of the ten most common verbs in the books corresponding to the top ten most frequently used verbs in English, although with proportional differences. Despite this partial overlap, which to some extent was expected, as many of the verbs are almost impossible to avoid in general speech, great differences in the lexical profile of adjectives and nouns indicate that the distribution of words in everyday language production has not been considered in the selection of words for the books. Only three of the adjectives and one of the nouns in the books are high-frequency words in general English. Compared to the words in NGSL, the absence of superordinate terms like *people, child, thing* and *man* is noticeable in the textbooks, although such terms would, as pointed out by Rixon (1999, p. 68), provide students with networks and hierarchies of meaning and enable them to make their meaning come across even when more precise words are missing.

As to the question whether the material analyzed shows a consensus in the vocabulary input, analyses of the vocabulary in individual books revealed that even if the material seems to focus on words important in children's lives, closer analysis of what words are shared by the books, showed that only two words are high-frequency words in all the seven books, and 62% of the top ten most frequent words (71 of 118) are exclusive to one book. Since different books were analyzed for years 3 and 4 in this study, it was not possible to investigate vocabulary progression within the same textbook series, but a comparative analysis of school years 4–6 books of two of the teaching materials part of this study (*Good stuff* and *New champion*; Nordlund, 2016) shows that only about 20% of the words recur throughout all three books and approximately 60% occur in one book only. A high proportion of these words are one-time words. Such figures indicate a low area of common consent as to vocabulary input in the books (cf. Konstantakis & Alexiou, 2012).

Evidently, there are words in the books which are not among the most frequently occurring words in general language use, but which nevertheless are important to know if progression beyond beginner level is to take place, and which are important to make communication possible in different thematic areas (Konstantakis & Alexiou, 2012, p. 36). Therefore, the representation of words like *dog, cat,* and color terms may fulfill a purpose, although one may question the enormous focus placed on them. Most Swedish students attending third and fourth year of school already know the names of the most common colors and pets in English from nursery rhymes and pre-school activities. A bit surprising is the complete absence of words related to school activities in the books. Not even the very word *school* is a high-frequency word in the books.

The authors of this study are fully aware of the fact that students need to know more than the 1,000–2,000 most frequent words in English to communicate and get their meanings across. It is also evident that it is not possible to produce a text consisting of only high-frequency words. Such texts would be both boring and awkward, and there are themes that require words more seldom used in English. The point is rather to find a good balance of high- and low-frequency words in texts to promote learning and make the texts enjoyable and meaningful to read and work with. To exactly determine what this balance should be to provide learners with the best learning possibilities is difficult to say, but including too many low-frequency words in materials for young learners in particular is not likely to help students acquire words and build up a basis for further learning. On the contrary, it may lead to both frustration and lack of interest, and as a consequence unsatisfactory results.

This study has provided empirical evidence of the lack of common criteria as to vocabulary size, lexical diversity and vocabulary profile in EFL textbooks used in Swedish primary schools. Like similar studies conducted on EFL textbooks used in other countries, it appears as if the books have not been constructed in consideration to research on how words are learned. Since the textbook is central in many language teaching activities, further studies focusing on vocabulary acquisition and the design of textbooks from different perspectives and contexts are encouraged. The present study has shown the value of using corpus linguistic approaches in analyzing textbooks, and how materials can be compared and analyzed by means of using free concordancing tools on the Internet. Future corpus-based studies on learning materials could, for instance, explore how national tests are constructed from a vocabulary perspective. If EFL textbooks used in Swedish schools do not explicitly show or mention what words students for a specific level of learning should learn, national tests, which are constructed to assess students' knowledge of words and language comprehension, would perhaps bring light to what is understood as passing knowledge for a particular school level.

Notes
1. Available at: www.natcorp.ox.ac.uk/docs/c5spec.html
2. Available at: www.lextutor.ca/vp/comp
3. MonoConc Pro does not take lemma into consideration, which means that the number of lemmas in Tables 1 and 2 was adjusted manually.

APPENDIX

TABLE 6.
THE TOP-TEN MOST FREQUENT LEXICAL VERBS IN THE TEXTBOOKS CONSTITUTING SWYLC.

Happy		*Lift off, Juniors*		*Lighthouse*		*Right on!*		*Good Stuff*		*New Champion*		*What's up?*	
like	1.77	see	0.80	like	1.01	go	1.13	like	0.88	see	0.61	got	1.10
play	1.10	like	0.76	come	0.83	let	0.74	go	0.76	got	0.54	go	0.72
read	0.59	sleep	0.56	got	0.72	wear	0.51	want	0.51	go	0.40	like	0.65
point	0.51	go	0.52	see	0.65	like	0.47	come	0.55	know	0.36	want	0.40
see	0.51	look	0.52	put	0.58	run	0.43	love	0.44	look	0.36	get	0.31
call	0.42	play	0.44	look	0.56	see	0.43	say	0.44	like	0.35	play	0.31
love	0.25	come	0.32	go	0.40	got	0.39	got	0.35	think	0.24	eat	0.25
jump	0.17	sat	0.32	say	0.38	miss	0.35	know	0.29	hear	0.21	live	0.24
open	0.17	scream	0.24	play	0.31	play	0.31	live	0.28	say	0.21	see	0.22
say[a]	0.17	chase[b]	0.20	thank	0.29	sing	0.31	see	0.28	take[c]	0.21	look[d]	0.18

[a] *climb* and *ask* also 0.17
[b] *eat* and *put* also 0.20
[c] *thank* and *want* also 0.21
[d] *love* also 0.18

Top-ten adjectives appearing in one book, two–four books, five–six books and all the books:
1 book (19): *ask, call, chase, get, hear, jump, let, miss, open, point, read, run, sat, scream, sing, sleep, take, think, wear*
2-4 books (10): *come, eat, know, live, look, love, put, say, thank, want*
5-6 books (3): *go, got, play*
All 7 books (2): *like, see*

TABLE 7.
THE TOP-TEN MOST FREQUENT ADJECTIVES IN THE TEXTBOOKS CONSTITUTING SWYLC.

Happy		*Lift off, Juniors*		*Lighthouse*		*Right on!*		*Good Stuff*		*New Champion*		*What's up?*	
big	0.76	red	1.00	big	0.76	red	0.89	red	0.31	good	0.43	new	0.24
old	0.76	black	0.92	good	0.47	hot	0.54	little	0.30	nice	0.28	old	0.20
blue	0.59	yellow	0.84	left	0.38	blue	0.51	big	0.23	big	0.23	great	0.19
black	0.51	little	0.76	yellow	0.36	good	0.47	best	0.16	great	0.21	nice	0.18
new	0.51	blue	0.72	bad	0.34	green	0.39	good	0.16	old	0.21	American	0.14
green	0.42	green	0.64	green	0.34	white	0.39	white	0.15	white	0.21	good	0.14
red	0.42	white	0.52	red	0.34	black	0.31	blue	0.13	sorry	0.19	hot	0.12
white	0.42	big	0.44	right	0.29	favourite	0.27	great	0.13	young	0.17	fun	0.11
small	0.34	quick	0.32	happy	0.27	runny	0.27	black	0.12	red	0.16	big	0.08
pink[a]	0.25	brown[b]	0.24	long	0.27	yellow	0.27	cool	0.12	black[c]	0.14	happy[d]	0.08

[a] *yellow* also 0.25
[b] *fat* and *long* also 0.24
[c] *hot* also 0.14
[d] *lost* and *sorry* also 0.08

Top-ten adjectives appearing in one book, two–four books, five–six books and all the books:
1 book (15): *American, bad, best, brown, cool, fat, favourite, left, lost, pink, quick, right, runny, small, young*
2-4 books (13): *blue, great, green, happy, hot, little, long, new, nice, old, sorry, white, yellow*
5-6 books (4): *big, black, good, red*
All 7 books (0)

TABLE 8.
THE TOP-TEN MOST FREQUENT NOUNS IN THE TEXTBOOKS CONSTITUTING SWYLC.

Happy		*Lift off, Juniors*		*Lighthouse*		*Right on!*		*Good Stuff*		*New Champion*		*What's up?*	
balloon	0.93	cat	0.92	finger	0.83	day	0.51	dad	0.51	mum	0.71	friend	0.86
cat	0.67	fish	0.76	dog	0.74	zoo	0.51	mum[d]	0.40	dad	0.57	name	0.59
dog	0.67	house	0.76	grandma	0.54	name	0.43	day	0.38	chapter	0.54	beep	0.53
house	0.67	mouse	0.76	nose	0.49	time	0.43	dog	0.31	morning	0.31	dad	0.22
book	0.59	witch	0.56	animal	0.43	turn	0.43	summer	0.30	grandad	0.19	house	0.19
name	0.59	dog	0.48	night	0.34	mum	0.35	friend	0.25	man	0.19	present	0.19
rabbit	0.59	bird	0.44	tree	0.34	morning	0.31	bed	0.24	boy	0.17	animal	0.16
red	0.59	dad	0.44	box	0.31	nose	0.31	name	0.24	reporter	0.17	cousin	0.16
bag	0.51	hat	0.44	cat	0.31	dentist	0.27	time	0.22	school	0.17	football	0.16
football[a]	0.51	mum	0.44	birthday[b]	0.27	brother[c]	0.23	doctor	0.21	bed[e]	0.16	family[f]	0.15

[a] *pet* also 0.51
[b] *sun* and *zoo* also 0.27
[c] *colour, dress, home* and *sister* also 0.23
[d] includes instances of *mom*
[e] *house* also 0.16
[f] *film* also 0.15

Top-ten nouns appearing in one book, two–four books, five–six books and all the books:
1 book (37): *bag, balloon, beep, bird, birthday, book, box, boy, brother, chapter, cousin, dentist, doctor, dress, family, film, finger, fish, granddad, grandma, hat, home, man, mouse, night, pet, present, rabbit, rat, reporter, school, sister, summer, sun, tree, turn, witch*

2-4 books (15): *animal, bed, cat, dad, day, dog, football, friend, house, morning, mum, name, nose, time, zoo*
5-6 books (0)
All 7 books (0)

REFERENCES

[1] Aitchison, J. (2012). Words in the mind: An introduction to the mental lexicon (4th ed.). Oxford: Oxford University Press.
[2] Bermheden, C., Sandström, L.-G., & Wahlgren, S. (2005). New champion 4. Stockholm: Bonnier Utbildning.
[3] Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, *24*(2), 199–208.
[4] Bowen, M., & Söderlund, K. (2003). Lighthouse 4. Malmö: Gleerups.
[5] Bowen, M., & Söderlund, K. (2007). Lift off, juniors. Malmö: Gleerups.
[6] Browne, C. (2013). The New General Service List: Celebrating 60 years of vocabulary learning. *The Language Teacher, 37*(4), 13–16.
[7] Burton, G. (2012). Corpora and coursebooks: destined to be strangers forever? *Corpora, 7*(1), 91–108.
[8] Cameron, L. (2001). Teaching languages to young learners. Cambridge: Cambridge University Press.
[9] Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology, 11*(3), 38–64.
[10] Council of Europe. (2001). Common European framework of reference for languages: Learning, teaching, assessment. Strasbourg: Council of Europe, Modern Languages Division.
[11] Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238.
[12] Criado, R., & Sánchez, A. (2012). Lexical frequency, textbooks and learning from a cognitive perspective: A corpus-based sample analysis of ELT materials. *Revista Española Lingüística Aplicada, 25*(1), 77–94.
[13] Foster, J., & Mackie, C. (2013). Lexical analysis of the Dr. Seuss Corpus. *Concordia Working Papers in Applied Linguistics, 4*, 1–21.
[14] Göransson, M., Hjälm, A., Widlund, K.-E., & Cowle, A. (2006). What's up? 4. Stockholm: Bonnier Utbildning.
[15] Granger, S. (2002). A bird's eye view of learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3–33). Philadelphia: John Benjamins.
[16] Hansson, C. (2010). Happy year 3. Textbook. Malmö: Gleerups.
[17] Hoey, M. (2000). The hidden lexical clues of textual organization: A preliminary investigation into an unusual text from a corpus perspective. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 31–42). New York: Peter Lang.
[18] Hunston, S., & Francis, G. (1999). Pattern grammar: A corpus-driven approach to the lexical grammar of English. Amsterdam: John Benjamins.
[19] Jiménez Catalán, R.M., & Mancebo Francisco, R. (2008). Vocabulary input in EFL textbooks. *RESLA, 21*, 147–165.
[20] Keay, C., Coombs, A., & Hoas, B. (2005). Good stuff 4. Stockholm: Almqvist & Wiksell/Liber.
[21] Keck, C.M. (2004). Corpus linguistics and language teaching research: Bridging the gap. *Language Teaching Research, 8*(1), 83–109.
[22] Konstantakis, N., & Alexiou, T. (2012). Vocabulary in Greek young learners' English as a foreign language course books. *The Language Learning Journal, 40*(1), 35–45.
[23] Laufer, B., & Ravenhorst-Kalovski, G.C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15–30.
[24] McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching, 41*(4), 563–574.
[25] Milton, J., & Vassiliu, P. (2000). Frequency and the lexis of low level EFL texts. In K. Nicolaidis & M. Mattheoudakis (Eds.), *Proceedings of the 13th International Symposium on Theoretical and Applied Linguistics* (pp. 444–455). Thessaloniki: Aristotle University.
[26] Nation, I.S.P. (1990). Teaching and learning vocabulary. New York, NY: Newbury House.
[27] Nation, I.S.P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 3–13). Amsterdam: John Benjamins.
[28] Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59–82.
[29] Nation, I.S.P. (2008). Teaching vocabulary: Strategies and techniques. Boston, MA: Heinle Cengage Learning.
[30] Nation, I.S.P. (2013). Learning vocabulary in another language (2nd ed.). Cambridge: Cambridge University Press.
[31] Nation, I.S.P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13.
[32] Nihlén, C., Gardenkrans, L., & Robinson Ahlgren, A. (2006). Right on! 3. Stockholm: Liber.
[33] Nordlund, M. (2016). EFL textbooks for young learners: A comparative analysis of vocabulary. *Education Inquiry, 7*(1), 47–68.
[34] Nordlund, M., & Norberg C. (forthcoming). Vocabulary recycling in EFL teaching materials for young learners.
[35] Rixon, S. (1999). Where do the words in EYL textbooks come from? In S. Rixon (Ed.), *Young learners of English: Some research perspectives* (pp. 55–71). Harlow: Longman.
[36] Roessingh, H., & Cobb, T. (n.d.). Profiling the vocabulary of K-2 learners. Retrieved on April 19, 2016 from http://www.uclagary.ca/langugaelearning/profilingvocabulary.
[37] Sánchez, A., & Cantos, P. (1997). Predictability of word forms (types) and lemmas in linguistic corpora. A case study based on the analysis of the *CUMBRE Corpus*: An 8-million-word corpus of contemporary Spanish. *International Journal of Corpus Linguistics, 2*(2), 259–280.
[38] Schmitt, N. (2000). Vocabulary in language teaching. Cambridge: Cambridge University Press.
[39] Scott, M. (2008). WordSmith Tools version 5. Liverpool: Lexical Analysis Software.
[40] Seong-sig, K. (2002). A corpus-based analysis of the words in the elementary school English textbooks. *English Teaching, 57*(3), 253–277.

[41] Shin, D., & Chon, Y.V. (2011). A corpus-based analysis of curriculum-based elementary and secondary English textbooks. *Multimedia-Assisted Language Learning, 14*(1), 149–175.

[42] 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.

[43] Skolverket. (2011). Curriculum for the compulsory school, preschool class and the recreation centre, 2011. Stockholm: Skolverket.

[44] Thompson, P. (2000). Citation practices in PhD theses. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 91–102). New York, NY: Peter Lang.

[45] Thompson, P., & Sealey, A. (2007). Through children's eyes? Corpus evidence of the features of children's literature. *International Journal of Corpus Linguistics, 12*(1), 1–23.

[46] Tyler, A. (2012). Cognitive linguistics and second language learning: Theoretical basics and experimental evidence. London: Routledge.

[47] Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader. *Reading in a Foreign Language, 15*(2), 130–163.

[48] West, M. (1953). A general service list of English words. London: Longman.

[49] Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review*, *57*(4), 541–572.

Cathrine Norberg is Associate professor of English at Luleå University of Technology, Sweden. Her main research interest is corpus linguistics with a focus on gender, and educationally related issues.

Marie Nordlund is Senior Lecturer in English at Luleå University of Technology, Sweden. Her main research interest lies within corpus linguistics and applied linguistics.