

Estonian, Latvian, Lithuanian, Russian and Swedish Young Learners' Written Production in EFL - Descriptions and Comparisons of Their Use of Vocabulary

Stellan Sundh
Uppsala University
Sweden

Abstract

English as a foreign language is learnt and used by many young people in Europe. It is therefore relevant to investigate the production in English by young learners from different geographical, cultural and language backgrounds. The present study is based on data collected over a period of nine months with 161 twelve-year-olds' writing. The vocabulary was analysed in terms of number of words in the texts, the proportions of high-frequency vocabulary and the type-token-ratios. The results show that generally the young learners in Estonia, Latvia, Lithuania, Russia and Sweden have similar uses of vocabulary no matter of topic or L1 and only tendencies of minor differences were identified.

Keywords: young learners, writing in English as a foreign language, vocabulary, Baltic Sea region

1. Background

English has more and more taken the role of a lingua franca in the world, particularly among young people. Since most young people in Europe start to learn English as their first foreign language, we may expect that English will be established as the most common means of communication in many contexts and places worldwide in the future. The attitude is generally positive to learning English and when there is a choice between languages in the early years, parents typically choose English, as they do for instance in Latvia (Murphy 2014: 133). This tendency is also found when the decision is taken by local authorities in Europe since they almost invariably choose English (Enever 2011). The learning of English takes place in different geographical, cultural and educational contexts all over the world. In addition to young people's language learning at school, their acquisition of English is influenced by factors outside school and in their spare time, for instance when they play computer games, watch films and listen to music. As a result, young learners in for instance Sweden are daily exposed to EFL (Murphy 2014: 154). For these reasons it is relevant to investigate young learners' free production in various regions in order to find out about similarities and differences in their use of English. The results could point at traits that are similar and which therefore facilitate communication but also at differences that could lead to misunderstandings. Since lexical errors tend to be judged more seriously than structural ones (Ellis 1994), we may expect that misunderstandings are most often due to erroneous lexical choices and not structural errors. There are thus good reasons to study the vocabulary used by young learners with different backgrounds. When analysing written productions, the age factor should be taken into account in order to make it possible to compare the quality in the productions. At the same time, 12-year-olds develop differently in several respects and variation in the uses of English can possibly be explained by individual differences since "all children are unique, and two children at the same chronological age can exhibit markedly different characteristics" (Nunan 2011: 13).

2. Learning vocabulary in English as a foreign language (EFL)

As stated above, vocabulary is relevant to investigate since the choice of words is important in communication. In addition, words play a special role in language learning for young language learners; learning words makes second language acquisition concrete. Vocabulary acquisition is also central in language learning, no matter if it concerns L1 or L2 (Schmitt 2010; Zimmerman 1997; Coady & Huckin 1997). Language learners learn words from an early stage and go on doing so up to advanced levels. At the elementary stages, the interest lies primarily on single words but gradually combinations of words in terms of idiomatic expressions, collocations, lexical bundles and multi-word units are introduced. In this way the process of learning words can be described as becoming more and more advanced when series of words gradually are treated together in the learning process instead of being focused on individual words.

Educated native speakers of English are expected to learn up to 20,000 words whereas 2,000 to 3,000 words out of the most frequent words in English are sufficient in order to understand and carry out everyday tasks (Nation & Waring 1997), or to minimally follow conversations (Nation 2006). By making young learners aware of the fact that by consistently working on learning a limited number of high-frequency words, they will eventually manage to communicate in everyday conversation. This will probably increase their motivation in the acquisition of vocabulary in a foreign language.

Two basic distinctions in discussions about the learning of vocabulary are passive and active lexical knowledge. The former refers to the knowledge of words which is required for the receptive skills: listening and reading. Knowing many words facilitates the listening or reading of a text. The latter distinction refers to the requirements which are needed to speak or write. The knowledge of words is thus related to the receptive skills, or in their more active usage to the productive skills. In the latter category, the young learners' lexical richness is identified in their writing and speaking and is an indication of their level of language use. Learners' proficiency in English is closely related to and correlates with their knowledge and use of vocabulary (Daller & Xue 2007: 150; Schmitt 2010: 4). It is clear that learners generally know more words in the receptive skills than in the productive ones, and that this knowledge gap decreases as the learner's language proficiency develops (Ortega 2009: 88). A learner who has the ability to use a wide range of words in adequate combinations and in addition is able to adjust the use to the required settings and situations in both writing and speaking has reached a high level of language proficiency. Several studies also conclude, which can be expected, that there is a correlation between language proficiency and lexical richness (See for instance, Ferrell Tekmen & Daloglu 2006).

Lexical knowledge can be described in terms of different layers in vocabulary knowledge (Meara 1996; Schmitt 2000). When learners state that they know a word, the depth of this knowledge is determined by a number of factors such as knowing (1) how the word sounds, (2) how it is spelled, (3) what other word parts (prefixes and suffixes) it can appear with, (4) what other words that often turn up together with the word (collocation and idiomatic factors), (5) the different meanings of the word, (6) and finally, in what contexts and registers the word occurs. Learning vocabulary is complex and comprises several aspects. One of these aspects is the fact that words often are learnt in a context and together with other words. A number of concepts are used to describe the phenomenon that words tend to occur with certain other words, for instance collocations, chunks, and prefabs (prefabricated patterns). This combining of words is here defined with the help of two of them: formulaic language and collocations. (*Formulaic* is defined as "made up of fixed patterns of words or ideas"; *collocation* is "a combination of words in a language that happened very often and more frequently than would happen by chance", Oxford Advanced Learner's Dictionary.) This implies that formulaic language is a set of words with a commonly recognized and inherent meaning, whereas collocations are words that are often used together and which sound adequate to a native speaker when combined. The learner who knows a word should additionally be aware of the different meanings of this word which then put more demands on the learner with deeper knowledge of the word which can be described in terms of "the depth of knowledge" (Schmitt 2010: 16). Finally, the learner who knows a word is expected to master the inherent grammar of this word by being capable of producing the different forms of the word.

The learners' L1 with the lexical resources acquired plays an important role also for the acquisition of vocabulary of L2. Consequently the sizes of vocabulary in L1 and L2 for learners at school correlate strongly. In addition, the lexical errors which can be explained by the interference from L1, often described as negative transfer, constitute a substantial part of the errors found in learners' production of L2 (Schmitt 2010: 25). Learning the meanings of words in L2 is thus related to this process in L1. Previous studies indicate that at an early stage of the learning of words in L2, the acquisition is filled with the conceptual content from L1. This is for instance shown in Jiang's study on young Koreans (2004), and the fact that L2 words have fewer meanings and are less elaborated than L1 words is emphasized by Finkbeiner & Nicol & Nakamura (2004) in a study on Japanese-English bilinguals. Learning words in L2 is furthermore related to the power of the memory capacity, which was investigated in a study with Greek 8 to 13 year-olds. The results of this study show that at an initial stage a superior memory capacity is of significance but that after some years, further learning of words is boosted by a more developed and large size of mental lexicon in L2 (Masoura & Gathercole 2005). The fact that young learners' memory capacity is linked to their acquisition of vocabulary in L2 has furthermore been confirmed in a study on Finnish elementary school children (Service & Kohonen 1995).

At the same time, there is great individual variability; some children can learn some 1,000 words in their first years of instruction whereas other children learn only a handful of words. This significant variability was noticed in a Hungarian study on 253 children's receptive knowledge of vocabulary (Orosz 2009).

Learning vocabulary is intentional or incidental (Schmitt 2010: 29) and the teaching of vocabulary can be described as being direct or indirect (Barcroft 2004: 201). These four terms at least partly overlap and are not clear-cut but are rather to be described along a continuum. The intentional learning of words is planned and the purpose of the activity is to learn new words. Evidently this often takes place in a school context. The incidental learning of words on the other hand can be described as an extra benefit while doing other things with language, such as reading or playing computer games in English. This incidental learning currently seems to play a significant role in the acquisition of vocabulary. Various studies point at this phenomenon with concepts such as Extra-mural English (Sundqvist 2009). Furthermore, this has been described with the help of results from simple case studies of students who made great progress, such as the 11-year-old Chinese girl with an interest in fan fiction in English (Black 2006) and with discussions of the impact of the cultures associated with the language learnt (Lindgren & Stevenson 2013). In the same way as with intentional and incidental learning, the two terms direct and indirect teaching can be described. The teaching of vocabulary is direct when instructions are given to pay attention to words and the learning of these words, as opposed to indirect when instructions are provided simply about the reading or writing of a text without any further details about working explicitly with the words in the text.

3. Young Learners' Writing in EFL

When studying young learners' writing in L2, it is relevant to highlight the complexity of their writing process and writing development in a foreign language. This complexity is a matter of the learners' intended interaction with the reader with the text, the meaning-making, the text construction, the accuracy and the formal properties of the L2 in use. Nevertheless, in spite of the complexity in the writing process, the conclusions drawn in a study on Swedish 11-year-olds' writing of letters in EFL are that if they have a strong willingness to express themselves, they can be successful in their attempt to create texts that communicate a meaning. In this context the genre used in the study (letter-writing) is highlighted as an important factor; letter-writing is "a simple and familiar genre for the students" (Lindgren & Stevenson 2013). In letter-writing, salutations and valedictions are central and these phrases were investigated in Spanish young learners' productions in EFL in a longitudinal study over three years which showed that stages in the development in the use of greetings could be identified in their letters (Jiménez Catalán 2016). Results from another study, namely a study on Czech 8 to 11 year-olds' writing in EFL, came to the conclusion that young learners can reach far in their writing and produce "clear, meaningful texts of their own" (Kovariková 2016). Results from previous studies with the advice provided in methodological literature thus indicate that giving learners the opportunity to produce texts in EFL can be expected to lead to progression in their written proficiency (Keaveney & Lundberg 2014).

The question is whether a time span and frequent writing lead to improvement in grammatical accuracy, text length, spelling, coherence and content. In a study on 42 Korean children aged 9 to 12 and practicing writing in EFL in a 315 hours long program called *English for Young Learners* (EYL) and tested three times, the results showed great improvements in a linear progression. The results were varying but remarkable across all the components investigated (Bae & Lee 2012). In a Swiss study on young learners' writing in EFL and teachers' teaching styles, the conclusion is that progress in the skill of writing in EFL is associated with some teacher factors and that in order to be successful in writing, learners need to start early with activities such as copying and letter-writing (Loder Buechel 2015).

When studying young learners' learning of writing in EFL, it is important to have in mind that the traditional division of language learning and teaching into the four skills is not relevant for very young learners as they do not see language as a linguistic system since their overall literacy skills neither in L1 nor in L2 are developed sufficiently. Learning to write in EFL naturally interconnects with other knowledge, such as vocabulary and discourse, and is gained as an effect of oral interaction (Cameron 2001). When the young learners have come to the stage in their learning that it is time to develop writing in EFL, this writing puts great demands on the learners with their attention on factors such as spelling and punctuation, linking ideas logically and thereby organizing the text, and in addition using vocabulary and sentence structures (Hedge 1988). But at the same time, writing can be a useful skill when teaching EFL to young learners provided that the children are ready and interested in using the English writing system (Pinter 2006).

The skill of writing in L2 is closely connected to their literacy in L1 and there is a relationship working both ways: early learning of L2 in addition works in favour for their literacy skills in L1 (Murphy 2014). The learners' L1 is a factor of interest when studying their production. It may be expected that when learners' L1 is closely related in lexis and language structure to English, the production is to be more developed and sophisticated. There are many extra-linguistic factors that play important roles for the learning of EFL which need to be considered however. These factors are for instance the learner's motivation, the teacher's methodology and the parents' attitudes.

Making learners write in EFL is a useful way of capturing language development; written production provides substantial evidence of the occurrence of lexis and syntactic complexity. Particularly free production gives a picture of the current state of learners' interlanguage. Writing in EFL is therefore used both for summative and formative assessments. Furthermore, the interest in the cognitive aspects of writing has increased so that writing is regarded to go hand in hand with learning. There has thus been a gradual change from viewing the writing process as a linear process with brainstorming, drafting, and revising and thereby organizing the text, to a cognitive process and a way for the learners to explore what they think and know (Lundahl 2014: 179). In this way writing becomes a more independent activity for the learners when they are free to explore their resources and potential for writing in EFL. Nevertheless, for many children early writing in EFL needs to be carefully scaffolded by prompts, questions, pictures or other stimulus, for instance when they are given the instruction to describe something, such as a place or a person (Lundahl 2014: 186).

4. Investigating learners' use of vocabulary in written production of EFL

Research on learners' vocabulary is a slightly new field of interest in studies of learner language. There are claims however that the scope in research on learner language has broadened in the last few decades and now also includes lexical studies (Barcroft 2004: 200). Traditionally studies on learners' English present results from error analyses (see e.g. Ellis & Barkhuizen 2005), strategies when learning grammatical accuracy (see e.g. Malmberg 2000), the assessment and testing of learner production (see e.g. Nikolov 2016) or when it comes to vocabulary, focus on the teaching of vocabulary itself (Ipek 2009; Mohebbi 2013; Nation 2011). Furthermore, investigations so far have seldom used data produced by young learners but rather analyses on advanced learners' vocabulary such as undergraduate students in contexts of academic writing. In studies on young learners' production, they are seldom compared with productions by native speakers. Studies have been carried out on written production by students of various L1s and are available for research thanks to international learner language corpora, such as the International Corpus of Learner English (ICLE) (Granger & Dagneaux & Meunier 2002). There is however a lack of corpus data concerning young learners' foreign language composition writing, which is emphasized with Greece as an example by Griva, Tsakiridou and Nihoritu (2009). In the same way the written production by primary school EFL learners is described as a field of investigations of an overlooked age group in Spanish research on learner language (Jiménez Catalán 2016). Studying learners' free productions, and then both oral and written, has thus currently become of interest particularly when studies are linked to assessments and grading in international comparisons and with the help of the CEFR (Common European Framework of References for Languages, Council of Europe 2001). In investigations of free production by students and particularly at the first levels of proficiency, such as A1 and A2 in the CEFR-scale, vocabulary plays an important role for successful communication. For this reason, there are claims that investigating learners' uses of vocabulary in free language production, and at different levels of language proficiency, is in need of more research (Schmitt 2010: 43).

As stated above, the learners' L1 is a factor of interest when analysing written production in EFL. Lindgren & Munoz (2013) state that the language distance between the L1 and the L2 is an important factor for L2 development for children. In a study on 30 Spanish 11-12-year-olds it is stated that the learners' background of English studies, together with the writing system in their L1 are factors of influence (Artero 2013). According to these results, variation in language use can be expected in texts written in EFL by learners who have different L1, and then particularly if there is a difference in language distance between their L1 and the English language.

5. Research questions

The description above on the state of English as a lingua franca in the world today, the discussion of young learners' vocabulary and writing in EFL and the account of previous studies on young learners' writing in EFL provide a background for the three research questions of the present study which are listed in what follows:

- In what ways does a period of nine months influence the length of 12-year-olds' written production and the vocabulary used?
 - What can be observed in the choice and complexity of vocabulary in written production in EFL by Estonian, Latvian, Lithuanian, Russian and Swedish young learners?
 - In what ways do different topics in essays affect the length of the texts and the choice of vocabulary?
- The purpose of the study is thus to find out about young learners' written production in English in five countries in north Europe with special interest in the vocabulary produced.

6. Material and Methods

6.1. The Data from the Baltic Young Learners' of English Corpus (BYLEC)

The present study is based on 12-year-olds' written production in EFL. In 2015, 2016 and 2017 texts written by 12-year-olds were collected in a project carried out at Uppsala University, Sweden in cooperation with schools and six universities in Estonia, Latvia, Lithuania, Poland, Russia and Sweden. The collection of texts in the BYLEC project was organized in two rounds: 491 students participated in the first round (2015-2016) by writing six different texts on six different occasions in nine months and then 700 other students participated in the second round (2016-2017) by writing about the same six topics as in the first round. Details about the learners' background, such as gender, L1 and years of English studies at school were also collected. The topics of the six texts, which were carefully scaffolded by prompts (cf. Lundahl 2014 above) were (1) *My best friend*, (2) *My pet*, (3) *A place I like*, (4) *An adventure/a journey of my dreams*, (5) *My favorite website/ computer game/app*, and (6) *Me in the world in the future*. The writing was carried out in September, October and November and then in February, March and April the following year and without any assistance can be described as unguided writing (Scrivener 2005: 193). BYLEC is a corpus by learners from different cultural and geographical backgrounds who all produced six texts with the same instructions. For the purposes of the present study, it is worth observing that the learners L1s were Estonian, Lithuanian, Russian and Swedish respectively. A majority of the Latvian young learners had Russian as L1 (67%). For more details of the collection of the texts and the young learners' profiles in BYLEC, see (Sundh 2016).

In the present study, only the learners from the first round in the BYLEC-project and only the ones who produced all the six texts are included in the investigation. As stated above, all in all 491 learners participated in the first round of the project and wrote between one and six texts. Since only the learners who wrote all the six texts in the academic year 2015-2016 are investigated, this leads to the fact that analyses are carried out on six different texts written by 161 learners: 14 Swedish, 85 Latvian, 37 Lithuanian, 36 Estonian and 19 Russian learners. The data comprise 966 texts with a total of 155,786 words. Table 1 shows the total number of words across the tasks and the country categories.

Table 1 The total number of words in each category of text (task) across the country categories

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Total
Estonia	4,902	5,599	6,937	7,428	6,148	6,343	37,357
Latvia	8,689	10,393	11,094	10,288	12,585	12,359	65,408
Lithuania	4,179	5,140	4,292	5,251	5,071	4,805	28,738
Russia	1,389	1,188	1,053	1,395	1,463	1,502	7,990
Sweden	2,771	2,561	3,010	2,990	2,654	2,307	16,293
Total	21,930	24,881	26,386	27,352	27,921	27,316	155,786

More than half of the students who participated in writing the six texts are Latvians (85 Latvians of the total of 161) and therefore there is a major part of Latvian student writing (42%) in the data investigated (65,408 words of the total 155,786).

6.2. Methods

In what follows the methods selected are presented to investigate (1) the length of the texts when nine months as well as the six topics are considered and (2) the choice and the complexity of the vocabulary. Firstly, the lengths of the texts were investigated with the average numbers of words per task and country category. Secondly, the learners' use of vocabulary and their lexical frequency profiles were analysed. This was carried out with three tools for frequency-based corpus studies.

The first tool is Vocab profile developed by Paul Nation, Alex Heatley and Averil Coxhead and it provides several ways of analysing vocabulary (Schmitt 2010: 205). The principle in the analysis is that vocabulary is classified in so called frequency bands based on different word lists (Lextutor 2015). In the present study, VP-Compleat is used together with the word list BNC-20, which corresponds to the 20,000 most frequent words in the British National Corpus and the words analysed are categorized in a number of frequency bands. Both the tool and the word list are recommended by Schmitt (2010:208) and used in previous research on learners' vocabulary use (see e. g. Kim & Ryoo 2011). The 1,000 most frequent words are found in the category K1, the following 1,000 words are in K2 and so forth. VP-Compleat also categorizes words in an off-list, and the words in this list do not occur in any the bands mentioned above. For this study it is relevant to concentrate on the K1, K2 and K3-levels since vocabulary research suggests that a realistic target for children learning a foreign language is 500 words a year with good learning conditions (Nation 1990) and the learners in the present study have learnt EFL for up to six years at schools.

The second tool is the Longman Dictionary Vocabulary Checker and displays the 9,000 most important words to learn in English, as well as words that are assigned to the Academic Word List (AWL). The category F1 contains the 3,000 most frequent words, F2 has the next 3,000 'mid-frequent' words, and F3 the last 3,000 'less-frequent' words. (Longman Dictionaries Online U.S.A. 2017).

The third tool is Wordsmith which is used for a type-token analysis (TTA) of the texts to obtain type-token-ratios of the texts (TTR). By calculating the TTR, a measure is obtained that provides a picture of the lexical richness and thereby an indication of the complexity of the text. What is problematic with this measure is that it is influenced by text length; short texts can have higher scores of TTR than long ones (Ellis & Barkhuizen 2005:155). Since the students' texts in the present study are quite short and vary in length to a great extent the standardized type-token ratio (STTR) will be used. The TTR-score is then calculated on a set number of words, in the present study 100 words, and then an average score is calculated that shows how many types that are found for every 100 tokens. In the present study, the students' texts are analyzed together for each category of country. This method provides an overall picture of the range of vocabulary in the learners' texts for each country category.

6.3. Methodological considerations

The texts of the six tasks are analysed separately since the topics in the texts may be expected to influence the learners' selections of vocabulary. In addition, only the students who wrote all the six texts in each country category are included in the analyses. In this way the risk is avoided that some students who participated only in a few tasks with their own individual style influence the results, for instance when comparing the productions across the tasks. The texts vary in length to a great extent and that is the reason why the texts are analysed together within each country category or task.

When studying the vocabulary in the production by learners of English who have different cultural and language backgrounds, as is the case with the data from BYLEC, the learners' L1 has to be considered. This is of interest since the occurrence of both positive and negative transfer from L1 could influence the results when identifying the high-frequency words in the texts. If a word is a high-frequency word in the learner's L1, but a similar word is a low-frequency word in English, it could imply that this English low-frequency word is easier to acquire for learners with this L1, in comparison with other learners for whom this is not the case. Previous research shows, however, that frequency tends to be quite parallel where L1-L2 counterparts are studied (Schmitt 2010:66).

The occurrence of common spelling errors in these young learners' production is a factor to take into account. Erroneous spelling is a very common feature in the texts and evidently, words that are wrongly spelt are not recognized or identified as English words with the help of the three tools described above. This means that the words that are wrongly spelt are not included in the frequencies but identified as, for instance off-the-list words in Vocabprofile. Since there are many wrongly spelt words in the texts, and that it is in some cases even impossible to comprehend what English word the student uses, it would not be feasible to correct all the erroneous spellings throughout the material to have data with no spelling errors. When this factor is taken into account in the analyses and the discussion, the results will nevertheless be useful and show tendencies in young learners' language use in the texts across the country categories.

7. Results

7.1. The Length of the Texts

The results in Table 2 show the average number of words in the six tasks with the different topics produced by the young learners in the five countries. The tendency is clear that on average the texts produced by the Estonian and Swedish learners are longer and the Russian learners' texts are shorter than the other learner categories' texts. This tendency is clear no matter the task. It is worth emphasizing however, that there are differences between the individual learners within each country category.

Table 2 The average number of words per student in the six tasks across the five country categories

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6
Estonia	136	156	193	206	171	176
Latvia	102	122	130	121	148	145
Lithuania	113	139	116	142	137	130
Russia	73	63	55	73	77	79
Sweden	198	183	215	214	190	165

Most of the 85 Latvian students in the present investigation of BYLEC have Russian as their L1; in the student questionnaire 79 out of the 85 Latvian students answered that Russian is "the language I speak at home with my parents". This remarkable difference in the length of the texts produced by Russian students in comparison with the other categories cannot therefore simply be related to these students' L1 and the language distance between English and Russian.

7.2. The Results from the Analyses using Vocabprofile

The lengths of the texts in terms of number of words as presented above are useful but we also need to investigate the occurring words in terms of their frequency in the English language. For this reason, the percentages are provided for the words in two bands (K1 and K2) across the six tasks and altogether per each country category. Words that are not found in the English language due to spelling errors or simply non-existent in English turn up in the Off-list (OL). Table 3 shows only minor differences both when comparing the production in the six texts, and when studying the results in the six country categories.

Table 3 The occurrence of high frequency words (token in %) in the bands K1 and K2 and in words occurring in the Off-list (OL) in the Vocabprofile-analysis (Classic version)

	Estonia			Latvia			Lithuania			Russia			Sweden		
	K1	K2	OL	K1	K2	OL									
Texts 1	86.3	5.6	12.7	84.8	9.7	9.2	84.5	4.5	10.5	84.5	6.5	8.2	82.6	4.4	12.8
Texts 2	84.5	7.8	7.1	80.8	9.7	9.2	84.3	7.1	8.5	81	9.8	9.2	83.2	4.0	12.4
Texts 3	87.6	5.3	6.5	83.0	6.4	10.0	83.7	6.5	9.5	80.2	5.6	13.2	84.3	5.2	9.7
Texts 4	84.6	6.8	8.3	81.8	7.5	10.2	84.3	6.7	8.7	80.6	4.2	14.9	84.3	4.2	11.2
Texts 5	85.6	3.7	8.8	84.6	4.7	7.8	84.2	3.8	8.7	79.1	3.2	15.8	86.3	2.7	9.6
Texts 6	85.4	4.3	9.2	86.3	4.1	8.1	86.1	4.2	8.3	87.9	2.7	9.0	87.2	3.4	8.4
Texts 1- 6	85.8	4.9	8.6	83.6	6.3	8.8	84.5	5.5	9.0	82.3	5.2	11.8	84.6	4.0	10.7

The percentages in K1 range from 79.1 to 87.9 and in K2 from 2.7 to 9.8. The greatest variation within country categories can be seen in the production by the Russian learners; the highest and lowest scores are found in this category. The explanation may lie in the fact that these texts were the shortest ones. The percentages for all texts per country category show similar tendencies with the K1-figures between 82.3 and 85.8, the K2-figures between 4.0 and 6.3, and the Off-list figures between 8.6 and 11.8. In what follows two examples are provided to illustrate the use of words of K1, K2 and more advanced words. In example (1) *cool* and *ice cream* are K2 words and all other words are K1. In example (2) *precisely* is a word that falls outside the K1 and K2 categories and is thus classified as less frequent in English.

(1) "When I am with my friends we buy some cool things and eat the ice-cream." (Estonian learner, text 3)

(2) "The place I like is my home and more precisely my room because I spend a lot of time there." (Estonian learner, text 3)

7.3. The results from the analyses with the Longman Dictionary Vocabulary Checker

The results in Table 4 show the distribution of the learners' vocabulary in percentages across the three bands F1 ("most frequent words"), F2 ("mid-frequent words") and F3 ("less frequent words") and in the six texts separately and altogether and in the five country categories.

Table 4 The percentages of the occurrence of words in the Longman Vocabulary Checker with words in F1 (top 3,000 words), F2 (the next 3,000 words and F3 ("the less frequent yet important 3,000 words) across the texts and country categories

	Estonia			Latvia			Lithuania			Russia			Sweden		
	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
Texts 1	85.5	9.1	2.7	83	8.3	3.1	84.4	9.5	3.1	81.8	7.2	2.5	80.4	7.5	2.4
Texts 2	85.6	9.3	2.7	83.6	10.6	2.7	84.5	8.4	3.6	85.7	8.6	2.9	81.5	7.8	3.1
Texts 3	83.7	14.4	3.4	83	14.4	2	81.9	14.8	1.4	78.4	13.6	1.5	79.8	13.7	1.6
Texts 4	84.6	14.1	2.1	82.4	11.5	1.9	84.5	12.8	1.7	76.6	9.6	1.7	81	11.5	1.8
Texts 5	77.6	10.8	2.6	81.4	10	3.5	81.5	9.4	2.6	78	8.7	1.9	81.2	9.7	2.1
Texts 6	80.7	9.7	2.6	82.5	9.5	2.3	81.8	9.7	2	82.7	7.6	2	80.7	8.1	3.6
Texts 1- 6	82.9	11.5	2.7	82.6	10.7	2.6	83.1	10.7	2.4	80.5	9	2.1	80.8	9.8	2.4

The overall tendency is that there are similar proportions of the top 3,000 words in English (F1) in the six texts written by learners in all the five country categories (between 80.5% and 83.1%). The texts 3 and 4 with the topics *A place I like* and *An adventure/a journey of my dreams* triggered more vocabulary in the band F2 for all country categories than in the other texts with a range in percentages from 9.6% to 14.8%. The topics of the texts thus play a role for whether learners choose or need to use more or fewer frequent words (band F1) and mid-frequent words (band F2) irrespective of their L1. Regarding the F3 words (the less frequent yet important 3,000 words), they are not frequently used in any of the texts of the six topics or by any of the learners in the five country categories. The percentages range from 1.5% to 3.6%. The frequent words of band F1 are illustrated by example 4 below with all the words of the sentence are in this band.

(4) My life will be very interesting. (Lithuanian learner, text 6)

The words *robot*, *room* and *washing* in example 5 are all classified as F2 and with the rest of the words in the example as F1.

(5) I will have a robot which will clean my room and washing. (Lithuanian learner, text 6)

Finally, *strategic* as used in example 6, belongs to the band F3.

(6) Often I'm playing action and strategic games.

7.4. The results from the analyses with Wordsmith

The results in Table 5 show the results from the type token analysis and thus the figures for types (distinct words) and standardized type-token ratios (STTR). The STTR give scores showing how varied the vocabulary in the text is with an average of types on every 100 tokens. It is worth highlighting here that since each different word will be counted as an individual item, this means that misspelled words will increase the score.

Table 5 The types and standardized type-token-ratio in the six texts and across the five country categories

	Estonia		Latvia		Lithuania		Russia		Sweden	
	Types	STTR								
Texts 1	742	63.2	996	61	677	61.7	284	39.2	608	63.4
Texts 2	835	62.1	1,193	62.3	801	61.9	292	42.2	569	63.6
Texts 3	994	58.5	1,472	59.2	755	59	319	39.6	568	59.9
Texts 4	1,069	61.3	1,534	62.1	887	60.7	421	36.5	607	62.8
Texts 5	875	60.8	1,611	61.4	919	61	447	34.4	578	61.5
Texts 6	1,036	59	1,478	59.6	747	59.9	386	40.8	513	60.9
Texts 1- 6	3,039	60.7	4,807	61.1	2,789	60.9	1,347	40.6	1,951	62.1

The results show that the STTR does not differ much in the texts written by learners in Estonia, Latvia, Lithuania and Sweden. The ratios range from 58.5 to 63.6 in all texts. When it comes to the Russian learners, the ratios are lower and ranging from 34.4 to 42.2 in their texts. As was shown in Tables 1 and 2, the Russian learners' texts are considerably shorter which overall provided a more limited data from this category. But even though the length of the texts is taken into account, the types of words used are fewer by these Russian learners.

8. Discussion and pedagogical implications

The results show that there were no differences in the quality of vocabulary as regards frequency level between the different categories of learners in spite of the fact that their L1s are different in terms of structural distance to the English language. This suggests that other factors than the learners' L1 play important roles in their acquisition of English vocabulary and the vocabulary they were to use in written production on certain topics.

In writing in EFL, when there is time to plan and work individually on the text, 12-year-olds with different educational and cultural backgrounds can have similar topics and tasks and produce texts on their own in these five countries. This finding is in line with the results by Kovariková (2016).

To a large extent the vocabulary used by the learners in the six texts belongs to the 3,000 most used words in the English language, irrespective of the learners' L1 and cultural background; approximately 80% of the vocabulary used was in Band K1 (Vocabprofile) and F1 (Longman). This implies that for young learners of English who are asked to write about everyday topics such as their friend, their pet, or their computer game, knowledge of basic and common words in English provides substantial support and comprises 80% of the words employed for this purpose. However, to a certain extent, the topic of the text is a factor that decides the vocabulary needed and thereby used. The topics of texts 3 and 4, for instance writing about *A place I like* required words that belong to a slightly higher band since all learners' vocabulary used in these texts was between 9.6% and 14.8 % in Band 2 (the next most common 3,000 words) according to the Longman bands and higher percentages than in the texts 1, 2, 5 and 6. This implies that even a limited range of vocabulary is enough for communication when writing about everyday matters for 12-year-olds in Estonia, Latvia, Lithuania, Russia and Sweden.

Learners from Russia show slightly different patterns in comparison with learners in the other four country categories. Their texts are shorter and the STTR is lower in all texts which imply that on average fewer types of words per 100 words were employed. The explanation does not lie in possible influence from L1 since we do not find the same tendency among the Latvian learners whose L1 also is Russian. One interpretation can simply be that Russian young learners have a more limited vocabulary when writing than learners of the other categories have due to their educational and cultural background. There may however be other explanations for this difference such as that some learners were more inclined to take risks in language production than the Russian learners and thereby they have many spelling errors which leads to a higher STTR. Another explanation may be that some learners were more used to free written production at school and particularly about the topics than the Russian learners and in this way they were more prepared to produce long texts and use a wide range of vocabulary. More research is needed on the vocabulary used in EFL by young learners with different educational and cultural backgrounds but with the same mother tongue and then with analyses of their depth of this lexical knowledge in terms of collocations (Meara 1996; Schmitt 2000). In addition, how the process of writing in EFL is taught and learnt at schools, teachers' teaching styles and whether the teaching of vocabulary is indirect or direct then need to be taken into account (Keaveney & Lundberg 2014; Loder Buechel 2015; Barcroft 2004).

9. Conclusion

The results show that a period of an academic school year does not influence the length of the texts or the variation in vocabulary for young learners with different backgrounds who are to write six texts on everyday topics. No great variation can be seen in the choice or quality of vocabulary in terms of frequencies in English of the words selected by young learners from different countries in the Baltic Sea Region. Longer time than a school year is required to identify developments in the lengths of written production in EFL by 12-year-olds which is in line with previous studies (Jiménez Catalan 2016).

Generally there were no major differences in terms of the proportions of frequent English vocabulary in the texts with different topics. To a limited extent the six topics put different demands on the words needed in the production however, since two of the six topics triggered a slightly higher proportion of words which are slightly less frequent in English. This leads to the significance of looking at the construction of the tasks in writing in EFL in order to trigger 12-year-olds' use of more advanced and uncommon vocabulary in English.

The learners' L1s and their distances to English do not seem to play a major role for the range of vocabulary produced in English. It is evident that other factors than L1 play a role for the length in and quality in the written production in EFL by Latvian and Russian young learners since they have the same L1 (Russian) but show different results in the analyses of vocabulary used. These factors could be related to the educational contexts, the learners' risk-taking or the exposure to the English language in their cultures. The tendency of a general variety of English lexis among Estonian, Latvian, Lithuanian and Swedish 12-year-olds is thus a predominant feature to have in mind when discussing EFL writing. The claim that the language distance between the learners' L1 and English, such as Swedish and Estonian, is an important factor for language development in EFL (cf. Lindgren & Munoz 2013 above) cannot be confirmed in the present study.

The results of the study point at the similarities that young learners share in their use of EFL in written production on everyday topics. When young learners in northern Europe who are from different educational and cultural backgrounds communicate in writing on everyday topics they all tend to use common English words to a large extent; some 80% of their vocabulary belong to the 3,000 most common words and this number is considered to be sufficient for these purposes and a level of proficiency of A1 or A2 according to the CEFR. Further research is needed to inform us about this specific variety of EFL in order to identify the specific vocabulary, the syntactic structures and the collocations that are characteristic of young learners' production in EFL.

References

- Artero, J. C. (2013) Writing in an EFL context: an insight into young English learners' writing skills. Paper. Departament de Filologia Anglesa i de Germinística, Grau de Estudis Anglesos, Universitat Autònoma de Barcelona
- Bae, J. & Y.-S. Lee (2012) Evaluating the Development of Children's Writing Ability in an EFL Context. *Language Assessment Quarterly*, 9:4, 348-374
- Barcroft, J. (2004) Second Language Vocabulary Acquisition: A Lexical Processing Input Approach. *Foreign Language Annals* 38 (2) 200-208
- Black, R.W. (2006) Language, culture and identity in online fanfiction. *E-learning*, 3(2), 170-184
- Cameron, L. (2001) *Teaching Languages to Young Learners* Cambridge: Cambridge University Press
- Coady, J., & T.N. Huckin, (1997). *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge: Cambridge University Press
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)* Cambridge: Cambridge University Press. Retrieved on September 26, 2017 from <http://www.coe.int/lang>
- Daller, H. & H. Xue (2007) Lexical Richness and the Oral Proficiency of Chinese EFL Students: A comparison of different measures. In *Modelling and Assessing Vocabulary Knowledge*. Ed. by H. Daller, J. Milton & J. Treffers-Daller. Cambridge: Cambridge University Press
- Ellis, R. (1994) *The Study of Second Language Acquisition* Oxford. Oxford University Press
- Ellis, R. & G. Barkhuizen (2005) *Analysing Learner Language*. Oxford: Oxford University Press
- Enever, J (2011) Chapter 1: Policy. *ELLIE: Early Language Learning in Europe*. Ed by J. Enever The British Council
- Ferrell Tekmen, E. A., & A. Daloglu (2006). An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. *Foreign Language Annals*, 39(2), 220-243.
- Finkbeiner, M. & K.F.J. Nicol & K. Nakamura (2004) The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language*, 51 1-22
- Granger, S. & E. Dagneaux & F. Meunier (2002) *The International Corpus of Learner English. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain
- Griva, E & H. Tsakiridou & J. Nihoritou (2009) "A Study of FL Composing Process and Writing Strategies Employed by Young Learners" in Nikolov, M. *Early Learning of Modern Foreign Languages* United Kingdom: Multilingual Matters 132-148
- Hedge, T. (1988) *Writing* Oxford: Oxford University Press
- Ipek, H. (2009) Comparing and Contrasting First and Second Language Acquisition: Implications for Language Teachers. *English Language Teaching* 2(2), 155-163
- Jiang, N. (2004) Semantic transfer and its implications for vocabulary teaching in a second language. *Modern Language Journal*, 88, 416-432

- Jiménez Catalán, R. M. (2016) Greetings in Letters by EFL Primary School Children: A Longitudinal Study. *Atlantis, Journal of the Association of Anglo-American Studies* Vol 38: 1, 11-32
- Keaveney, S. & Lundberg G. (2014). *Early Language Learning and Teaching A1-A2*. Lund: Studentlitteratur
- Kim, S.-Y. & Y.-S. Ryoo, (2011) Learners' Vocabulary Use in Reading-based Writing: According to Topic and Learner Proficiency. *English Teaching* 2011, Vol 66, Issue 1, 91-109
- Kovaríková, M. (2016) Teaching Writing to Primary School Learners. Bachelor Thesis, Masaryk University, Faculty of Education, Department of English Language and Literature
- Lindgren, E. & C. Munoz, (2013) The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, 10(1), 105-129
- Lindgren, E. & Stevenson M. (2013) Interactional resources in the letters of young writers in Swedish and English. *Journal of Second Language Writing* 22, 390-404
- Lextutor. (2015) "Vocabprofile" Accessed: March 24, 2017. url: <http://www.lexutor.ca/vp/comp/>
- Loder Buechel, L. (2015) Young Learner Writing Performance in Swiss Elementary Schools – Which Teacher Variables Matter? *RELC journal* Vol 46:3, pp.275-292 ISSN: 0033-6882 DOI: 10.1177/0033688215598437
- Longman Dictionaries Online U.S.A. 2017. "Longman Vocabulary Checker"
Accessed: April 12, 2017. url: http://www.longmandictionariesusa.com/vocabulary_checker
- Lundahl, B (2014) *Texts, Topics and Tasks – Teaching English in Years 4-6*. Lund: Studentlitteratur
- Malmberg, P. (ed) (2000) *I huvudet på en elev: projektet STRIMS: strategier vid inläring av moderna språk* Stockholm: Bonnier Utbildning
- Masoura, E.V. & S.E. Gathercole (2005) Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory*, 13, 422-429
- Meara, P. (1996) The dimensions of lexical competence. In: *Performance and Competence in Second Language Acquisition*. Brown, G., Malmkjær, K & J. Williams, 35-53 New York: Cambridge University Press
- Mohebbi, H. (2013) Investigating Vocabulary Learning in Second Language Classroom Context: Recent Findings, Future Outlook. *Advances in Asian Social Studies (AASS)*, Vol 4, No 3, 882-886
- Murphy, V. (2014) *Second Language Learning in the Early School Years: Trends and Contexts*. Oxford: Oxford University Press
- Nation, I.S.P., (1990) *Teaching and Learning Vocabulary* New York: Heinle and Heinle
- Nation, I. S. P. (2011). Research into practice: Vocabulary. *Language Teaching*, 44(4), 529-539.
- Nation, I.S.P., (2006) How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63, 59-81
- Nation, I.S.P. & Waring R (1997) Vocabulary size, text coverage and word lists. In: *Vocabulary: Description, Acquisition and Pedagogy*. McCarthy, M. & Schmitt, N. (eds) 6-19, Cambridge: Cambridge University Press
- Nikolov, M. (2016) *Assessing Young Learners of English: Global and Local Perspectives*. Cham: Springer
- Nunan, D. (2011) *Teaching English to Young Learners* Anaheim: Anaheim University Press
- Orosz A. (2009) The growth of young learners' English vocabulary size . In M. Nikolov (ed) *Early Learning of Modern Foreign Languages: Processes and Outcomes*. Bristol: Multilingual Matters
- Ortega, L. (2009) *Understanding Second Language Acquisition*. London: Hodder Education
- Pinter, A.M. (2006) *Teaching Young Language Learners* Oxford. Oxford University Press
- Schmitt, N. (2000) *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press
- Schmitt, N. (2010) *Researching Vocabulary: a Vocabulary Research Manual* London: Palgrave MacMillan
- Scrivener, J. (2005) *Learning Teaching: A Guidebook for Language Teachers* Oxford: Macmillan
- Service, E. & V. Kohonen (1995) Is the relation between phonological memory and foreign language learning accounted for by acquisition? *Applied Psycholinguistics*, 16, 155-172
- Sundh, S. (2016) A Corpus of Young Learners' English in the Baltic Region - Texts for Studies on Sustainable Development *Discourse and Communication for Sustainable Education* Vol 7, no. 2, 92-104, doi: 10.1515/dcse-2016-0018
- Sundqvist, P. (2009) *Extramural English matters: out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary*. Karlstad: Karlstad University Studies
- Zimmerman, C. (1997). Do Reading and Interactive Vocabulary Instruction Make a Difference? An Empirical Study. *TESOL Quarterly*, 31(1), 121-140.