

The Impact of Free-time Reading on Foreign Language Vocabulary Development

Päivi Pietilä

Department of English, University of Turku, Turku, Finland

Riika Merikivi

Department of English, University of Turku, Turku, Finland

Abstract—The study reported in this article was conducted to investigate the influence of free-time reading on foreign language vocabulary development. The subjects were sixth and ninth graders in content and language integrated learning (CLIL) and in regular mainstream classes, with Finnish as their L1 and English as their L2. The vocabulary sizes of the learners were measured with the Vocabulary Levels Test and the Productive Vocabulary Levels Test. Students who reported reading in English in their free time turned out to have larger vocabularies, both receptive and productive, than those who did not. The influence of reading on L2 vocabulary skills was particularly strong in the case of learners who read on a daily basis. Participants in the CLIL programmes had larger vocabularies than their mainstream peers, which can be seen as a result of the effective learning environment but also of the fact that they read substantially more than pupils in regular classes.

Index Terms—vocabulary size, free-time reading, L2 English, CLIL, language learning

I. INTRODUCTION

The relationship between foreign language vocabulary knowledge and reading is at least twofold and reciprocal: on the one hand, vocabulary knowledge has an effect on reading comprehension, and on the other hand, reading undoubtedly influences vocabulary knowledge and growth. It has been established (e.g. Laufer, 1997) that vocabulary is the most important factor in reading comprehension in general, far more important than syntax or general reading strategies. The influence of reading on lexical knowledge, the focus of this research report, has also been demonstrated in several studies (e.g. Coady, 1997; Wesche & Paribakht, 1999). As Gass (1999) states, there are still a number of unsettled questions related to vocabulary acquisition, although there is general consensus about the fact that at least some vocabulary is acquired *incidentally*, i.e. as a by-product of some other cognitive activity, especially reading.

Studies examining the influence of reading on the acquisition of vocabulary have mainly concentrated on finding out whether such an influence exists in the first place (e.g. Coady, 1997), or contemplating how many exposures to a word are necessary for successful acquisition (see Nation's 1990 survey of this issue). Some studies (e.g. Waring & Takaki, 2003; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006) have also considered the incremental nature of vocabulary learning and investigated the emergence of different aspects of lexical knowledge. It seems unquestionable that there are a great number of variables involved in the acquisition process, including various features of the word itself, not to mention individual characteristics of the learner. The study presented in this article differs from those conducted earlier in that it sets out to investigate the impact of the amount of free-time reading on foreign language vocabulary, measured in terms of productive and receptive vocabulary size. The subjects of the study represent learners from two different learning environments, the general mainstream education (GEN) and CLIL (content and language integrated learning).

II. LEARNING THROUGH READING

Nation and Coady (1988) explain that learning words by reading requires, first, that a sufficient number of the vocabulary in a text are familiar in order for understanding and contextual guessing to be possible and, second, that there are at least some unfamiliar words to be learnt. Unsurprisingly, if the contextual cues are unusable or non-existent, a reader cannot guess the meanings of unfamiliar items, and misunderstandings are probable (cf. Laufer, 1997). If, on the contrary, the context is too rich, the reader does not have to dwell on the meanings as he can nevertheless reach a satisfying level of understanding when it comes to the whole text. That is to say, the proportion of both unfamiliar items and cues needs to be adequate but not excessive for acquisition to occur. According to Nation (2001), a very high ratio of familiar words in a text, i.e. 99 to 100%, strengthens the existing vocabulary, whereas growth is still possible at lower levels of coverage, i.e. 95 to 98%.

It is not settled how much exposure to adequate texts is needed for learning to take place (cf. Huckin & Coady, 1999) but, basically, acquisition through reading is considered fragile and slow. Each meeting with a word typically involves only small gains: the estimates of the proportion of unfamiliar words learnt when encountered in a text vary between

20% and 5% (Nagy, 1997; Horst, Cobb & Meara, 1998), which means that only one unfamiliar item in every five to one unfamiliar item in every twenty may be learnt to some extent when reading. Nation (2001) explains that, for instance, the initial exposure to a new word may only give rise to a vague knowledge of its form and awareness of its novelty. Moreover, if the knowledge is not reinforced soon enough by another meeting, the learning will in all likelihood be lost.

Advanced learners are, nonetheless, indisputably found to profit from meaning focused reading (see e.g. Krashen, 1989; N. Ellis, 1994; Coady, 1997; Wesche & Paribakht, 1999) in that their vocabulary becomes wider. It appears that small gains may become large gains if a learner does a multitude of reading. Nagy's (1997) train of thought is very interesting: if one encountered a million tokens a year, and if at least two per cent of these tokens were unfamiliar, it would mean meeting 20 000 unfamiliar words per year. Moreover, if at least one in twenty of them is acquired (cf. above), the annual gain would be 1000 more or less learnt words, which corresponds to a native-like rate. One million running words is roughly equivalent to ten to twelve novels (Kennedy, 1987), or approximately 65 graded readers (Nation & Waring, 1999), or three to four undergraduate textbooks (Sutarsyah, Nation & Kennedy, 1994).

As pointed out earlier, the processes of lexical learning in general are as yet rather ill-understood. Consequently, it is not fully known how learning through extensive reading takes place either (cf. Gass, 1999). The presumption of furthering vocabulary growth through rich exposure to written texts is essentially based on input-oriented language acquisition theories which assume that development occurs when learners are processing meaningful input and involve themselves in inferring word meanings from context (e.g. Lee & VanPatten, 2003). Learning of this kind is often referred to as incidental or implicit because it appears to take place while learners are focused on something other than word acquisition itself, that is, a new item seems to be learnt without intent to do so.

Krashen (1989) is one of the firmest advocates of extensive reading, for he states that the majority of vocabulary knowledge is learnt automatically simply by receiving enough comprehensible input in the form of printed materials, entailing that reading a great deal of adequate texts is both crucial and sufficient for acquiring a rich lexicon. However, as R. Ellis (1994) remarks, it may not be comprehensible input but rather comprehended input that is central. Lawson and Hogben (1996) take a yet more radical stand in arguing that even successful understanding may not be the same as successful acquisition, i.e. comprehension alone only seldom appears to translate into learning.

Research conducted by Paribakht and Wesche (1993; 1997; 1999) provides evidence for the above position. They found that many learners simply ignored unrecognised words if they could grasp the message of the whole text without understanding them. In other words, it seems that learners do not necessarily pick up new vocabulary automatically when engaged in reading activities. More recently, similar findings have been achieved by Waring and Takaki (2003), whose subjects learned some new words by reading a graded reader but forgot most of them, as shown in a post-test one week later and especially in a second post-test administered three months after the reading experiment. This led the researchers to conclude that the number of occurrences required for a word to be retained may well be much higher than estimated so far: even after more than 18 encounters, there was only a 10 to 15% chance of the word meaning to be remembered. Waring and Takaki (2003) used different tests to tap different aspects of word knowledge (word-form recognition, prompted recognition of meaning, and unprompted recognition of meaning). According to their results, recognition of word-forms was the easiest part of word knowledge, whereas meaning appeared to be more difficult. Pellicer-Sánchez and Schmitt (2010), in contrast, found meaning recognition to be the best learned aspect of word knowledge in a study where the subjects read an authentic novel and were tested immediately afterwards. However, these two studies are not easily comparable, as they had very few similarities. What is more, the words tested were rather unusual in both: Waring and Takaki (2003) had substituted each test word with a non-existing but English-looking word, whereas Pellicer-Sánchez and Schmitt (2010) used only African words from their otherwise English novel.

A number of studies have indeed shown that second language learners do learn vocabulary, albeit slowly and in small quantities, incidentally, through extensive reading. It has also been proposed (e.g. by Sonbul and Schmitt, 2009) that the most effective method is a combination of incidental and explicit learning. Incidental learning may occur as a by-product of any language-related activity, for example reading, but learning would be enhanced by explicit vocabulary instruction. In summary, it seems evident that vocabulary knowledge comprises countless features and that learners do not immediately learn all of them but, rather, may incrementally add new properties until a mature entry is achieved, if it ever is achieved. Extended reading of adequate texts has been found to promote lexical learning particularly well, although the acquisition process itself is not clear as yet. Contrary to input-oriented learning theories, both conscious and unconscious procedures may be present and enhance all L2 word acquisition, whether incidental or intentional.

III. VOCABULARY SIZE

The question of *vocabulary size* is directly linked to the ability of a language learner to read and understand different types of texts. It has been estimated that a text cannot be comprehended even for gist if less than 80% of the running words are familiar (Nation, 2001; Milton, 2006). According to Takala (1989), 90% familiarity is required to understand the main ideas (one unfamiliar word in every ten words), and 95 % familiarity to comprehend the details of a text (one unfamiliar word in every twenty words). Laufer (1989) also sets the limit of adequate comprehension at 95%. However, Hu and Nation (2000) estimate that even one unknown token in every twenty is too much. They compared the effect of different densities of unfamiliar items and measured comprehension. At the 95% coverage level, some of the partici-

pants gained adequate understanding but most did not. For sufficient comprehension to be likely to occur, they arrived at suggesting a density as low as one unknown token in every fifty, meaning that learners would need to know around 98% of the running words in a text. This is in line with Nation and Coady (1988), who approximate that unfamiliar words should not cover more than 2 % of any text for the reader to be able to reach sufficient understanding.

The relationship between vocabulary size and text coverage (or the proportion of familiar words in a text) is affected by text type. With the 98% familiarity condition, youth novels would require the knowledge of around 5000 words (Hirsch & Nation, 1992), adult fiction as many as 8000 to 9000 word families, and newspaper articles around 8000 word families (Nation, 2006), to be understood. If, however, the 95% coverage reported by Laufer is adequate, substantially smaller lexicons are sufficient for comprehension. A vocabulary of the 2000 most frequent words would be enough for reading youth fiction and 4 000 for comprehending newspapers and adult fiction. As a rule of thumb, Laufer (1997) reports that one needs a minimum vocabulary of around 3000 word families to gain 90 to 95% coverage of any text, which she considers a general threshold for reading unsimplified texts.

As for vocabulary sizes of language learners, they are generally substantially smaller than those of native speakers. According to Nation & Waring (1997), native speakers of English typically learn about 1000 word families a year until they reach about 20000 word families as young adults. At the ages of thirteen and sixteen, the ages of the subjects in the present study, they would presumably know well over 10000 and round 15000 to 16000 word families, respectively. After six years of study, learners of English, however, have been found to know receptively about 3500-4000 words and productively about 2600 words (Schmitt & Meara, 1997; Laufer, 1998).

IV. CLIL VS. MAINSTREAM

Teaching and learning through a foreign or second language is not a novel phenomenon; on the contrary, it has occurred throughout the history of education. Interest in the matter has recently increased considerably, which may at least partly be explained by the current societal situation. The on-going European integration and internationalisation in general continuously lay more and more demanding objectives on learning and teaching additional languages. Content and language integrated learning (CLIL) attracts educators and researchers, for it may be an ideal environment for achieving good proficiency in an L2.

CLIL programmes vary a great deal but nonetheless they share a number of key features. CLIL invariably involves dual focus in that the basic objective of any type of CLIL is to integrate the learning of a non-language subject matter with the acquisition of a language other than the learners' L1. Further, all CLIL programmes explicitly focus on the development of learners' thinking skills (Räsänen, 1994; Sjöholm, 1999), entailing the development of the learning strategies and cognitive skills necessary for academic progress on all levels of education. In the light of the latter point, CLIL and GEN education do not differ much in the Finnish context, as objectives of a similar kind are written down in the mainstream curriculum as well (Finnish National Board of Education, 2004).

As for learning results in CLIL, compared with mainstream classes, studies that have investigated the effects of CLIL on the learners' mother tongue skills (a severe concern for some parents and educators) have shown that CLIL does not have a negative impact on the development of L1, some actually suggesting that the situation may be quite the opposite. As an example, Merisuo-Storm (2002) examined pupils' L1 skills in reading and writing in elementary instruction and found that CLIL learners generally outperformed their peers attending the monolingual stream. Another concern has been related to the acquisition of content knowledge. According to the results of Seikkula-Leino (2007), there were no marked differences between classes taught in their L1 Finnish and those taught in L2 English. Similarly, Xanthou (2011) found that both groups, CLIL (with English as the language of instruction) and mainstream (Cypriot Greek), gained equal amounts of content knowledge in science classes, in addition to which the CLIL pupils demonstrated a marked increase in L2 vocabulary.

Also research concentrating on the development of L2 skills in CLIL has produced encouraging results. Järvinen (2005; see also 1999) tracked the syntactic development of relativisation, as measured by elicited imitation and grammaticality judgements, in CLIL and mainstream education during elementary school in Finland. Her results suggest that foreign language acquisition is substantially faster and more versatile in CLIL than in general classes in that the CLIL group could produce sentences significantly longer, more complex and more accurate than the control group, in the elicited imitation task. Valtanen (2001) investigated overall English proficiency at the end of lower secondary school and measured English skills in reading, listening, writing, speaking, and grammar. CLIL learners scored on average higher than their mainstream peers in each section of the test. The results imply that CLIL has a clear positive effect on the development of English competence as a whole. In fact, the CLIL learners were so advanced that a ceiling effect was found in most of the tasks, that is, the level of their proficiency could not actually be discovered with the national test battery aimed at the regular ninth-graders.

Karonen's (2003) study is one of the few aimed at investigating vocabulary learning in CLIL. She explored lexical organisation by conducting a word association test on both CLIL students and students following the mainstream curriculum. She hypothesised that CLIL learners' lexicons would be more organised in nature and that they would give more paradigmatic responses typical of native speaker adults, whereas the learners in formal language instruction were expected to give more syntagmatic responses typical of native speaker children. However, no marked differences were found between the classes in this respect. That is to say, no developmental syntagmatic-paradigmatic shift was found in

either of the groups. One reason might be that the learners were simply too young (5th, 7th, and 9th graders) for the organisation dimension of their lexicons to have developed sufficiently. The case might be different with more advanced learners.

Another small-scale study on CLIL pupils' vocabulary knowledge in the Finnish comprehensive school is Viinikkala (2006). In that study, the development of passive and active vocabularies was investigated with Nation's Levels Tests (see the subchapter on Methods and Procedures) and lexical richness in the students' free writing was evaluated. As with previous research on reception and production, the two vocabulary systems turned out to be strongly related to one another, i.e. learners with wider passive vocabularies had also wider active vocabularies. Receptive knowledge also appeared to develop the fastest and lexical richness the slowest.

In conclusion, research has indicated that learning in a CLIL environment does not typically prevent students from learning content, nor does it influence negatively the development of their mother tongue proficiency. As for TL development, CLIL seems to have a more favourable effect on pupils' L2 skills than the monolingual stream, although it is not yet clear exactly how and why this is. As for vocabulary acquisition, CLIL may in fact be ideal, as it combines explicit and implicit learning conditions (see e.g. R. Ellis, 1990). In addition to formal instruction, the pupils are constantly exposed to authentic input and have a multitude of opportunities to communicate in the target language.

V. THE STUDY

A. *The Aim of the Study*

The purpose of the study was to examine the impact of free-time reading on foreign language (L2 English) vocabulary development. As the subjects came from two different learning environments, content and language integrated learning (CLIL) and general mainstream education (GEN), the study was also intended to compare the learning outcomes of these two groups of learners and thus shed light on the effectiveness of these two learning contexts.

B. *Subjects*

The subjects of the study were Finnish comprehensive school pupils, aged 13 to 16. At the onset, 367 pupils took the vocabulary tests, but the number was reduced to 330 on the basis of a background questionnaire. To ensure the homogeneity of the subjects, only those whose native language was Finnish, who did not speak English at home, and who had not lived in an English-speaking country for more than two weeks, were chosen. 150 of the subjects were in 6th grade and thus finishing primary school, and the rest, 180, were in 9th grade, i.e. finishing lower secondary school. Of the 6th-graders, 75 were in CLIL classes (CLIL6) and 74 in mainstream instruction (GEN6), whereas the 9th-graders consisted of 88 CLIL students (CLIL9) and 93 mainstream students (GEN9). The subjects came from altogether 17 classes from 11 different schools in southwest Finland.

It is worth noting that in Finland, pupils start a CLIL programme without any knowledge of the foreign language. They take an aptitude test which measures school readiness and skills in their native language, so in a sense the CLIL students may represent a group with certain academic propensities, but this does not automatically mean that they have an upper hand compared with their mainstream peers, most of whom were never tested and might well have obtained similar results if they had taken the test. We are aware that studies involving CLIL classes are often criticised because of the allegedly unfair initial situation of the mainstream groups, but we would like to emphasise the fact that many pupils opt for non-CLIL education, not because they feel they would not be able to follow a bilingual programme but simply because they want to follow the regular curriculum.

Attending a CLIL class nevertheless inevitably differentiates pupils from their mainstream peers in various ways. The sheer amount of foreign language input is, by definition, greater in CLIL instruction, and, at least in Finland, CLIL pupils start being exposed to the foreign language, albeit in small doses, from the beginning of the first grade, whereas formal foreign language teaching in regular classes does not start until the third grade. Therefore, the CLIL subjects of the present study had been exposed to English at school two years longer than the mainstream subjects.

C. *Methods and Procedures*

Two vocabulary tests were used to investigate the vocabulary size of the subjects. Receptive vocabulary knowledge was measured with the Vocabulary Levels Test, originally designed by Nation (1983). The initial test has later been replaced by improved versions 1 and 2 by Schmitt, Schmitt and Clapham (2001). This study used Version 1 (see Schmitt, 2000). Productive vocabulary knowledge was measured with the Productive Vocabulary Levels Test developed by Laufer & Nation (1999). The test has two parallel versions; this study exploited Version 2. Both tests have been widely used and validated by various scholars (e.g. Schmitt, Schmitt & Clapham, 2001; Laufer & Nation, 1999; Beglar & Hunt, 1999).

The scores obtained in the vocabulary tests do not directly correspond to a certain vocabulary size; rather, they indicate the proportion of words a learner knows at the various frequency levels. Therefore, the size of the learner's lexicon had to be calculated separately. Our formula was largely the same as that used by Laufer (1998).

A questionnaire was used to find out about the subjects' use of English during their free time, the main focus being on whether or not they read English language texts outside school hours. The participants were asked to announce, first of all, if they read English language books, magazines or other texts outside class time at all. Furthermore, they were

asked to specify how often they were involved in these activities in English, the options being *monthly*, *a few times per week*, *many times per week* and *daily*.

A self-report is not necessarily the most reliable way to collect data as, for instance, the subjects may give disinformation and as there is always room for understanding the questions differently from the intended aim. However, the participants were encouraged to answer as truthfully as they possibly could. The testing was carried out under normal classroom circumstances during schooldays on two different occasions, both lasting 45 minutes. On the first occasion, the participants filled in the questionnaire and took the receptive test, and on the second occasion they took the productive test.

VI. RESULTS AND DISCUSSION

The participants' receptive and productive vocabulary sizes were measured using the Vocabulary Levels Test and the Productive Vocabulary Levels Test, as stated above. A more thorough account of the vocabularies, in terms of sizes, the productive-receptive ratios, and lexical frequency levels can be found in another, forthcoming, article by the same authors. The subjects' receptive and productive vocabulary sizes can be seen in Table 1 (sixth-graders) and Table 2 (ninth-graders). The figures represent word families.

TABLE 1.
SIXTH-GRADERS' RECEPTIVE AND PRODUCTIVE VOCABULARY SIZES

	GEN6 (N=74)	CLIL6 (N=75)	t	df	p
Receptive voc.	1853	4505	11.39	116	<0.001
Productive voc.	788	2271	9.22	102	<0.001

As Table 1 indicates, the CLIL6 pupils had statistically larger vocabularies, both receptive and productive, than the GEN6 group. In other words, the CLIL environment seems to have produced better learning results than general education.

TABLE 2.
NINTH-GRADERS' RECEPTIVE AND PRODUCTIVE VOCABULARY SIZES

	GEN9 (N=93)	CLIL9 (N=88)	t	df	p
Receptive voc.	5161	6379	4.82	179	<0.001
Productive voc.	2565	3742	5.26	179	<0.001

Similarly to the 6th-graders, significant differences were found in the receptive and productive vocabulary sizes of the 9th-graders, with the CLIL learners outperforming the mainstream group. Taken as a whole, the vocabulary sizes apparent in this study are in line with earlier studies on vocabulary sizes of language learners (cf. Nation & Waring, 1997; Laufer, 1998). However, the passive vocabularies elicited in this study are larger than those discovered in the above-mentioned studies, i.e. the active-passive ratio is lower. This would seem to suggest that the Finnish educational system strongly encourages passive vocabulary learning but fails to activate items as effectively, an observation supported by Järvinen's (1999) findings.

As regards the differences between the CLIL and GEN groups, these results are also in accordance with many earlier studies, implying that CLIL as a learning environment is more effective and gives better opportunities also for lexical development than the traditional classroom. However, children and young people of today, at least in countries like Finland, are in constant contact with the English language also outside the classroom. It is therefore likely that they also acquire English vocabulary incidentally when engaged in free-time activities. The rest of the results reported here are based on the information gathered with the questionnaire that the subjects filled in regarding their encounters with English in their free-time activities, particularly reading. The question the subjects had to answer about their reading habits in general was the following:

In your free time, do you read English books, magazines or other texts?

If the answer was positive, the testees were asked to specify how often they read in English on their own, the options being *monthly*, *a few times per week*, *many times per week* and *daily*. When analysing the results, the option *many times per week* was combined with *a few times per week* to constitute *weekly*, for it was selected only rarely.

About 60% (N=196) of the participants reported that they read English texts outside class time and about 40% (N=134) stated that they do not. The CLIL9 learners announced reading the most (70% of them) and the GEN6 learners the least (40% of them). CLIL6 and GEN9 groups were identical to one another in that the proportion of readers was sixty per cent in both of them. Both readers' and non-readers' receptive and productive vocabulary sizes and the ratios between the vocabularies are displayed in Table 3. The differences between the two groups of learners are presented in terms of t-test results.

TABLE 3.

DIFFERENCE BETWEEN READERS AND NON-READERS IN RECEPTIVE AND PRODUCTIVE VOCABULARY SIZE AND IN PRODUCTIVE-RECEPTIVE RATIO

	Readers (N = 196)	Non-readers (N = 134)	Difference		
	Vocabulary size		t	df	p
Receptive	5275	3601	7.07	328	<0.001
Productive	2923	1659	7.83	326	<0.001
Ratio (%)	52.9	44.4	5.04	328	<0.001

Table 3 shows that pupils who reported reading in English in their free time scored significantly better in the vocabulary tests than those pupils who reported not reading. Also the active-passive ratio is higher for them, suggesting that they could use a larger proportion of their passively known words in production.

Of those 196 participants who read English texts in their free time, 34% announced reading a few times a month (N=67; 20% of the entire sample), 40% a few times a week (N=78; 24% of the entire sample), and 26% every day (N=51; 16% of the entire sample). These three options were chosen surprisingly evenly within the CLIL6, GEN9 and CLIL9 groups, i.e. each option was selected by around 30 per cent of the readers. Of the GEN6 readers, on the other hand, approximately 10 per cent reported reading daily, 50 per cent weekly and 40 per cent monthly. Table 4 presents the readers' passive and active vocabulary sizes as well as the proportion of actively known items, according to the reported frequency of reading.

TABLE 4.

RECEPTIVE AND PRODUCTIVE VOCABULARY SIZE AND PRODUCTIVE –RECEPTIVE RATIO BY FREQUENCY OF READING

Frequency of reading	N	Receptive vocabulary	Productive vocabulary	Ratio (%)
Monthly	67 (34 %)	4613	2390	49
Weekly	78 (40 %)	5023	2717	53
Daily	51 (26 %)	6527	3965	58

Table 4 implies that vocabulary knowledge is the greater the more often one reads in the target language. Moreover, the proportion of active words seems to increase the more often one reads. The possible significance of these differences was investigated by a one-way ANOVA, which indicated that the frequency of reading activity significantly affected receptive and productive vocabulary knowledge ($F=13.8$, $p<0.001$ and $F=15.5$, $p<0.001$, respectively) and the relationship between the vocabularies ($F=5.7$, $p=0.004$). The post hoc Scheffe test results are displayed in Table 5.

TABLE 5.

DIFFERENCE IN RECEPTIVE AND PRODUCTIVE VOCABULARY SIZE AND IN PRODUCTIVE-RECEPTIVE RATIO BY FREQUENCY OF READING

	Receptive vocabulary		Productive vocabulary		Ratio	
Frequency of reading	Mean difference	p	Mean difference	p	Mean difference	p
Monthly – Weekly	-410	0.482	-327	0.465	-3.8	0.275
Weekly – Daily	-1504	<0.001	-1248	<0.001	-5.2	0.137
Monthly – Daily	-1914	<0.001	-1574	<0.001	-9	0.004

Table 5 indicates that the frequency of reading has a significant effect on vocabulary size especially if one reads on a daily basis. The same is true for the active-passive ratio, i.e. among those who read, vocabulary seems to become active significantly better if one reads every day. Reading weekly or monthly does not seem to be as beneficial.

In sum, what learners do in their free time appears to have a significant effect on their target language word proficiency, as those pupils who read in English on their own scored significantly higher than those who reported not to read. Moreover, the more often they reported reading in English outside class time, the more vocabulary they knew both receptively and productively. More precisely, those who read on a daily basis turned out to be lexically stronger than those who read more infrequently.

The observation about the benefits of reading large quantities of text is in line with previous research stating that learning from reading is extremely fragile and slow (see e.g. Nagy, 1997; Horst, Cobb & Meara, 1998). As discussed in Section II, each meeting with a word may typically involve only small gains and the learning will in all likelihood be lost if not reinforced soon enough by another meeting (cf. Nation, 2001), which is probably why the frequency of exposure was found to make such a decisive difference in the results of our participants also. On the other hand, as some researchers (see e.g. N. Ellis, 1997; Paribakht & Wesche, 1997; Carter, 2001) suggest, lexical learning could probably be advanced by deeper cognition, that is to say, by learners paying attention to individual items and investing conscious effort into concluding their meaning. In addition to reading more and more, learners may profit from being taught different inferring and memorising strategies, as well as other explicit ways to commit word knowledge into memory while reading.

It appears, then, that learners who receive English input in the form of written texts outside school are those who acquire the largest lexicons. Then again, the streams were not equal in that the CLIL pupils reported reading to a larger extent than their GEN peers, possibly suggesting that a CLIL context itself may encourage different reading habits than

the general context. By the end of elementary school, sixty per cent of the CLIL pupils and forty per cent of the GEN pupils had started to read in English on their own, which is a striking difference. By the end of lower secondary school, the corresponding proportions are seventy per cent for CLIL and sixty per cent for GEN, the CLIL pupils thus taking to reading English texts outside of school more frequently than the GEN pupils during the comprehensive school. Taken the evidence in support of the power of reading, we believe that the clear difference in the reading habits between the two learning environments may be one of the main reasons for CLIL having yielded greater vocabulary sizes (cf. Sylv  n, 2010). Of course, our conclusions would be based on more solid grounds if we had been able to compare the subgroups in more detail and control for the differences in type and amount of input and in some other factors which differentiate CLIL and GEN classes.

Finally, as the majority of the participants claimed to read outside the school environment, it is interesting to view how well their vocabularies support the reading of distinct types of English language texts. Based on the idea that one needs to know 98 per cent of the vocabulary in a text to be able to comprehend it and to guess the meanings of unfamiliar words from context, Nation (2006) estimates that understanding newspapers and adult fiction would require a vocabulary of 8000 to 9000 words. In the present study, only around twenty CLIL9 pupils and a handful of GEN9 pupils knew this much vocabulary, suggesting that most of the readers would have considerable difficulties in trying to comprehend newspapers or novels. On the other hand, Hirsh and Nation (1992) report that comprehending youth novels necessitates a lexicon of only 5 000 items, meaning that the average ninth-grader would be able to read them without major difficulty. The average CLIL6 pupil was familiar with some 4500 word families, which provides fairly good readiness for reading youth fiction.

On the other hand, Laufer (1989) concludes that knowing 95 per cent of the running words in a text suffices for adequate understanding, entailing that a vocabulary of 4 000 words would be enough for reading newspapers as well as adult fiction (cf. Nation, 2006) and a vocabulary of only 2 000 words for reading youth novels (cf. Hirsh & Nation, 1992). This holds that the three strongest groups would be able to read all of these text types, provided they can tolerate a level of ambiguity. As regards the weakest group, GEN6 learners' knowledge does not directly suffice to read any of the above-mentioned text types, although they knew almost enough vocabulary to read youth fiction. As forty per cent of them nonetheless reported reading, they are probably able to comprehend the simplest texts to a satisfying degree. Free-time reading does not necessarily require the same level of understanding as, for instance, reading for studying, free-time readers conceivably standing relatively well the disturbance that arises from not being able to infer the meaning of each and every unfamiliar item.

To conclude, what the pupils do in their spare time appears to be crucial as regards their success rate in mastering vocabulary. Frequent target language use in general and reading target language texts outside of schoolwork in particular may be decisive in that the pupils who announced reading on their own were those who scored the best, regardless of group. That being said, the CLIL learners reported reading to a larger extent than their control peers, the CLIL context conceivably encouraging different reading habits than the monolingual environment.

VII. CONCLUSION

Our results implied that the Finnish comprehensive school can provide a good setting for foreign language vocabulary acquisition in that the pupils were shown to learn receptively an average of around 700 word families and productively about 400 word families per year during their comprehensive education, estimations which are in line with previous research on English vocabulary learning in a foreign language context. It does, however, seem to matter whether a pupil attends a CLIL class or a traditional class as far as the overall size of his/her English lexicon is concerned. In the sixth grade, the average CLIL participant knew receptively some 4500 and productively around 2300 word families, the corresponding figures for the mainstream participants being only 1800 and 800, respectively. In the ninth grade, the CLIL pupils could on average comprehend about 6400 and produce around 3700 word families, the GEN pupils understanding approximately 5200 and being able to produce around 2600 words.

It was not possible to attribute the positive receptive results of the CLIL participants to the CLIL method alone as, in the light of our findings, the CLIL learners may have understood more words partly because they had started their English education two years earlier than the mainstream learners, giving them more time to accumulate lexical knowledge. On the other hand, the extra English input to which all of the learners had been exposed outside school hours had possibly ironed out the probable differences in the word learning opportunities the GEN and CLIL pupils have at school. Reading target language texts, especially, was proven to be decisive and the participants who announced reading on their own were those who scored the best, regardless of group. That being said, the CLIL learners reported reading to a larger extent than their control peers, the CLIL context conceivably encouraging different reading habits than the monolingual environment.

ACKNOWLEDGEMENT

The authors wish to thank the anonymous reviewers for their precise and insightful comments.

REFERENCES

- [1] Beglar, D. & A. Hunt. (1999). Revising and validating the 2000 Word Level and University Word Level Vocabulary Tests. *Language Testing* 16.2, 131-162.
- [2] Carter, R. (2001). Vocabulary. In R. Carter & D. Nunan (eds.), *The Cambridge guide to teaching English to speakers of other languages*. Cambridge: Cambridge University Press, 31-57.
- [3] Coady, J. (1997). L2 vocabulary acquisition through extensive reading. In J. Coady & T. Huckin (eds.), *Second language vocabulary acquisition*. Cambridge: Cambridge University Press, 225-237.
- [4] Ellis, N. (1994). Vocabulary acquisition: The implicit Ins and Outs of explicit cognitive mediation. In N. Ellis (ed.), *Implicit and explicit learning of languages*. London: Academic Press, 211-282.
- [5] Ellis, N. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In N. Schmitt & M. McCarthy (eds.), *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press, 122-139.
- [6] Ellis, R. (1990). Instructed second language acquisition. Oxford: Basil Blackwell.
- [7] Ellis, R. (1994). Factors in the incidental acquisition of second language vocabulary for oral input: A review essay. *Applied Language Learning* 5, 1-32.
- [8] Finnish National Board of Education. (2004). Perusopetuksen opetussuunnitelman perusteet (The national core curriculum for basic education). Helsinki: Opetushallitus.
- [9] Gass, S. (1999). Incidental vocabulary learning. *Studies in Second Language Acquisition* 21.2, 319-333.
- [10] Hirsh, D. & I. S. P. Nation. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language* 8.2, 689-696.
- [11] Horst, M., T. Cobb & P. Meara. (1998). Beyond a Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language* 11.2, 207-223.
- [12] Hu, M. & I. S. P. Nation. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language* 13.1, 403-430.
- [13] Huckin, T. & J. Coady. (1999). Incidental vocabulary acquisition in a second language. A review. *Studies in Second Language Acquisition* 21.2, 181-193.
- [14] Järvinen, H.-M. (1999). Acquisition of English in content and language integrated learning at elementary level in the Finnish comprehensive school. Turku: University of Turku.
- [15] Järvinen, H.-M. (2005). Language learning in content-based instruction. In A. Housen & M. Pierrard (eds.), *Investigations in instructed second language acquisition*. Berlin: Mouton de Gruyter, 433-456.
- [16] Karonen, S. (2003). Semantic organization of L2 learners' mental lexicon: comparing pupils in mainstream education and CLIL classes. Unpublished MA Thesis. University of Turku.
- [17] Kennedy, G. (1987). Expressing temporal frequency in academic English. *TESOL Quarterly* 21.1, 69-86.
- [18] Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal* 73. 4, 440-464.
- [19] Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Laurén & M. Nordman (eds.), *Special language: from humans thinking to thinking machines*. Clevedon: Multilingual Matters, 316-323.
- [20] Laufer, B. (1997). The lexical plight in second language reading. Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (eds.), *Second language vocabulary acquisition*. Cambridge: Cambridge University Press, 20-34.
- [21] Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics* 19, 255-271.
- [22] Laufer, B. & I. S. P. Nation. (1999). A vocabulary-size test of controlled productive ability. *Language Testing* 16.1, 33-51.
- [23] Lawson, M. & D. Hogben. (1996). The vocabulary-learning strategies of foreign-language students. *Language Learning* 46.1, 101-135.
- [24] Lee, J. F. & B. VanPatten. (2003). Making communicative language teaching happen. Boston: McGraw-Hill.
- [25] Merisuo-Storm, T. (2002). Oppilaan äidinkielen lukemisen ja kirjoittamisen taitojen kehittyminen kaksikielisessä alkuopetuksessa (The development of L1 reading and writing skills in bilingual elementary instruction). Turku: University of Turku.
- [26] Milton, J. (2006). Language lite? Learning French vocabulary in school. *French Language Studies* 16, 187-205.
- [27] Nagy, W. (1997). On the role of context in first- and second-language vocabulary learning. In N. Schmitt & M. McCarthy (eds.), *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge University Press, 64-83.
- [28] Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines* 5, 12-25.
- [29] Nation, I. S. P. (1990). Teaching and learning vocabulary. Rowley, MA: Newbury House.
- [30] Nation, I. S. P. (2001). Learning vocabulary in another language. Cambridge: Cambridge University Press.
- [31] Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review* 63.1, 59-82.
- [32] Nation, I. S. P. & J. Coady. (1988). Vocabulary and reading. In R. Carter & M. McCarthy (eds.), *Vocabulary and language teaching*. New York: Longman Inc., 97-110.
- [33] Nation, I. S. P. & K. Waring. (1999). Graded readers and vocabulary. *Reading in a Foreign Language* 12.2, 355-380.
- [34] Paribakht, T. S. & M. Wesche. (1993). The relationship between reading comprehension and second language development in a comprehension based ESL program. *TESL Canada Journal* 11.1, 9-29.
- [35] Paribakht, T. S. & M. Wesche. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin. (eds.), *Second language vocabulary acquisition*. Cambridge: Cambridge University Press, 174-200.
- [36] Paribakht, T. S. & M. Wesche. (1999). Reading and "incidental" L2 vocabulary acquisition. An introspective study of lexical inferencing. *Studies in Second Language Acquisition* 21.2, 195-224.

- [37] Pellicer-Sánchez, A. & N. Schmitt. (2010). Incidental vocabulary acquisition from an authentic novel: Do *Things Fall Apart?* *Reading in a Foreign Language* 22.1, 31-55.
- [38] Pigada, M. & N. Schmitt. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language* 18.1, 1-28.
- [39] Räsänen, A. (1994). Language learning and bilingual education: Theoretical considerations and practical experiences. In A. Räsänen & D. Marsh (eds.), *Content instruction through a foreign language*. Jyväskylä University of Jyväskylä 9-31.
- [40] Schmitt, N. (2000). Vocabulary in language teaching. Cambridge: Cambridge University Press.
- [41] Schmitt, N. & P. Meara. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition* 19.1, 17-35.
- [42] Schmitt, N., D. Schmitt & C. Clapham. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing* 18.1, 55-88.
- [43] Seikkula-Leino, J. (2007). CLIL learning. Achievement levels and affective factors. *Language and Education* 21.4, 328-341.
- [44] Sjöholm, K. (1999). Education in a second or foreign language: An overview. In K. Sjöholm & M. Björklund (eds.), *Content and language integrated learning: Teachers' and teacher educators' experiences of English medium teaching*. Vaasa: Åbo Akademi University, 15-39.
- [45] Sonbul, S. & N. Schmitt. (2009). Direct teaching of vocabulary after reading: is it worth the effort? *ELT Journal* 64.3, 253-260.
- [46] Sutarsyah, C., I. S. P. Nation & G. Kennedy. (1994). How useful is EAP vocabulary for ESP? A corpus-based study. *RELC Journal* 25.2, 34-50.
- [47] Sylvén, L. (2010). Teaching in English or English teaching? On the effects of content and language integrated learning on Swedish learners' incidental vocabulary acquisition. Gothenburg: Acta Universitatis Gothoburgensis.
- [48] Takala, S. (1989). Sanaston opettamisen uudet haasteet (New challenges in vocabulary instruction). In S. Takala (ed.), *Sanaston opettaminen ja oppiminen (Vocabulary instruction and learning)*. Jyväskylä University of Jyväskylä 1-11.
- [49] Viinikkala, N. (2007). The development of vocabulary knowledge in content and language integrated learning in the Finnish comprehensive school. Unpublished MA Thesis. University of Turku.
- [50] Waring, R. & M. Takaki. (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.
- [51] Wesche, M. & T. S. Paribakht. (1999). Incidental L2 vocabulary acquisition: theory, current research, and instructional implications. Introduction, *Studies in Second Language Acquisition* 21.2, 175-180.
- [52] Xanthou, M. (2011). The impact of CLIL on L2 vocabulary development and content knowledge. *English Teaching: Practice and Critique* 10.4, 116-126.

Pävi Pietilä was born in Pori, Finland. She received her MA degree, majoring in English, from the University of Turku in 1978; a *maîtrise ès lettres* degree from l'Université de Franche-Comté, France, in 1979; a Licentiate in Philosophy from the University of Joensuu, Finland, in 1983; and completed her PhD at the University of Turku, Finland, in 1990.

Her permanent position is that of Professor of English at the University of Turku, Finland, but in 2013, she divides her time as a visiting scholar at the University of Cambridge, UK, and the City University of New York, USA.

Her research interests include second/foreign language learning, second language attrition, L2 speaking skills, L2 academic writing, vocabulary development and the lexis-grammar interface.

Riika Merikivi comes originally from Huittinen, Finland. Her MA degree, with English as her major subject, is from the University of Turku, Finland (2012). She has formerly worked as a CLIL teacher and is now teaching English in a vocational institute in Turku.

She is interested in researching foreign language learning, particularly vocabulary acquisition in a CLIL environment. Her research interests also include the relationship between receptive and productive vocabularies in different target languages.