

FINNISH EFL LEARNERS' SIZE AND DEPTH OF RECEPTIVE
VOCABULARY KNOWLEDGE

Master's thesis
Venla Lahtikallio

University of Jyväskylä
Department of Languages
English
March 2016

JYVÄSKYLÄN YLIOPISTO

Tiedekunta - Faculty Humanistinen tiedekunta	Laitos - Department Kielten laitos
Tekijä - Author Venla Lahtikallio	
Työn nimi - Title Finnish EFL learners' size and depth of receptive vocabulary knowledge	
Oppiaine - Subject Englanti	Työn laji - Level Maisterintutkielma
Aika - Month and year Maaliskuu 2016	Sivumäärä - Number of pages 81 + 1 liite
<p>Tiivistelmä - Abstract</p> <p>Sanasto-osaamisen tärkeys vieraan kielen oppimisessa ja hallitsemisessa on nykyään itsestäänselvää. Sanaston oppimista pidettiin kuitenkin pitkään toissijaisena, kun yleinen näkemys oli, että pelkkä kieliopin osaaminen mahdollistaa sujuvan kielenkäytön. Sanaston oppimiseen, opettamiseen ja tutkimiseen onkin kiinnitetty enemmän huomiota vasta muutaman viimeisen vuosikymmenen aikana. Suomessa on tutkittu monia vieraan kielen oppimisen ja osaamisen osa-alueita ja tarjolla onkin paljon laadullista tietoa esimerkiksi opetusmenetelmistä ja yksilöiden oppimiseen vaikuttavista tekijöistä (ks. esim. Pietilä ja Lintunen (toim.) 2014). Määrällistä tutkimusta vieraan kielen oppimisesta ja erityisesti sanaston osaamisesta on kuitenkin tehty hyvin vähän.</p> <p>Tämän tutkimuksen tarkoitus oli ottaa selvää suomalaisten yhdeksäsluokkalaisten sekä lukion toisen vuoden opiskelijoiden englannin kielen sanasto-osaamisesta. Tutkimus oli luonteeltaan määrällinen ja sen aineisto koostui yhteensä 130 oppilaan testituloksista, jotka analysoitiin erilaisia tilastollisia menetelmiä käyttäen. Osaamista tutkittiin reseptiivisen sanaston laajuutta ja syvyyttä arvioivien testien avulla, jotka suoritettiin kirjallisesti oppilaiden englannin kielen tunneilla. Sanaston laajuuden mittaamiseen käytettiin X-Lex-testiä, joka arvioi englannin 5000 yleisimmän sanan osaamista, kun taas Word Associates Test (WAT) mittaa sanaston syvyyttä kiinnittämällä huomiota synonyymien ja kollokaatioiden osaamiseen. Oppilaiden keskimääräistä sanaston osaamista arvioitiin ikäryhmittäin ja ikäryhmiä vertailtiin keskenään. Lisäksi tutkittiin sanaston laajuuden ja syvyyden välistä korrelaatiota ja yksittäistapauksina muutamia oppilaita, joiden oppimistausta erosi muista oppilaista.</p> <p>Tutkimuksesta selvisi, että kaksi vuotta vanhemmat oppilaat olivat selvästi nuorempia oppilaita edellä molemmilla sanasto-osaamisen osa-alueilla. Nuorempien oppilaiden sanaston laajuuden kasvu oli kuitenkin ollut suhteessa hieman nopeampaa. Molemmilla ryhmillä oli enemmän tietoa sanojen syntagmaattisista suhteista eli kollokaatioista kuin paradigmaattisista suhteista eli synonyymeista. Tulosten perusteella suomalaisten oppijoiden englannin sanaston laajuus on kasvanut sitten Takalan (1984) tutkimuksen ja on hyvällä tasolla muiden maiden englannin oppijoihin verrattuna. Jatkotutkimukselle on kuitenkin tarvetta, sillä tietoa tarvitaan niin eri ikäluokkien sanasto-osaamisesta kuin sanasto-osaamisen kehittymisestäkin.</p>	
Asiasanat - Keywords EFL, vocabulary, upper level of comprehensive school, upper secondary school	
Säilytyspaikka - Depository JYX	

TABLE OF CONTENTS

1	INTRODUCTION	3
2	VOCABULARY KNOWLEDGE	5
2.1	What is a word?.....	6
2.2	Unit of counting.....	10
2.3	What does it mean to know a word?.....	11
3	SIZE AND DEPTH OF VOCABULARY KNOWLEDGE	16
3.1	Vocabulary size.....	16
3.2	Vocabulary depth	22
3.3	Development and relationship of size and depth.....	25
4	ASSESSING VOCABULARY SIZE AND DEPTH	30
4.1	Ways of assessing vocabulary size	31
4.2	Ways of assessing depth of vocabulary.....	35
4.3	Previous studies on vocabulary size and depth.....	39
4.3.1	Studies on vocabulary size	40
4.3.2	Studies on vocabulary depth.....	41
5	THE PRESENT STUDY.....	42
5.1	Research questions.....	43
5.2	Data collection.....	44
5.3	Participants	47
5.4	Methods of analysis	48
6	FINNISH EFL LEARNERS' VOCABULARY KNOWLEDGE	50
6.1	Vocabulary size.....	51
6.2	Vocabulary depth	57
6.3	Correlations	63
6.4	Pupils with a longer or shorter history of EFL instruction.....	66
6.5	Summary	68
7	DISCUSSION.....	72
8	BIBLIOGRAPHY	78
9	APPENDIX: The vocabulary tests	82

1 INTRODUCTION

Studying and assessing vocabulary knowledge with respect to second and foreign language acquisition is a fairly new area of research. For the most part of the last century, not much attention was paid to vocabulary learning, and this important aspect of language knowledge was somewhat neglected as far as language learning, teaching and testing was concerned (Milton 2009: 1). One of the reasons for this was that structural and other approaches, which were dominant in the 20th century, emphasised the role of language rules and thought that knowing words was secondary in mastering a foreign language (ibid.). Fortunately, the importance of words in learning a language is now being acknowledged and it is evident that words are vital for communication in second and foreign language as well. Research has shown that it is impossible to be a proficient language user if one's vocabulary is very small (Milton 2009: 3). In other words, if a second or foreign language learner wants to achieve a higher level of language proficiency, s/he has to acquire new words and expand her/his knowledge of words. A good knowledge of grammar alone is not enough.

Because of the significance of vocabulary in learning and mastering a language, it is also very important to do qualitative and quantitative research on the word knowledge of language learners in order to gain information about what sort of and how much knowledge they have about words, and how their learning of those aspects that need improving could be reinforced. Research findings can help teachers and other people involved in language teaching develop more suitable materials and methods for promoting learners' word knowledge. Despite the importance of the subject, and the fact that there are several studies on second and foreign language learners' vocabulary knowledge conducted in other countries, very little research has been done on Finnish EFL (English as a foreign language) learners' word knowledge as far as size and depth of vocabulary knowledge are concerned. Takala's (1984) study of the Finnish ninth graders' vocabulary size, i.e., how many words the learners know, is one of the few studies conducted but the EFL teaching and the

status of the English language in Finnish society have changed considerably in the last 30 years, which means that Takala's results are undoubtedly outdated. New information about the vocabulary knowledge of ninth graders and EFL learners of different ages is therefore desperately needed.

Since the national core curricula and the Common European Framework of Reference (CEFR) do not take any stand on vocabulary knowledge, there are no goals set for Finnish EFL learning and teaching regarding vocabulary size or depth. Thus, there must be a lot of variation in how and what kind of vocabulary is taught and learnt in schools. Based on my own experience, vocabulary can be taught very differently depending on how important the teacher considers vocabulary and how s/he thinks is the best way to learn new words. Unfortunately, a very common way of teaching and testing vocabulary knowledge has been to ask pupils to learn a list of separate decontextualized words, which are then quizzed in a word test. The tests are often very straightforward and one-dimensional because they ask learners to give a Finnish or an English equivalent to the quizzed items. Usually learners are expected to use only the words found in the word list they were told to memorise for that particular test. The tests therefore emphasise the knowledge of certain word meanings and more or less neglect the fact that learners might know other meanings of the words, too. In addition, the tests do not pay much attention to whether or not learners know where and how to use the words. Furthermore, this type of word tests with ten or twenty words in total do not give a good overview of learners' vocabulary skills.

All things considered, there is very little objective knowledge of Finnish EFL learners' vocabulary skills, which is why more extensive and versatile tests are needed. The present study attempted to fill at least a part of the void there is in this field of research. I explored the vocabulary knowledge of Finnish ninth graders (N=59) and upper secondary school pupils (N=71), who learn EFL. To be precise, the study measured the size of their vocabulary, or how many words the learners know, and their vocabulary depth, i.e. what they know about the words. This study focused only on the receptive sides of these two aspects of word knowledge, which means

that just the learners' ability to recognise words was being assessed, while the ability to produce written or spoken words was left aside. The participants' receptive vocabulary size and depth was measured using two vocabulary tests, the X-Lex by Milton and Meara (2003, cited in Milton 2009: 254-256) that is a vocabulary size test and the Word Associates Test (WAT) by Read (1998, cited in Cobb 2015) that is concerned with vocabulary depth.

The next three chapters form the theoretical background of the study. In Chapter 2, I will go through various ways to define the word, deal with counting words and discuss what it means to know a word. In the third chapter, I will explain the two key terms of the present study, that is, vocabulary size and depth. The development and the relationship of these two aspects are also considered in Section 3.3. The fourth chapter will introduce ways to assess and test vocabulary knowledge. The two test used in the present study, the X-Lex and the WAT, are also discussed in this chapter. Moreover, I will cover some previous studies on size and depth of vocabulary knowledge in Section 4.3. In the fifth chapter, I will then describe the present study in detail: I will introduce the research questions and explain how the data was gathered, who the participants were and what methods were used to analyse the data. In Chapter 6, the findings will be presented and linked to the theoretical background. Lastly, I will consider the limitations of the study, discuss the findings and their implications for vocabulary learning and teaching as well as give some suggestions for further study in Chapter 7.

2 VOCABULARY KNOWLEDGE

When dealing with vocabulary knowledge, two essential issues need to be covered before it is possible to measure vocabulary size and depth of vocabulary knowledge. It is necessary to know, firstly, what the word is, and secondly, what it is to know a word. Words are often taken for granted; everybody uses, sees and hears them every day in various contexts. There are, however, several ways to define what words are. I will discuss this issue in Section 2.1. With respect to this, in Section 2.2, I will briefly

go through what kind of units can be used to count words. In the third section, I will then cover the issue of what it means to have knowledge of words and what kind of aspects knowledge includes.

2.1 What is a word?

If asked, everybody can give some sort of definition to the *word*. Words are an integral part of communication and, because of that, everybody has some kind of an idea of what the word is. There is, however, no unambiguous definition of the word, but instead differing approaches, which apply different criteria for the word. One way is to make a distinction between so-called lexical words, or content words, and grammatical or functional words (Carter 2012: 23). Lexical words, that is adjectives, adverbs, nouns and verbs, are described as words that have a meaning even when they stand alone outside a context, while grammatical words, i.e. pronouns, articles, auxiliary verbs, prepositions and conjunctions, “have little or no independent meaning” and “have a largely grammatical role” (Carter 2012: 23-24; Singleton 1999: 11). The division into lexical and grammatical words is not so straightforward since some words classified as grammatical, such as *although*, cannot be said to be devoid of semantic meaning (Singleton 1999: 11). Almost all of the words in the vocabulary size test X-Lex and all the words in the vocabulary depth test WAT employed in the present study are considered as lexical words. The X-Lex test includes a few words, such as *which* and *that*, which could be referred to as grammatical words, but it is possible to say what they mean and give an example of how they are used in a sentence. Test takers can therefore recognise these words and imply that they know their meanings. Thus, the division into lexical and grammatical words does not play a role in the present study.

The word can also be defined in orthographic, phonetic, phonological, semantic or grammatical terms (Singleton 1999: 11-13). The orthographic word is commonly understood as a group of letters between two blanks. For instance, the phrase *I would like to have some ice cream* consists of eight orthographic words. The concept of the orthographic word seems relatively easy and is widely used, but it has its

deficiencies. The concept simply does not work in languages such as Chinese and Japanese, which do not consistently indicate word boundaries, or, for example, in indigenous languages, which have never been written (Singleton 1999: 12).

Compound words (*ice cream*) and phrasal verbs (*show up, run across*) might also pose a problem. Additionally, it does not take into account the several meanings of a word or grammatical functions (Carter 2012: 20).

To define the word phonetically is rather difficult since it is hard to separate individual words from the flow of speech and to say where one word ends and the next one begins since there are no distinct pauses between words when they are uttered (Singleton 1999: 12). For instance, it is not very simple to say how many phonetical words the greeting *How do you do?*, which more or less sounds like *Howdoyoudo?*, includes. The phonological word, on the other hand, might be easier to define, but the definitions are not definite and are very language-specific. In English, for example, words normally have only one stressed syllable. (Singleton 1999: 12-13.) This means that *ice* is a word and *cream* is a word because they both have a stress on their first and only syllable, but when the two words are put together (*ice cream*) only the first syllable (*ice*) is stressed and the two orthographic words are therefore considered as one phonological word. In addition, many orthographic words, such as *and* or *but*, are not recognised as words in phonological terms because they do not typically have stress in ordinary speech (ibid.).

The semantic word can be defined as “the minimum meaningful unit of language” (Carter 2012: 21). Nevertheless, this semantic definition is not unproblematic either. In relation to the English language, it is not clear whether compound words (e.g. *ice cream* or *coffee maker*) should be interpreted as one word or two since, apart from creating a meaning together, both parts of the compound word also have their own separate meanings. What is more, the definition does not take into account that there are also units below words, that is bound morphemes (e.g. inflections of tense), which are semantically meaningful (Singleton 1999: 13). The semantic definition, however, makes it possible to differentiate the various meanings an orthographic word might have (Carter 2012: 21). For example, *book* can refer to a written work or it

can mean making a reservation. In semantic terms, these two uses of the orthographic word *book* would be separate words.

The fifth way of defining the word is the grammatical one, according to which the word is “positionally mobile” and has “internal stability” (Singleton 1999: 13). Positional mobility indicates that words do not have a certain place in a sentence but the order of words can be permuted, for instance, *I will gladly help you, gladly I will help you, you I will gladly help*. Internal stability, on the other hand, refers to the order of morphemes within a word, i.e. the places of morphemes are stable in relation to one another. Thus, the two morphemes in *gladly* cannot be put the other way around, **lyglad*. (Singleton 1999: 13-14.) Nonetheless, based on these criteria, the English definite article *the*, for instance, would not be a word because it does not actually have positional mobility but is dependent on the position of the noun it modifies (Lyons, 1968, cited in Singleton 1999: 14). According to Singleton (1999:14), the grammatical definition of the word is least problematic and least language-specific of the definitions despite its flaws.

As explained above, the word can be defined in several ways based on different characteristics, but there is no universal concept of the word, which would suit all languages and situations. Despite the critique, many researchers employ the traditional view that words are separated by spaces, that is, apply the orthographic definition, because in some situations, such as counting words in essays or in corpora, this sort of idea of the word is practical (Daller et al. 2007: 2). It has nevertheless been said that the orthographic word should not be used in a foreign-language learning context because it does not really describe the task learners face when learning word meanings. Instead, “a more functionally differentiated unit is needed”, and the so-called lexical unit, which can be compared to the semantic word, could suit the purposes better. (Bogaards 2001: 324-325.) A lexical unit “must be at least one semantic constituent [...], at least one word” and it is “the union of a lexical form and single sense” (Cruse 1986, cited in Bogaards 2001: 325-326). In short, a monosemic word, such as *necessarily*, which is an orthographic word with only one sense, can be a lexical unit, whereas an orthographic word with multiple meanings

embodies several lexical units (Bogaards 2001: 326). An example of this could be the word *table*, which can be used as a noun, a verb and an adjective and has several meanings. Sometimes a lexical unit can also consist of several words, which have one mutual meaning (ibid.). For instance, *ice cream* matches this definition.

Since the present study aims to measure the participants' knowledge of single written words, the orthographic definition is the best to describe the words in the vocabulary tests employed in this study. That is to say, the vocabulary tests only include words which consist of only one component, i.e. all words presented in the tests are single orthographic items and there are no compound words or phrasal and prepositional verbs, for instance. The use of simple one-part items prevents situations where a testee does not actually know the meaning of *coffee maker* but implies that s/he knows it just because s/he knows the meanings of the two separate orthographic words, that is *coffee* and *maker*, which might falsify the results.

The present study also considers the orthographic definition of the word practical especially in relation to the X-Lex, which measures how many word forms testees recognise, in other words, what the size of their receptive vocabulary is. The X-Lex test is not created to assess how many different meanings of a word testees know, which is why using the lexical unit as the unit of measurement is not relevant. If the knowledge of all lexical units included in each orthographic word was tested, it would either take much more time or only a few orthographically different words could be included in the test, which would mean that the test results would not represent the test takers' overall vocabulary knowledge very well. The WAT is designed to measure depth of vocabulary knowledge by targeting at test takers' knowledge of multiple meanings and collocations. There are actually lexical units in the WAT since test takers have to link the given orthographic word with synonyms or partial synonyms, all of which therefore represent one meaning of the given word. Nevertheless, the lexical unit is not helpful in this case either because the objective of the WAT is not to assess how well one knows all the meanings of a single orthographic word but instead it tries to form an overall impression of test takers' vocabulary knowledge.

2.2 Unit of counting

Based on the reasons provided above, this study uses the orthographic definition of the word. However, there are different ways and different terms for counting these words. Depending on the context, words can be referred to as tokens, running words, types, lemmas or word families (Daller et al. 2007: 2). The number of tokens or running words in a text is equal to the total number of orthographic words, whereas types of words indicate the number of different word forms in a text (Read 2000: 18). That is to say, the words *cat* and *cats* would be considered two different words. Lemmas are groups of words that include a headword and its most common inflected forms, which belong to the same word class. For instance, *sing*, *sings*, *singing*, *sang* and *sung* form a lemma. Word families are also groups of words but contain even more inflections and derivations, which do not have to be of the same part of speech as the headword. (Milton 2009: 10-11.) This means a word family can include words such as *book*, *books*, *booking*, *booked* and *bookshelf*.

When measuring size of vocabulary, it is essential to take the unit of counting into consideration because the results can vary a lot depending on the unit used. Using an inappropriate unit can lead to overestimation or underestimation of vocabulary size (Nation 2007: 39). So far, researchers have not reached a consensus on which unit to use. Usually the choice is made between lemmas and word families. Nation and Webb (2011: 197), for instance, prefer using word families when counting receptive vocabulary and justify their opinion by saying that “word family members require little or no additional learning for listening or reading if the learner already has control of the important affixes of the language”. They (2011: 212) say that using lemmas as the unit of measuring vocabulary size may be problematic because low-frequency derivations, which are related to high-frequency words, might give too high an estimate of words known at low-frequency levels.

On the other hand, Milton (2009: 11), who does not set productive and receptive vocabulary knowledge apart, disagrees with Nation and Webb and states that lemmatisation has been proved useful in estimating vocabulary sizes of elementary

and intermediate level language learners. He says this is based on the assumption that learners at this level are familiar with the most common inflections and derivations only. In other words, a word family would include too many derivations and inflections an intermediate learner would not know and so the size of his/her vocabulary knowledge would be overestimated. Thus, it might be advisable to use lemmas with lower-level language learners and word families with advanced learners, who can be expected to have very good word formation skills (Milton 2009: 12).

Since there are contradicting opinions on which unit of counting to utilise and in which situation, researchers have to make the decision and choose what best suits their research purposes. In the present study, I choose to use lemmas as the unit of counting because the participants of the study are intermediate EFL learners, who are not likely to know a wide variety of derivations and inflections of words. I believe that the vocabulary sizes of the participants would be overestimated if word families were used. However, when referring to studies which have used word families, I will present the vocabulary sizes in word families and, when relevant, in lemmas to enable comparison between previous studies and the present study. One just has to bear in mind that vocabulary sizes reported using different units are not directly comparable with each other, but, for example, with Milton's (2009: 12) formula sizes in word families can be converted into roughly equivalent sizes in lemmas: the word family size multiplied by 1.6 is the size in lemmas.

2.3 What does it mean to know a word?

As difficult as it is to define what the word is, it can be even more complicated to explain what it is to *know* a word. There are several viewpoints to this issue. To begin with, vocabulary knowledge can be divided into receptive and productive knowledge. Sometimes these two are also referred to as passive and active knowledge. Receptive knowledge refers to being able to deal with words and recognise them in the context of reading or listening, whereas productive knowledge of words means that one is able to recall words when one needs to use them in

speaking or writing. The relationship between the two types of knowledge is unclear, however, which could be due to variance in learners' vocabulary skills and different kinds of tests used to assess the knowledges (Daller et al. 2007: 6). Milton (2009:13) notes that receptive and productive knowledge are not so clearly distinct because when one reads or listens one also actively anticipates certain words to come next. In other words, there is no distinct step from receptive to productive vocabulary skills, but vocabulary knowledge could rather be seen as a continuum. Additionally, instead of a learner either having or not having these knowledges, s/he could also be thought to have some knowledge of a word. In spite of this new type of interpretation, the division of receptive and productive vocabulary is still very commonly used to describe word knowledge in vocabulary studies. It is easier to discuss and measure receptive and productive knowledge as two separate items rather than as a continuum in relation to quantitative research in particular. Thus, the present study, which concentrates solely on receptive vocabulary knowledge, chooses to follow the conventional view of receptive and productive knowledge.

Nation (2009: 27) remains faithful to the separation of receptive and productive vocabulary knowledge in his table *What is involved in knowing a word* where he splits word knowledge into several different sections:

TABLE 1 WHAT IS INVOLVED IN KNOWING A WORD (NATION 2009: 27)

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

Note: In column 3, R = receptive knowledge, P = productive knowledge.

There are three main categories, *Form*, *Meaning* and *Use*, which all have been divided into three subcategories. In addition, each subcategory has been split into receptive and productive knowledge. Daller et al. (2007: 4) find Nation's list the "most comprehensive incarnation" of word knowledge so far and many other researchers (e.g. Greidanus and Nienhuis 2001, Milton 2009, Li and Kirby 2012) have taken the advantage of it. Nonetheless, this model is also founded on the traditional black-and-white notion of word knowledge: a learner either has or does not have knowledge of certain aspects of a word when s/he answers the questions asked in the list. That is, the model recognises that a learner does not have to know all the aspects to have knowledge of a word, but it as if ignores the possibility that a learner might have partial knowledge of some aspects. Regardless of Nation's model describing word knowledge somewhat insufficiently, the list of aspects is very extensive and thus the present study utilises it when discussing the issue of what is involved in word knowledge.

Unlike Nation, Ringbom (1987: 36-37) describes aspects of word knowledge as something that have various levels and he brings forward a model of different dimensions of vocabulary knowledge:

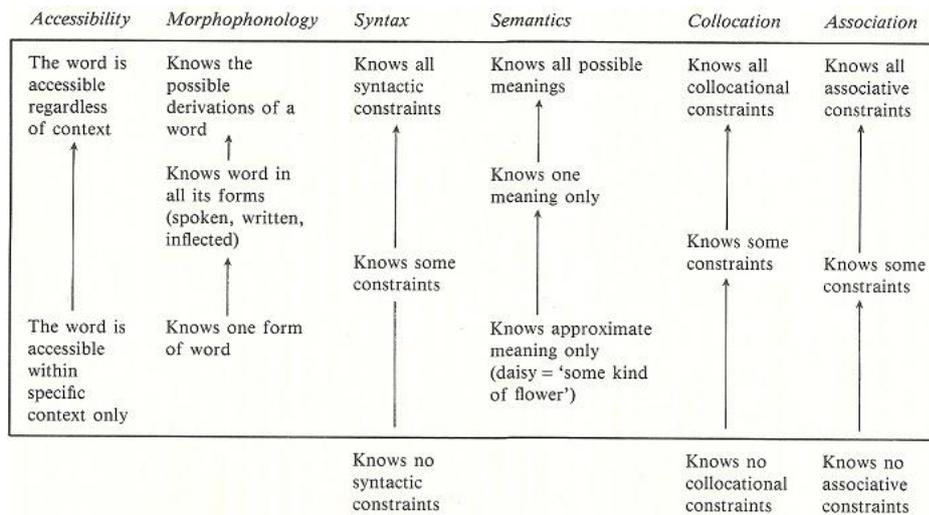


FIGURE 1 LEXICAL KNOWLEDGE (RINGBOM 1987: 37)

According to Ringbom (1987: 35-36), lexical knowledge is a system of various continua, which range from having no knowledge to having full knowledge of a specific aspect of a word and can be used to describe both receptive and productive word knowledge. The continua of word knowledge mentioned in the figure are *accessibility*, *morphophonology*, *syntax*, *semantics*, *collocation* and *association*.

Ringbom's concept acknowledges the fact that knowing a word is not an either-or situation. Instead, there are several aspects of word knowledge and a learner's knowledge of those aspects can vary between the two extremes of the continua. It is important to note that the continua are separate from each other and being near the top of one of them does not automatically mean that one would be near the top of other continua, too. What is more, one has to remember that all the aspects of knowledge, as with other vocabulary knowledge models, are always word-specific, that is, a learner's position on each continuum depends on the word in question. Nobody can have full command of a language but a native or a native-like speaker should be quite close to the top of each continuum (Ringbom 1987: 36). Although Ringbom's and Nation's models share the idea of word knowledge consisting of several aspects, the concepts are still fundamentally different in the way they approach and assess learners' abilities concerning words. As mentioned earlier, this study uses Nation's list in describing aspects of vocabulary knowledge, but Ringbom's model and the idea of continua might actually be a better representative

of the reality of vocabulary knowledge because it is more flexible in acknowledging learner's vocabulary skills.

Daller et al. (2007:8), on the other hand, present a model of vocabulary knowledge, which they call the lexical space. It also uses continua as well as utilises the elements of Nation's list:

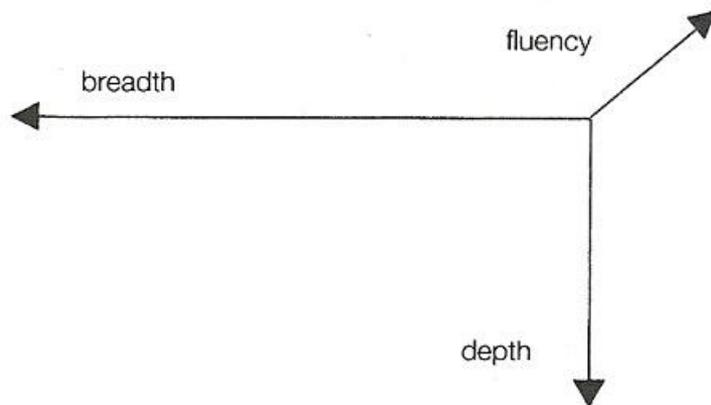


FIGURE 2 THE LEXICAL SPACE (DALLER ET AL. 2007: 8)

The lexical space has three dimensions, *breadth*, *depth* and *fluency*, which describe different aspects of learners' vocabulary knowledge. It is important to remark here that although Daller et al. (2007) manage to present vocabulary knowledge in the form of a simple-looking figure, it does not make the issue any simpler than the models of Nation and Ringbom, because there are several aspects of word knowledge included in the axes. The first dimension, *breadth*, which is also called *size*, includes the *Form* category and the sub-category *form and meaning* of Nation's list, and the second dimension, *depth*, consists of the sub-categories *concepts and referents*, *associations*, *grammatical functions*, *collocations* and *constraints on use*. The third dimension is *fluency*, which refers to the capability to recall and use words automatically and accurately. Thus, the idea is to locate a learner somewhere within the lexical space depending on the size of their vocabulary, their knowledge of words and the ease to recall and use those words. (Daller et al. 2007: 7-9.) Milton (2009: 16) suggests that in this three-dimensional model, *breadth*, or *size*, and *depth* could represent receptive word knowledge while *fluency* would describe productive knowledge. This is not a common view, however, since usually both *size* and *depth*

are considered to have a receptive and a productive side. The terms depth and size are widely used in the field of vocabulary research, whereas fluency has not yet established its position next to them as the third aspect of vocabulary knowledge. The present study is also interested only in size and depth of vocabulary knowledge and chooses to study their receptive sides. The two terms will be elaborated in Chapter 3.

3 SIZE AND DEPTH OF VOCABULARY KNOWLEDGE

Because of the centrality of size, or breadth, and depth in the present study, it is necessary to discuss these terms further. Anderson and Freebody (1981) were the first to define the two aspects of vocabulary knowledge. Breadth of vocabulary knowledge they described as “the number of words for which the person knows at least some of the significant aspects of meaning” and depth of vocabulary knowledge as “the quality [...] of understanding” (Anderson and Freebody 1981: 93). These definitions might sound simple, but in reality, they are not that straightforward. In order to understand the concepts of vocabulary size and depth better, one has to have an idea of what words are and what it is to know a word. These issues have already been discussed in Chapter 2. The present chapter will specify what kinds of issues size and depth include and how these two aspects relate to each other.

3.1 Vocabulary size

As mentioned earlier, vocabulary size can be considered to consist of the *Form* and *form and meaning* components of Nation's list (see Table 1) (Daller et al. 2007). The *Form* component includes sub-categories, which ask how a word sounds like, how it is pronounced, what it looks like, how it is written as well as what word parts are included in it and what parts are needed to express meaning. The *form and meaning* component, on the other hand, asks *What meaning does this word form signal?* and *What word form can be used to express this meaning?* This component could also be thought to be part of vocabulary depth, because it is about knowing a word (Milton 2009: 14).

Daller et al. choose to include it in vocabulary breadth, however, which is understandable: this makes it easier to assess whether a testee actually recognises a word since s/he can be asked to provide an example regarding the word. Nevertheless, a person's ability to answer any of these questions mentioned depends on his/her knowledge of words and whether that knowledge is receptive or productive by nature. The present study concentrates on receptive knowledge and is particularly interested in finding out how many word forms learners can recognise. That is, the study mainly covers the question *What meaning does this word form signal?* in relation to vocabulary size.

The size of receptive vocabulary tends to be bigger than that of productive vocabulary, although the exact relation between the two is still uncertain; studies on foreign language learners' vocabulary have shown variation in the ratio of receptive and productive vocabulary size, which may be due to learners' language levels and differences in their learning experience (Milton 2009: 125). In their study, Eyckmans et al. (2007: 74) discovered that productive vocabulary is about half of the size of receptive vocabulary, which is quite a common view (Milton 2009: 123). Other studies (e.g. Erigna 1974, Burns 1951, cited in Milton 2009: 123) have got results where the breadth of productive vocabulary varies between 40% and 80% of receptive vocabulary. Takala's (1984) findings even suggest that at intermediate and low levels receptive and productive vocabulary might be almost equal in size. Nation (2009: 371) drawing conclusions from a few studies, says that when learners' vocabulary becomes wider, their receptive lexicon grows in relation to their productive lexicon. However, although the ratio of the two might vary, the number of receptively known words is usually bigger at all language learning levels. One factor that promotes the larger size of receptive vocabulary is that a learner can utilise contextual and other clues with written and spoken texts to deduce word meanings. In connection to that, a learner first gains receptive knowledge of a word and after more frequent and versatile encounters with the word possibly learns how to use it when writing and speaking, in other words, gets productive knowledge of it. Since the present study deals with receptively known words, I will, from now on, mainly discuss and refer to receptive vocabulary.

The English language has several roles and is used for many kinds of purposes: Some people speak it as their first language, some as a second or as a foreign language depending on what kind of a status English has in their country. Sometimes English can be the only common language, i.e. the lingua franca, between people from different countries and cultural backgrounds. Some people use the language in their everyday lives, some study or work through the medium of English, while some others use English only when they take a holiday abroad. That is to say, language learners' and speakers' needs to use the language can vary significantly. This should be taken into consideration when vocabulary sizes, as well as vocabulary depth, are concerned because language users' differing needs also affect their vocabularies: even native speakers differ from each other in their vocabulary knowledge, which is due to their educational and cultural backgrounds, for instance. A very interesting question is, however, how many words an English language learner needs to know in order to cope with the language in various situations. This issue will be discussed next.

There is no certainty how many words the English language actually contains, but some estimates have been made. For example, Goulden et al. (1990: 356) suggest that the number could be around 110,000 word families, but the estimate is based on the number of words in one dictionary, *Webster's Third New International Dictionary*, only. This means that the actual number of English words is probably much bigger. Nevertheless, it would be utterly impossible as well as unnecessary for a non-native speaker to try to learn all of the words in English, since even native speakers have only a part of all the words in their vocabulary. Many of the words are related to a specific area of study and are technical words, which is why they are not very frequent in the language and thus unknown and useless to many speakers. Many words are also specific to a certain variety of English, which means that only the speakers of that variety are likely to know those words. What is more, new words are coined all the time to answer the requirements of the changing world, while old words die away or get new meanings. It is therefore impossible even for a native speaker to master all the words and to keep up with the development of the English vocabulary.

It has been estimated that well-educated native speakers of English know around 20,000 word families (or some 32,000 lemmas), which would be equivalent to 1,000 acquired word families per year up to the age of 20 (Nation 2006: 60). Nation (2009:9) notes that to achieve a vocabulary equal to that of native speakers is not a realistic goal for most foreign language learners, although it is possible if the learner in question is very advanced or uses English as a second rather than as a foreign language. In his study, Nation (2006: 60) actually found out that non-native speakers of English, who were highly educated and studied through the medium of English, had a receptive vocabulary of around 8,000 to 9,000 word families, that is, some 12,000 to 14,400 lemmas. On that account, it could be suggested that the vocabulary size of very advanced non-native speakers could be about half of the vocabulary size of well-educated natives. Since it is also possible for other than well-educated people to come along in an English-speaking environment with their foreign language skills, the actual number of words a learner needs to know might be lower than 8,000 word families.

It has been suggested that a coverage of 95% or 98% is needed in order to comprehend written and spoken texts (Laufer 1989, cited in Laufer and Ravenhorst-Kalovski 2010: 17; Hu and Nation 2000: 422). The two different figures are due to the differences in defining what *adequate* comprehension is. Laufer and Ravenhorst-Kalovski (2010: 25-26) suggest that both 95% and 98% can be considered lexical thresholds. The optimal threshold that refers to “functional independence in reading” is at 98% and represents the knowledge of some 8,000 word families (12,000 lemmas), while the minimal threshold of 95% “enables learners to read with some guidance” and is equivalent to 4,000 to 5,000 word families (6,400 to 8,000 lemmas). (ibid.) Hu and Nation (2000: 419), however, only support the idea of 98% because most learners need to achieve this coverage level to be able to comprehend texts. To achieve the coverage of 98%, Nation says (2006: 79) one has to know receptively 8,000 to 9,000 word families, which is in line with Laufer and Ravenhorst-Kalovski’s results. This is also the number of word families that highly educated language learners might have in their vocabulary. In order to achieve the same coverage with spoken texts, a person needs to know 6,000 to 7,000 word families, in other words,

around 9,600 to 11,200 lemmas (Nation 2006: 79). Due to the repetitive and formulaic nature of spoken language and the use of gestures and facial expressions when speaking, the frequency of words is different than in written language and fewer words are needed to comprehend spoken texts (Milton 2009: 55).

It depends greatly on the text type and the situation, however, how wide a vocabulary a person needs; reading a novel is very different to reading newspapers and listening to a lecture differs greatly from hearing informal spoken conversations. Nonetheless, it is said that a coverage of at least 80% is needed to understand the main points of a text (Milton 2009: 52, 54). Being able to understand 80% of any written texts, including novels and newspapers, somewhat corresponds to the 2,000 most frequent lemmas, which is about 1,250 word families (Nation 2009: 15). Reading informal texts, such as personal messages, notes or blogs, often presumes a much smaller vocabulary. In comparison, understanding 80% of conversational speech requires knowledge of only some 200 word families, which is equal to 320 lemmas (Adolphs and Schmitt 2003: 431), whereas around 1,000 word families (1,600 lemmas) cover 80% of context-governed texts including lectures and meetings (Milton 2009: 58). In conclusion, a foreign language speaker can manage with a relatively small receptive vocabulary in informal day-to-day encounters with the language but has to have quite a large vocabulary if s/he wants to be able to perform fluently in a variety of language situations and environments (Milton 2009: 75-76).

Using the scores of the X-Lex test, which assesses learners' knowledge of the 5,000 most frequent words, Milton and Meara (2003, cited in Milton 2009: 186) have linked EFL vocabulary size with the levels of the Common European Framework of Reference for Languages (CEFR) as Table 2 shows:

TABLE 2 MEAN EFL VOCABULARY SIZE SCORES AND THE CEFR (ADAPTED FROM MILTON 2009: 186)

CEFR level	X-Lex
A1	< 1,500
A2	1,500 – 2,500
B1	2,500 – 3,250
B2	3,250 – 3,750
C1	3,750 – 4,500
C2	4,500 – 5,000

This connection is somewhat indicative of learners' capability to manage in different language situations when they master a certain number of the most frequent words. Learners at the A level are defined as *basic users*, B level learners are considered *independent users*, whereas those who have achieved the C level are *proficient users* (Council of Europe 2001: 23). Basic users at the A2 level, for example, are expected to be able to "understand sentences and frequently used expressions related to areas of most immediate relevance" (Council of Europe 2001: 23-24). This level would correspond to some 1,500 to 2,500 words, in other words, lemmas. At the B1 level, which is also called the *threshold level*, learners "can understand the main points of clear standard input on familiar matters" and they can also "deal with most situations likely to arise whilst travelling in an area where the language is spoken." (ibid.) B1 learners should possess a vocabulary of around 2,500 to 3,250 words. As proficient users at Level C1, learners can already "understand a wide range of demanding, longer texts, and recognise implicit meaning", which would mean they have a vocabulary of 3,750 to 4,500 words.

When the CEFR levels and the corresponding EFL vocabulary size are compared with the suggestion that the 2,000 most frequent lemmas are needed to understand 80% of any text, one can notice that 2,000 lemmas would correspond with the A2 level. It is unlikely, though, that basic users would be able to understand even the main points of novels and newspaper articles, since they only understand topics close to their own life. It is more probable that B2 or C1 learners, who are can understand more complex texts and abstract topics, reach the threshold of

understanding any text. Consequently, there is a discrepancy between threshold of understanding and the size of vocabulary at each level of the CEFR, because a B2 level learner has a vocabulary of 3,250 words or more instead of 2,000 words. It is of course possible that a learner does not know all the 2,000 most frequent words, but has knowledge of other words instead, which partly explains the disparity.

Additionally, the slight differences in the word frequency lists used to estimate the threshold of 2,000 lemmas and to create the X-Lex test play a small part as well.

Nevertheless, the table of the CEFR levels and vocabulary sizes can be a useful tool and enables the comparison of X-Lex results and the expected CEFR level of learners, since the CEFR itself does not include recommendations on vocabulary size.

3.2 Vocabulary depth

Depth of vocabulary is quite a complex concept and comprises several components concerning meaning and use of a word listed by Nation (see Table 1). From the point of view of receptive knowledge, *concepts and referents* deals with the multiple meanings of words, *associations* is about different meanings and associations words might have, while *grammatical functions* concerns the pattern in which a word appears. *Collocations* refers to words and word types with which a specific word usually occurs and *constraints on use* to where, when and how often a word is seen or heard. Vocabulary depth is therefore a multifaceted construct, which consists of several pieces of vocabulary knowledge.

Some researchers prefer the concept of word association network structures to depth, although they concern with some of the same aspects of vocabulary knowledge.

Vocabulary networks refer to the connections between words people have in their vocabularies, but the concept is said to differ from that of depth so that it considers vocabulary as a structure rather than as a group of separate words with meanings (Meara 2009: 62, 71). Meara (*ibid.*) defends the idea of networks and is of the opinion that depth is only “a property of individual words, not a property of the vocabulary as a whole”. He says the reason for this is that research on depth concentrates on only a few words and tends to ask too detailed questions about those words.

It seems to me, however, that the network model has the same problems Meara claims vocabulary depth to have: It can also be seen only as a property of individual words because the connections between words are always dependent on how much and what kind of knowledge of a particular word one has. The different types of connections will be elaborated in Section 3.3. Additionally, it is impossible to map all the connections in a learner's lexicon with a network test, since that would simply take too much time and effort. Researchers must therefore concentrate on fewer words in network tests, as is the case with depth tests as well. What is more, the concept of word association networks could also be said to be rather detailed because it deals with even fewer aspects of vocabulary knowledge than depth. It is mainly comprised of the aspects of *associations* and *collocations*, which are also included in the concept of depth. The two concepts are thus partially overlapping and describing the same phenomena. I do not see why using one model should exclude the use of the other. Instead, network structures could be used as a tool to describe some of the word properties included in depth. However, despite the fact that network structures could provide interesting information about collocations, which is one of the aspects measured in the present study, I prefer the concept of depth because it is better established, widely used and it offers a wider perspective on vocabulary knowledge as a whole.

As already mentioned above, the present study attempts to measure collocational knowledge as part of learners' depth of vocabulary knowledge, which is assessed with the WAT. There is thus a need to elaborate briefly the term collocation. To have knowledge of words involves knowing what words usually occur with each other. This is referred to as collocational knowledge and it is one of the aspects mentioned in Nation's list (see Table 1). Varying definitions can be attached to the term collocation. Cambridge Dictionaries Online (2016) defines it as "the combination of words formed when two or more words are often used together in a way that sounds correct". This description is quite loose and can include many types of word combinations. That is, depending on the point of view, idioms and compound words can be either included in or excluded from the term. The WAT used in the present study does not include idioms or compound words.

Another definition says that the collocation is “a group of words which occur repeatedly in a language” and is either lexical or grammatical (Carter 2012: 62). Lexical collocations are described as consisting of verbs, nouns, adjectives and adverbs, i.e. lexical words, while grammatical collocations consist of a dominant lexical word and a preposition or a grammatical structure, such as an infinitive or clause (Benson et al. 2010: XIX, XXXI). In the present context, all collocations are composed of two lexical words, that is, of an adjective and a noun (e.g. *bare feet*), which means they are lexical collocations. Examples of grammatical collocations are *on purpose*, *look at* and *It was a pleasure to see you*, for instance. The present study leaves this type of collocations aside simply because the employed vocabulary tests do not contain such items.

Additionally, collocations differ from each other in their size, the closeness of collocates and the possible range of collocates as well as in the semantic unpredictability they have. Size means the number of words in a collocation. Grammatical collocations with a clause or an infinitive in particular can comprise a number of words (e.g. *to consider someone to be something* - *They considered him to be a good teacher*), but in the WAT, the size of each collocation is two words. As for closeness of collocates, it refers to the fact that the distance between collocates in a sentence may vary: sometimes collocates are positioned next to each other and occasionally there can be several words between them, e.g. *make a decision* (*I had to make a very difficult decision*). However, this aspect is not relevant in the present study, because the words are presented without context, whereas the possible range of collocates plays a role. Some words tend to collocate with only a few words, while some other words can form various collocations with other words. (Nation 2009: 56, 330-331.) For instance, the noun *wish* can be called a strong collocator because it occurs only in few collocations, such as *make a wish*. By contrast, the adjective *beautiful* can collocate with multiple words (e.g. *face*, *weather*, *house*) and thus, it is a weak collocator. (Cambridge Dictionaries Online 2016.) This aspect is also called collocational restriction, which ranges from unrestricted collocations to restricted collocations (Carter 2012: 78-79). The WAT is mainly comprised of weak collocators, i.e. unrestricted collocations, because the words can co-occur with a range of other

words.

When it comes to semantic unpredictability, there are differing opinions about whether collocations should be semantically unpredictable or not. Nation (2009: 317), for example, thinks that the meaning of a collocation should not be directly deducible from the meanings of its parts, but should be unpredictable to some extent. Carter (1987, cited in Singleton 1999: 23), on the other hand, does not set such strict requirements but says that collocations may be “more or less transparent semantically” and uses a continuum with fixed points to categorise collocations: the semantic opacity of collocations can range from transparent to opaque (Carter 2012: 79-80). The collocations of the WAT are rather transparent since their meanings could more or less be said to be the combination of their parts, for instance, *dull + knife* or *fresh + water*.

Collocational knowledge plays a very significant role in mastering a language because “stored sequences of words are bases of learning, knowledge and use.” It has been said that familiar word combinations actually form a major part of the language people use, while entirely new combinations are in minority. Thus, when words are learnt and stored in long-term memory as chunks, it helps one to understand and produce the language more effectively. (Nation 2009: 321, 323.) By having chunks of language in the memory, a foreign language user can select the most suitable words from his/her vocabulary for each situation and use the language fluently (Pawley and Syder 1983, cited in Nation 2009: 323). Knowing collocations, as well as other word combinations, is therefore very advantageous for a language learner, which is why it makes sense to pay attention to and test this type of word knowledge.

3.3 Development and relationship of size and depth

Anderson and Freebody (1981: 93) assume that “a person has sufficiently deep understanding of a word if it conveys to him or her all of the distinctions that would be understood by an ordinary adult under normal circumstances.” This kind of description of vocabulary depth sounds somewhat problematic, however. An

ordinary adult refers to a native speaker, but what is an ordinary native speaker like and what exactly are normal circumstances. Ringbom's (1987) view, as cited above, is that nobody can have a fully developed vocabulary knowledge or a full command of all the aspects of language, because even native speakers do not know even receptively all the words, let alone all the meanings of the words they can use. Additionally, native speakers' vocabulary knowledge reflects their education, occupation as well as social and cultural background among other things (Read 2004: 213). That is to say, there is no ordinary native speaker, and it is even difficult to say what an average native speaker's vocabulary knowledge includes.

Despite the variety among native speakers, learners' vocabulary knowledge skills are generally compared with those of native speakers. When doing so, an important point to remember is that the vocabulary skills of first language (L1) learners and foreign language (FL) learners develop at a very different pace, which is mainly due to the differences in the language environment. The language surrounds L1 learners in their everyday life, which means that the input frequency of words is high, and L1 learners can acquire many new words in a short time. FL learners, on the other hand, do not usually encounter or have the chance to use the language and its words nearly as frequently. The participants of the present study learn English as a foreign language in Finland. Although the English language does not have an official status in Finland, media, social media and games, for instance, offer EFL learners opportunities to encounter English outside the language classroom and use it receptively and productively. It, nevertheless, depends on a person whether s/he takes advantage of these opportunities. However, the amount of English used in Finnish society is altogether much lower than it is in an English-speaking country, which substantially affects the learning of the language and differentiates Finnish FL learners from L1 learners.

According to Vermeer (2001: 231) the input frequency of words, alongside with a person's ability to learn words, has a significant effect on word knowledge, and for this reason, L1 learners (or speakers) are ahead of FL learners concerning vocabulary size. L1 learners are also ahead in vocabulary depth, because when people learn new

words they also learn more of the words they already have in their vocabulary (Read 2004: 221). Referring to Ringbom's model of lexical knowledge, adult L1 speakers tend to be near the top of each continua, while FL learners, who have to work their way up from the very bottom, only seldom reach the levels of native speakers, although it is of course possible if one is a highly proficient learner (Ringbom 1987: 36).

There might also be differences in how native speakers' and foreign language learners' vocabulary knowledge is structured and organised. It is generally thought that the FL lexicon does not have as many links between words as the L1 lexicon and many of the links may even be wrong and inappropriate (Milton 2009: 149). What is more, Meara (1983: 1-2) found out that FL learners provide more varied and less homogenous responses to stimulus words than L1 speakers, although one would expect their smaller vocabulary size to limit the connections between words. FL speakers also tend to produce so-called clang responses like young L1 children usually do, while older L1 children prefer syntagmatic responses and adult L1 speakers normally produce paradigmatic associations. Clang responses are phonological associations, which can be based on, for example, rhymes (*blade - shade*) or initials sounds (*share - sheer*), but which have no obvious semantic relationship to stimulus words. (Meara 1983: 1-2.) Syntagmatic and paradigmatic responses, on the other hand, are semantic word associations. Syntagmatic, or combinatory, associations link together words that usually collocate with each other, such as *grass - green*, or *dog - barks*, whereas paradigmatic, i.e. substitutional, associations include links based on synonymy, antonymy and hyponymy; *dog - wolf - cat - animal* (Singleton 2000: 170, 181-182; Meara 1983: 1).

It has been argued, however, that L1 and FL speakers producing different types of responses has nothing to do with a particular language being their first or foreign language. Instead, the associations one produces depend on how familiar one is with the word in question. Thus, when L1 or FL speakers encounter a new word, they both tend to produce form-focused associations before they become familiar with the word meaning and are able to produce semantically motivated links. (Singleton 2000:

182.) The familiarity of a word is naturally related to the frequency of that word in one's language environment. The more frequently one encounters the word, the more knowledge of that word one is likely to have. However, because native, second and foreign language speakers can have very different language environments, meaning that they use English dissimilarly and for different purposes, the most frequent words in one environment are not likely to be the most frequent words in the other. This should be taken into account when using vocabulary tests based on word frequency. For instance, results suggesting that FL learners have fewer semantic links between words than native speakers might actually mean that the stimulus words are not as frequent in the FL context as in the L1 environment. What is more, native, second and foreign language speakers are by no means homogeneous groups, but every individual encounters and uses the language in differing contexts: L1 speakers of different varieties of English are likely to have divergent knowledge of words, for example, and a person's education and occupation play a great role in what kind of knowledge of words s/he possesses. In conclusion, the way L1 and FL speakers form links between words is essentially the same, but the word associations one is able to produce depend on how familiar one is with the words in question, which is affected by the frequency of those words in one's language environment.

Naturally, there are also differences between FL speakers with respect to the development of their vocabulary knowledge. The time used to study the language and the intensity of study as well as the teaching provided all affect learners' vocabulary skills. In relation to vocabulary size, there have been arguments for advancing the uptake at the beginning of language learning (e.g. Nation 2009). Feeding learners with high-frequency words as early in their language learning careers as possible could help them achieve a certain level of independence as language users relatively soon (Milton 2009: 76-77). After this, learners could carry on expanding their vocabulary size with less frequent words and their vocabulary depth.

There has also been discussion about whether the proficiency level of learners affects their vocabulary uptake. Some have proposed that the more advanced the learners are, the easier and quicker they can adopt new words because of their existing knowledge of lexis and morphology. At the same time, it is said that advanced learners might actually develop their depth and fluency at the expense of size to become more proficient language users. (Milton 2009: 76-77.) This, on the other hand, would mean that the pace of vocabulary uptake does not increase with the proficiency level, but the aspect of depth becomes more important when the proficiency increases. So far, there is no univocal evidence for how FL learners actually acquire words and whether learning them would be easier or quicker at the outset of learning or at some other level.

As discussed above, native speakers and FL learners are thought to differ from each other in the way their lexicon organise words, but FL learners are not a homogeneous group regarding the structure of their vocabulary either. On one hand, some can have a large vocabulary size, but the organisation of their lexicon can be quite poor, which means the words in their vocabulary are not connected to each other that well. On the other hand, there are people who do not know many words but have an organisation in their lexicon, which is closer to that of native speakers. This might be the cause for differences in being able to perform with the foreign language in academic contexts and communication situations. (Milton 2009: 150.) Such variance in vocabulary skills also raises a question about whether size and depth are entirely separate dimensions or are they somehow connected to each other.

Since Anderson and Freebody, as Read (2004: 221) points out, the tendency to separate and somewhat contrast breadth and depth has prevailed, although there is now some evidence of a close relation between them. Vermeer (2001) studied the size and depth of word knowledge of 50 Dutch monolingual and bilingual kindergartners by using a receptive vocabulary task and a word description task to measure vocabulary size, and a guided association task to assess vocabulary depth. She found evidence for size and depth being closely related when it comes to children. Based on her findings, Vermeer (2001: 231) suggests that “the denser the network around a

word, the richer the set of connections around that word, the greater the number of words known, and the deeper the knowledge of that word." In addition, she concludes that a child who knows more words usually also knows more of each word.

Qian (1999: 299), who studied 74 ESL (English as a second language) learners of Korean and Chinese backgrounds in Canada, found out that the scores of breadth and depth tests correlated strongly and suggested that the two dimensions overlap and are probably interconnected and interdependent. He measured vocabulary size with Nation's Vocabulary Levels Test and vocabulary depth with Read's Word Associates Format. Nurweni and Read (1999) assessed the English vocabulary knowledge of Indonesian university students (N = 350) by using a translation test and a word associates test and by interviewing a smaller group of students (N = 39). They got similar findings, but they also found out that the correlation between size and depth was stronger the more advanced the learners were, based on their achievement in the national examination of English. The findings by Akbarian (2010), who measured Iranian ESP (English for specific purposes) learners' (N = 112) word knowledge with Nation's Vocabulary Levels Test and Read's WAT, also support the view that when learners are advanced the two dimensions might be closely related to each other, but at lower levels of language knowledge, the dimensions may be more separate.

4 ASSESSING VOCABULARY SIZE AND DEPTH

Daller et al. (2007: 15-17) state that the results of tests measuring receptive vocabulary knowledge are very reliable, in other words, the tests are consistent and treat people with the same abilities the same way. Test reliability refers to the accuracy of a test, whereas test validity is about how a test measures the features it ought to measure. However, Daller et al. (ibid.) want to point out that in language studies it is rather impossible to say that a test is clearly valid or invalid, since language knowledge is

difficult to quantify and there are many aspects of validity, which need to be considered.

As a part of language knowledge, vocabulary knowledge is also very difficult to quantify, which makes its assessment problematic. Despite there now being some evidence of the relationship between size and depth, there is, however, no test that would be able to measure vocabulary knowledge as a whole. It is therefore easier to divide the knowledge into size and depth, because these aspects are easier to define and measure. The present study will also consider them as two separate aspects. In this chapter, I will give an overview of different ways to assess size and depth and present some tests used by researchers. I pay special attention to the two tests, the X-Lex and WAT, which are employed in the present study.

4.1 Ways of assessing vocabulary size

According to Nation and Webb (2011: 196-212), there are three main approaches to assessing vocabulary breadth. First, one can count all the words testees produce, second, one can test how many of the words in a dictionary testees are familiar with, and third, one can sample frequency lists from several frequency levels to get an estimation of word knowledge at each level. Nation and Webb (ibid.), however, give several reasons for why the first approach does not measure vocabulary size very well, the main reason being that test takers are likely to produce only a part of the words they know and could produce. They go on to mention that there are many issues related to sampling, which need to be considered if dictionaries were used to measure breadth: one has to decide which dictionary suits the particular purpose, how to count the total number of entries and what is included in a word family, for instance. It would therefore be possible to use the second approach and get rather good estimates of vocabulary size, but Nation and Webb (ibid.) say that by using vocabulary frequency lists, one could actually avoid many of these issues.

Using frequency lists in measuring vocabulary knowledge has gained a lot of support from researchers and many have utilised vocabulary size tests, such as the

X-Lex and the Vocabulary Levels Test, which take advantage of frequency lists. Based on the thought that the more frequently certain words appear in the language, the more useful it is to know them, it is easy to justify the importance of learning and teaching these words, and hence it makes a lot of sense to test and assess the knowledge of the words as well. Nation (2009: 16), for instance, considers the 2,000 words found in the first two frequency bands of English corpora so important that language learners should do anything they possibly can to learn these words. Assessing learners' vocabulary size in relation to frequency lists is therefore reasonable and provides accurate results efficiently (Milton 2009: 22).

Li and Kirby (2012: 339-340) elaborate the topic further and list six widely used test types to assess vocabulary size. The first one is the written multiple-choice format vocabulary test, which measures a test taker's ability to substitute synonyms, match translations and complete definitions. The second and third test type are oral tests that assess receptive and productive vocabulary breadth by asking a test taker to point to the picture that represents the word spoken by the examiner or to name orally the pictured verb or noun. The fourth type of test is about translating L1 words into FL words or the other way around, whereas in the fifth kind, a testee has to link definitions with words. A well-known example of this fifth test type is the Vocabulary Levels Test (VLT) by Nation.

Although the VLT was originally meant as a tool for teachers to develop suitable vocabulary learning and teaching methods, many researchers have used it to assess non-native speakers' vocabulary size in the absence of a better-developed measure (Read 2000: 118). The test consists of five levels, which are based on frequency bands: 2,000-word, 3,000-word, 5,000-word, University word and 10,000-word levels. In Nation's (1990, cited in Cobb 2015) own version of the test, each level has six groups of six words, out of which only three have matching definitions. That is, test takers have to show their vocabulary knowledge by matching the definitions to right words. The test thus measures receptive vocabulary knowledge. Each correct answer is worth a point the maximum score being 18 at each level. In the improved version of Schmitt et al. (2001), the form of the test is the same but some items have been

revised and the number of items tested has been increased to ten groups of six words, which makes the total score 30 points per level. Although the test only includes 18 or 30 words at each level depending on the version, Nation (1990: 262) says the results can be generalised to vocabulary size. In other words, if a test taker scores 15 out of 30 at 2,000-word level, s/he can be estimated to know somewhat half of the words of the first two frequency bands, i.e., around 1,000 words or lemmas.

- 1 business
- 2 clock _____ part of a house
- 3 horse _____ animal with four legs
- 4 pencil _____ something used for writing
- 5 shoe
- 6 wall

FIGURE 3 AN EXAMPLE ITEM FROM THE REVISED VERSION OF THE VLT (SCHMITT ET AL. 2001: 82)

Despite the VLT being generally accepted as a good measure of vocabulary size, I choose not to measure vocabulary size with it but utilise another test instead, which represents the sixth type of test mentioned by Li and Kirby (2012: 339-340), that is, a checklist test. The X-Lex test by Milton and Meara (2003, cited in Milton 2009: 254-256) is an orthographic vocabulary recognition test that measures the size of receptive vocabulary knowledge. It is a simple checklist test that asks test takers to tick the words they know or can use. The test includes 120 lemmatised words in total, 20 randomly selected words from the first five 1,000 word frequency bands and 20 pseudo-words, which look and sound like English words but are actually made up. (Milton 2009: 12, 26, 71-74.) All the words are presented separately and without context or any other clues, which might help testees deduce word meanings (see Table 3).

TABLE 3 EXAMPLE ITEMS OF THE X-LEX (MILTON 2009: 255)

had	govern	system	interval	mosquito	warboy
which	industry	position	overcome	proceed	cordonise

The frequency bands are based on Nation's general frequency lists and Hindmarsh's lists that are tied to the vocabulary of EFL books and exams (Milton 2007: 49-50). The

use of pseudo-words reduces the effect that guessing has on the scores. The estimated size of test takers' vocabulary knowledge is calculated by counting the number of real words ticked and multiplying it by 50 to get a raw score out of 5,000. Then the number of ticked pseudo-words is multiplied by 250 and subtracted from the raw score to get an adjusted score out of 5,000. (Milton 2009: 74.) The score thus gives an estimation of how many words out of the most frequent 5,000 a person knows and also indicates the number of words s/he knows from each of the five frequency bands. It is important to note, however, that the score does not tell the total number of words known by a testee because s/he probably also knows words that are not included in these frequency bands. For example, words that are frequent in an EFL classroom or relate to the culture or the environment in which EFL is learnt might not be so frequently used in the English-speaking world. Nonetheless, it makes sense to assess the knowledge of the most frequent words because of their usefulness in the language.

The X-Lex has many advantages compared to other types of tests. First, the test has a clear format and can be completed fast, which means test takers are not likely to get bored or lose concentration in the middle of the test (Milton 2009: 72). Second, the checklist format test enables testing a larger number of words than other test formats, and therefore the results can be more reliable compared to other kinds of tests (ibid.). In addition to that, Daller et al. (2007: 16) point out that the X-Lex has a good content validity because the test does not "waste time and effort testing words which learners will never know" and because it uses "frequency information to target testing in areas where there is likely to be knowledge."

As any other tests, the X-Lex also has its disadvantages. One problem, which this type of Yes/No-test has, is that test takers do not have a common definition for what knowing a word means (Li and Kirby 2012: 340). These different conceptions might affect the results of the test; one test taker does not tick a word, which s/he remembers seeing or hearing a couple of times but does not know the meaning of, whereas the other ticks the word based on the same criteria. To get as reliable results as possible when comparing the scores of the participants of the present test, it is

thus very important to explain the testees what is meant by knowing a word. That could be, for example, knowing one of the meanings of the word and either use the word in a sentence or find a corresponding L1 word for it.

Another issue is the possibility that a test taker might not know a word but guess correctly instead. This problem is at least partially solved by including pseudo-words into the test, which compensate guesswork. If a testee ticks a pseudo-word and so claims to know its meaning, it is quite probable that s/he has guessed meanings of actual words, too. Each ticked pseudo-word therefore lowers the total score by 250 to compensate the possible guessing of real words. Nevertheless, it is important to note that test takers respond to the checklist test very differently, and there are even some national characteristics in relation to guesswork, some nations being more conservative in estimating their knowledge than others (Milton 2009: 73).

The X-Lex is the most suitable test for the present study because of its simplicity and clear format as well as the short amount of time needed to complete it. These features make the test easy to administer to a relatively large number of testees. The use of pseudo-words decreases the effect of guessing and since all the participants in this study are of the same nationality, national characteristics related to guesswork do not affect the results of the study.

4.2 Ways of assessing depth of vocabulary

Since the concept depth of vocabulary knowledge includes so many different aspects and there is no generally accepted definition of the term, it is quite difficult to measure this knowledge as a whole. There are many tests and test types, however, that try to measure this knowledge or at least some of its aspects. For instance, Read (2004: 214-216) mentions three test methods used to operationalise the quality of vocabulary knowledge. First, researchers can use tests that require more specific word knowledge from testees. An example of this type of test is Dolch and Leeds' (1953, cited in Read 2004: 214-215) multiple-choice questionnaire, which demands more than general understanding of the stimulus word:

A disaster is ruin which happens

- a. suddenly
- b. within a year's time
- c. to all people
- d. gradually

FIGURE 4 AN EXAMPLE ITEM OF DOLCH AND LEEDS' VOCABULARY TEST (READ 2004: 215)

Second type of depth test asks test takers to give explanations and definitions to stimulus words either orally or by selecting the right option from multiple-choice items. The quality of test takers' answers is then evaluated and scored. This test type has mainly been used to assess children's developing vocabulary in their L1 or L2.

Third test method is self-report tests. A well-known example of this is the Vocabulary Knowledge Scale (VKS) by Paribakht and Wesche (1993). The self-report scale has five categories to score word knowledge, which are ranging from I *I don't remember having seen this word before* to IV *I know this word* and V *I can use this word in the sentence*. A testee has to show his/her knowledge of a stimulus word by choosing the category that best suits his/her knowledge of that word. The first two categories test merely the receptive knowledge of the written form of the word, but in the last three categories testees have to show they know the meaning of the words and they have to demonstrate their ability to use the word in a sentence. (Nation and Webb 2011: 228-229.) Thus, the VKS combines different aspects of knowledge but does not measure any of them very thoroughly. Read (2000: 138) says that it is understandable that one simple scale cannot measure word knowledge very well because of the complexity of vocabulary knowledge. Wesche and Paribakht (1996, cited in Read 2000: 135) seem to agree with this and point out that their scale does not aim to measure general vocabulary knowledge but is designed to assess knowledge of specific words, which the test takers have recently been taught in a particular situation.

The three test methods mentioned by Read are the main ways to assess a person's understanding of words. There are, of course, other tools and tests, which can be used to measure vocabulary depth, as well. One just needs to choose which aspects of vocabulary depth s/he wants to assess. I decided to utilise the Word Associates

Test, which concentrates on knowledge of synonyms and collocations. I discuss the particular test further below.

Read's (1998, cited in Cobb 2015) Word Associates Test (WAT) measures test takers' depth of receptive vocabulary knowledge by focusing on semantic and collocational relations between words (Nassaji 2004: 114). The WAT assesses three aspects of Nation's list (see Table 1) of what is involved in knowing a word, that is, *form and meaning, concepts and referents* and *collocation* (Nation and Webb 2011: 227). The *Form and meaning* category is actually associated with breadth, whereas the latter two categories are about depth. Although the test is called Word Associates Test, it does not assess the knowledge of *associations* mentioned in Nation's list. The WAT includes 40 stimulus words, all of which are adjectives to avoid any confusion. All the stimulus words are followed by eight words, half of which associate with the stimulus word, while the other half are just distractors (see Figure 5). In addition, the eight words have been divided into two groups of four, adjectives and nouns. Read (2000: 184) explains that the associative adjectives in the first box are either synonyms or partial synonyms of the target word, which means that the adjectives and the target word have a paradigmatic relationship. The nouns in the second box, on the other hand, have a syntagmatic relationship with the stimulus word, that is, they tend to collocate with it. The frequency and the usefulness of the words in the language were the reasons why the particular words are included in the test (Read 2000: 185).

beautiful

<input type="checkbox"/> enjoyable	<input type="checkbox"/> expensive	<input type="checkbox"/> free	<input type="checkbox"/> loud	<input type="checkbox"/> education	<input type="checkbox"/> face	<input type="checkbox"/> music	<input type="checkbox"/> weather
------------------------------------	------------------------------------	-------------------------------	-------------------------------	------------------------------------	-------------------------------	--------------------------------	----------------------------------

FIGURE 5 AN EXAMPLE ITEM OF THE WAT (COBB 2015)

Read (2000: 184-185) advises test takers to identify one to three words in each box so that they choose a total number of four words per target word. Each right answer is scored with one point the maximum score being 160. Wrong answers do not decrease the score. Read (ibid.) chooses to vary the number of associates in each box in order to reduce guessing, but he admits that guessing still has an effect on test takers'

scores. He, however, highlights the importance of informing testees to choose exactly four associates per stimulus word; otherwise, even a native speaker might find only two or three associates and move on without finding all four, which might then lower the scores.

As Read (1993: 357) notes, it is thought that tests where words are presented in a context are better than tests with isolated words, since it is important for language learners to be able to recognise words in discourse. However, sometimes researchers have to compromise: introducing words in a context would demand testees to understand the context, which might cause problems if they cannot comprehend what they are reading. Testing contextualised items takes more time and that is why it may not be possible to get a coverage that is wide enough. For these reasons, the WAT measures vocabulary knowledge with isolated items.

The WAT is based on the thought that people with deeper knowledge of vocabulary have more information about words and are therefore better at finding the right associates of the words (Read 1993: 359). Good vocabulary knowledge is nevertheless not the only factor affecting the scores. "A certain resourcefulness in seeking possible associates, as well as the confidence to make guesses" are two other attributes affecting the scores, which Read (1993: 367) mentions. Depending on how the results of the test are interpreted, the latter two features can be either positive or negative. According to Bogaards (2000: 496), if one only wants to assess whether or not testees truly know the stimulus word and its associates, then guessing and resourcefulness are a problem and could cause overestimation. Nevertheless, if one wants to find out whether test takers are able to use different strategies to find relations between words, these features can be considered very profitable.

Some consider underestimating also a problem. Schmitt et al. (2011: 117-118) name two reasons for possible underestimation of the score: first, a testee has a fixation on "a particular meaning sense of a word and neglect other associates", and second, s/he is not able to "discriminate the best answers from possible options" especially in relation to syntagmatic options. This is, nonetheless, once again a matter of interpretation: instead of seeing these personal attributes as something that affects

the reliability of the test, it could be suggested that a person with such problems might not be very resourceful or good at deducing, which is why they do not score well in the test.

Read (2004: 223) wants to remind that the word associates format is just one way of measuring depth of vocabulary knowledge, albeit a rather promising one. As discussed in Section 3.2, vocabulary depth consists of several aspects of knowing a word. It is thus important to underline that even though the WAT is called a depth test, it only assesses some aspects of depth. Nevertheless, despite its possible deficits, Nation and Webb (2011: 227) state that the WAT is the best-known measure of depth of vocabulary knowledge, and several researchers have used the same format in their studies (e.g. Greidanus and Nienhuis 2001, Nassaji 2004, Akbarian 2010). According to Schmitt et al. (2011: 106), the reason for the popularity of this format is its “potential to economically measure learners’ familiarity with target words’ meanings and some of their uses as well”.

The present study decides to employ the WAT because, as the X-Lex test, it has a clear format and it does not take too much time to complete the test. Using decontextualized items also enables the test to cover a larger number of words than tests with contextualised words. I have, however, decided to not to follow the suggestion of Read in relation to the number of words test takers are told to choose. Because the participants of the present study are foreign language learners who have intermediate English language skills on average, I do not think it is meaningful to tell them to pick exactly four items per each stimulus word. It is probable that the participants will not recognise all the right associates with most of the words, and thus, asking everybody to identify four words would only increase guesswork.

4.3 Previous studies on vocabulary size and depth

Comparing the findings of different studies is quite difficult because of the variety of methods employed to assess vocabulary size and depth. Other variables, such as participants’ L1 and age, hours of instruction and the learning context, have an effect

on the results as well. It is thus important to pay attention to details when referring to previous studies. As the studies, which are presented below, demonstrate, there is quite a lot of variation in the results gained with size and depth tests. It is somewhat possible to compare scores of various size tests with each other when one just remembers to consider all the details. Comparing depth tests with each other is, nevertheless, quite impossible if they concentrate on measuring different aspects of depth. For these reasons, I will first deal with some studies that have utilised various tests to count the number of words learners know, after which I will go through studies all of which have used the WAT to measure vocabulary depth.

4.3.1 Studies on vocabulary size

Milton and Meara (1998) assessed 80 German EFL learners' receptive vocabulary size by using Meara's LLEX Lingua Vocabulary test, which could be considered some kind of a predecessor of the X-Lex test, since the idea behind both tests is the same. The participants were 14 to 15 years old and had received some 400 hours of instruction in EFL. Milton and Meara estimated the mean vocabulary size to be 1,200 lemmas. Milton (2006, cited in Milton 2009: 79-80) also studied Greek learners in a private EFL school from the beginner level to the upper-intermediate level, that is, the B2 level of CEFR. Milton, who utilised the X-Lex, reported that the participants learnt about 500 words, i.e., lemmas, every year. After seven years, which is equal to some 650 hours of tuition, the FCE level learners were familiar with approximately 3,500 words out of the most frequent 5,000.

Canga Alonso (2013) conducted a study of 92 Spanish EFL learners' vocabulary size and utilised the 2,000-word level of the Vocabulary Levels Test. The learners were 15 to 16 years of age and had received a total amount of 1,049 hours of tuition. Canga Alonso found out that there was a significant difference between the scores of male and female participants, males scoring an average of 1,028 lemmas and females an average of 854 lemmas. The average score of all learners was 935. He (2013: 70) pointed out that the vocabulary size of the participants was quite low compared to

that of EFL learners in other countries, who had received even fewer hours of instruction (e.g. Takala 1984, Milton and Meara 1998).

Takala (1984) assessed over 2,400 Finnish ninth graders' total receptive and productive vocabulary size. The participants had studied English for seven years, which is equal to about 450 hours of instruction. The learners were asked to provide Finnish or English equivalents of decontextualised stimulus words, which were all found in the two English textbooks used in the Finnish schools. Takala concluded that the fast learners, that is, the ones who studied in the advanced group, had a receptive and a productive vocabulary of some 1,500 words. The average learners had a vocabulary of around 900 words and the slow learners that of 450 words. The numbers refer to types of words. Takala went on to suggest that if the learners' limited word formation skills and ability to deduce word meanings in a context were taken into account, the advanced group could be said to have a receptive vocabulary of 2,200 words and an productive vocabulary of 2,000 words. The size of the average group's vocabulary would then be around 1,000 words, whereas the slow learners were not considered to have skills that would add to their number of words they know.

Vocabulary size has not been a big area of research in Finland. Thus, much has not been written about Finnish EFL learners' vocabulary knowledge. Takala's comprehensive study is one of the few studies conducted in Finland with relation to vocabulary size. Other, more recent studies have been interested in different aspects of vocabulary knowledge, for example productive vocabulary (e.g. Kallela 2014), or have focused on university level students (e.g. Makkonen 2008, Pirilä 2012).

4.3.2 Studies on vocabulary depth

Unlike vocabulary size that has received some attention in Finnish studies, vocabulary depth has not really been studied in Finland. Fortunately, there are studies conducted in other countries that give an idea of how learners score in the WAT. For instance, Qian (1999), who assessed the roles of size and depth in reading

comprehension in English as a second language context, used the WAT as one of the tools in his study. All 74 participants were adult academic ESL learners in Canada and everyone's vocabulary size was at least 3,000 word families, which is around 4,800 lemmas. The mean score the participants gained in the WAT was approximately 120 out of the possible 160.

Farvardin and Koosha (2011) conducted a study on the same subject as Qian, but with 78 Iranian university students, who majored in teaching English as a foreign language. Their mean score was 112 out of 160. Akbarian's (2010), who studied the relation between size and depth, participants were also Iranian university students but learnt English for specific purposes. The 112 students scored an average of about 51 out of 160. When the participants were divided into high and low proficient groups based on whether they mastered or did not master the Vocabulary Levels Test's 2,000-word level, the mean score of the high proficient group was 84, whereas the low proficient group had a mean score of 42.

Hellman (2011) was interested in the vocabulary size and depth of people who had started learning English a second language in adulthood but who now were highly proficient in English. She compared the non-native speakers to monolingual and bilingual native speakers and used the WAT as one of the tests employed. The WAT scores of the three groups did not differ significantly from each other: the non-native speakers' mean score was some 147, bilinguals' score 149 and monolinguals' score 151 out of 160. What is noteworthy, however, is that even native speakers did not score full marks. People think and use the language differently, which is why the connections between words, which the WAT suggests, do not seem so obvious to all people.

5 THE PRESENT STUDY

The present study aims to find out about Finnish EFL learners' size and depth of vocabulary knowledge. The research is quantitative in nature: the tests used to assess vocabulary knowledge in this study give scores in numbers and all the results are

therefore represented and compared numerically. Moreover, typical of a quantitative study is a great number of participants, which enables the use of statistical methods when analysing data and makes it possible to generalise the results (Vilkka 2007: 17). That is what the study also seeks to do. In addition, quantitative studies typically do not pay much attention to individuals' features and skills but are rather more interested in characteristics of a group of people (Dörnyei 2007: 33). That is to say, the present study mostly takes notice of the participants only as members of either of the two age groups. Apart from that, however, some participants who differ from the rest as regards their results or background are studied as separate cases.

In this chapter, I describe the present study in detail. Section 5.1 presents the research questions and hypotheses. Section 5.2 elaborates how the data was gathered and specifies the instruments used to collect the data. Section 5.3 gives a description of the participants and their background, and finally, in Section 5.4, the methods used to analyse the quantitative data are explained.

5.1 Research questions

Words have an integral role in communication and the larger the vocabulary and the more knowledge of the words, the better the ability to manage in diverse language situations. Since vocabulary knowledge is such an important part of language skills, it is obvious that foreign and second language learners need to develop their word knowledge in order to be able to use the language in various ways. Researchers worldwide have paid quite a lot of attention to different aspects of vocabulary learning and teaching. A lot is known about how language learners acquire new words, what is needed for new words to be learnt and what teachers could do to promote the learning. Vocabulary size and depth have also been subjects of study for some time, but they have not received as much attention. It would be very logical, however, to study size and depth together with other subjects associated with vocabulary. By measuring language learners' vocabulary knowledge, one can find out how well learning and teaching have actually succeeded and whether learners and teachers have reached their goals.

As discussed in Section 4.3, not much attention has been paid to Finnish EFL learners' vocabulary skills and their development. Thus, the present study wants to fill this void and aims to answer the following questions:

1. What is the size of Finnish EFL learners' receptive vocabulary and what can be said about their depth of vocabulary knowledge?
2. How do two age groups of Finnish learners differ from each other in their vocabulary size and depth?
3. How do the two aspects of vocabulary knowledge correlate with each other?

The object of the first question is to get information about the participants' vocabulary knowledge, in other words, to see how wide vocabularies the participants have and what and how much they know about words. Every learner is an individual and has his/her own way and pace of acquiring vocabulary knowledge, which is why it can be presumed that the scores of size and depth tests vary within the age groups. The main attention is, however, on the average scores of the groups. The second research question aims to find out how the two groups of Finnish EFL learners, i.e., ninth graders and two years older upper secondary school pupils, differ from each other in their vocabulary size and depth. It is hypothesised that the older group of EFL learners, which has received more hours of EFL tuition and is expected to be on a higher CEFR level, is likely to be ahead in both size and depth on average and so score better in both tests. Although the exact relation between vocabulary size and depth is unknown, studies have shown that the two aspects of vocabulary knowledge are likely to go hand in hand. That is, when vocabulary size increases, the knowledge of the known words grows as well. The third question addresses this issue related to the correlation of vocabulary size and depth.

5.2 Data collection

The data collection took place in a comprehensive school and in an upper secondary school in Central Finland in January 2016. The data was gathered using two

vocabulary tests, that is, the X-Lex test, which measures size of vocabulary, and the WAT, which assesses depth of vocabulary knowledge. The participants were informed about the aim of the present study and made clear that the test does not affect their grades or course evaluation. Before taking the tests, the participants were given instructions on how they should complete the tests. They also had the chance to ask about the instructions. The tests were treated anonymously, but the participants were asked to provide some background information (i.e. sex, first language, when they had started learning EFL and whether or not they had lived or otherwise spent a longer period in an English-speaking country or area) about themselves.

There are several ways to assess learners' language skills including vocabulary knowledge. For example, researchers and teachers can give learners tests to measure certain aspects of their language knowledge, ask them to evaluate their own performance and language proficiency or observe their work in class in order to get information about their various language skills (Huhta and Tarnanen 2011: 201). Nevertheless, conducting vocabulary tests was found the only suitable way of collecting the type of quantitative data needed to answer the research questions. It would not have been possible to get this sort of and a sufficient amount of quantitative data using surveys or interviews, for instance.

In the tests used by the present study the test takers had to tick the words they recognised (X-Lex) and the words they thought had the same meaning or appear with the given word (WAT). These tests are not dependable on any syllabi but are based on the usefulness and importance of the frequent words, and they measure the test takers' vocabulary knowledge in general (Huhta and Tarnanen 2011: 202). Other test types, such as cloze tests or differing multiple-choice questionnaires, could be used to measure vocabulary knowledge as well (Martin 2011: 168). As Martin (*ibid.*) notes, however, one has to weigh up which of the options corresponds best with the aims of the study, and how it would be possible to get valid and reliable information in a sensible way. The X-Lex and WAT met the criteria of the present study best.

The X-Lex and WAT have already been elaborated in Chapter 4. However, the tests and their instructions had to be slightly altered because of the limitations and requirements of the present study. Firstly, the tests were conducted in 45-minute lessons using pen and paper. To make certain that every participant had time to finish both tests, the tests could not have included too many items. If there had been too many items, the participants could have also been more likely to lose their concentration and interest and might not have finished the tests properly. It is often recommended that when the X-Lex is used to estimate vocabulary size, more than one and preferably three tests should be used to get results that are more reliable. In this case, only one X-Lex test was included in the test paper due to the limited time. Furthermore, the WAT, which usually consists of 40 items, was reduced to include only the first 20 items of the original test for the same reason. This means that the maximum score in the X-Lex test was still 5,000, but in the WAT 80, instead of 160.

Secondly, in order to avoid possible misunderstandings, all instructions were given in Finnish instead of English. This way it was more probable that the participants understood what they were expected to do in the tests. In addition, a lot of time was used to make the instructions as clear and comprehensive as possible. The X-Lex instructions followed more or less the original instructions, but the instructions of the WAT were somewhat revised, as mentioned in Section 4.2. I am of the opinion that the guesswork would only have increased and thus affected the scores if the participants, who were intermediate English learners, would have been asked to find exactly four right answers per each stimulus word. The participants were therefore told to tick as many words as they thought met the given criteria as long as they ticked four words at most. The participants were also told that there was a maximum of four right options per item and the right options could be divided between the two sides whichever way.

The instructions as well as the actual tests were piloted in December 2015 with two people, who represented the same age groups as the participants. After the pilot test, some adjustments were made to somewhat simplify the instructions but overall, the pilot testers found the tests very comprehensible.

5.3 Participants

In the present study, there were participants from two age groups. Comprehensive school ninth graders (N = 59) aged 15 to 16 formed the younger group. Majority (80%) of the ninth graders had started learning EFL in the third grade, which means they had received EFL tuition for about 6.5 years. This is approximately equivalent to 550 lessons or 413 hours (Ministry of Education and Culture 2012). The other group consisted of second-year pupils in upper secondary school (N = 71), who were 17 to 18 years of age. 93% of them had also learnt English since the third grade. In other words, they had learnt EFL for 8.5 years, which is equal to 760 lessons, i.e. 570 hours (Ministry of Education and Culture 2014).

These two age groups were chosen to be the subjects of the study because both groups were soon expected to reach certain CEFR levels. By the end of grade 9, the younger group would have received 608 lessons (i.e. 456 hours) of instruction in EFL. The aim of EFL instruction is that by the end of the ninth grade as many pupils as possible will reach the target level set in the national core curriculum. The goal is to attain the B1.1 level of the CEFR in three different skill categories, which are “the ability to interact with other people”, “the ability to read or interpret texts” and “the ability to write or produce texts” (Finnish National Board of Education 2015a: 351-352). As discussed in Section 3.1, the B1 level is a so-called threshold level, which refers to B1 learners' ability to use the language somewhat independently (Council of Europe 2001: 23-24). The upper secondary school pupils (USS pupils), on the other hand, are expected to reach the level B2.1 in the same skill categories after they have finished all six compulsory EFL courses (Finnish National Board of Education 2015b: 108). The older participants of the study had thus about two EFL courses left, which equals 76 lessons or 57 hours, before they were expected to be at that B2.1 level (Ministry of Education and Culture 2014). I was therefore interested in finding out how ninth graders' and USS pupils' lexical knowledge would correspond with their expected language levels, and whether the relation of pupils' receptive vocabulary size measured with the X-Lex test and CEFR levels would match as Milton and Meara (2003, cited in Milton 2009: 186) have suggested (see Table 2).

The total number of participants in the present study was 133, out of which 62 were ninth graders and 71 USS pupils. The tests of three ninth graders had to be left out from the results, however, because they had not either followed the instructions carefully enough or completed the entire test paper, which would have had an effect on the results. In other words, the analysed data comprises the tests of 130 participants, 59 ninth graders and 71 USS pupils. All background information reported by the participants can be found in Table 4.

TABLE 4 NINTH GRADERS' AND USS PUPILS' BACKGROUND

		N (%)	N (%)
Ninth graders (N = 59)	sex	male 26 (44.1)	female 33 (55.9)
	L1	Finnish 58 (98.3)	bilingual 1 (1.7)
	English instruction	before third grade 6 (10.2)	since third or fourth grade 53 (89.8)
	abroad	yes 5 (8.5)	no 54 (91.5)
USS pupils (N = 71)	sex	male 26 (36.6)	female 45 (63.4)
	L1	Finnish 70 (98.6)	bilingual 1 (1.4)
	English instruction	before third grade 3 (4.2)	since third or fourth grade 68 (95.8)
	abroad	yes 8 (11.3)	no 63 (88.7)

5.4 Methods of analysis

The X-Lex, which was used to measure the participants' vocabulary size, and the WAT, which was used to assess the participants' depth of vocabulary knowledge, were treated separately and every participant was given separate scores from both tests. Overall, all the participants were given four different scores. First score was the vocabulary size score. The maximum score of the test was 5,000, since the X-Lex test aims to measure the knowledge of the 5,000 most frequent words. There was a total of 100 actual words and 20 pseudo-words mixed up in the test. That is, every real word ticked by the participants was worth 50. Each pseudo-word ticked, on the other hand, lowered the score by 250. The second score was the vocabulary depth score

with the maximum of 80 points. Every right answer was rewarded with one point, while wrong answers did not influence the score. The third and fourth scores were sub-scores of the depth score, that is, synonymy score and collocation score. These sub-scores were based on the two types of possible answers the stimulus words had in the WAT. That is to say, the right words ticked from the left-hand side of each stimulus word formed the synonymy score, while ticking right words from the right-hand side added to the collocation score. The maximum score for the synonymy part was 39 and that for the collocation part was 41. I decided to calculate these two sub-scores to get more precise information about the participants' word knowledge, and to find out if there were such qualitative differences between the age groups that could be seen between language users at differing proficiency levels (see Section 3.3).

When counting the scores, two issues needed to be considered. Firstly, there was one pseudo-word in the X-Lex test that had to be left aside when scoring the test. It turned out that 20% of the 130 participants had ticked the pseudo-word *warboy*. This could be explained by the participants' first language, i.e. Finnish, which has a word that could be considered a direct translation of the pseudo-word *warboy*, that is, *sotapoika*. It is obvious that this Finnish equivalent of the pseudo-word has caused confusion among the participants. I therefore decided to ignore *warboy*, because it would not have been reasonable to subtract 250 from the scores of 20% of the test takers because of it. I believe that it would have falsified the results of the test.

Secondly, although the testees were clearly advised to tick four words per stimulus word at most in the WAT, some had occasionally ticked five words per stimulus word. Such incidents were rather rare, however: only eight persons had done so with ten items altogether. Since there were 20 items in the test, these participants did not considerably benefit from their actions, which they probably did by accident, and thus, I did not see any reason to exclude their results from the study.

After scoring each test paper, the data was entered into SPSS, which was used to statistically analyse the quantitative data. To find out how the results inside the two age groups varied and to compare the two groups with each other, the mean values and standard deviations in relation to all four scores, or variables, were calculated for

both groups. The distributions of the variables were calculated with the One-sample Kolmogorov-Smirnov test, which indicated that the depth scores and the synonymy scores followed a normal distribution, and the collocation scores were also very close to a parametric distribution, whereas the size scores were not normally distributed. The Independent-samples t-test was used to compare the differences between the groups' scores. The t-test for independent samples is a method that compares the means of two different groups while considering the variability of the data. It is typically applied to normally distributed data. (Rasinger 2008: 179, 182.) Regardless of the non-parametric distribution of some of the data, the relatively large sample size enabled the use of the t-test in the present study, since the t-distribution can be thought to be close to normal (Johnson 2008: 50). What is more, to compare the synonymy score with the collocation score within the groups to discover whether the differences between the scores were statistically significant, paired-samples t-test test was applied. This t-test can be used to compare two sets of data of one group, instead of comparing the data of two different groups (Rasinger 2008: 185-186). Lastly, the Pearson correlation coefficients were defined to see how the vocabulary size scores correlated with the depth scores and, additionally, with the sub-scores.

6 FINNISH EFL LEARNERS' VOCABULARY KNOWLEDGE

This chapter presents and discusses the results of the analysed data with relation to the research questions and the theoretical background, which is linked to the results whenever relevant. The results will be covered in the following order: To start with, Section 6.1 deals with the results of the vocabulary size test. Means and other significant values of both age groups are first presented separately, after which the groups are compared with each other. Next, the main concern of Section 6.2 is the results of the vocabulary depth test. The results will be treated in the same way and order as the results of the size test, apart from the fact that there are three variables related to vocabulary depth in this study. That is, the total score of the depth test is the first variable, the synonymy score is the second and the collocation score is the third variable. All of these are used to assess the participants' knowledge of words

and to find differences between and within the age groups. Since the relationship between size and depth of vocabulary knowledge has been widely discussed, the present study wanted to contribute to the discussion as well. For this reason, the correlation between the two aspects of vocabulary knowledge will be covered in Section 6.3. Moreover, Section 6.4 will briefly deal with such participants who had received more or fewer hours of EFL tuition than the average learner of the same age. Lastly, Section 6.5 summarises the results of the study.

6.1 Vocabulary size

One of the aims of the present study was to find out what the Finnish EFL learners' receptive vocabulary size is and how the two age groups differ from each other in this respect. It was expected that there are differences between the learners of the same age, because learners are always individuals who learn in their own way and at their own pace. Furthermore, it was hypothesised that the older group is likely to be ahead as far as vocabulary size is concerned, since they have learnt English for a longer period, which in this case was two years on average. Because the USS pupils were also expected to be at a higher CEFR level, they should, according to Milton and Meara (2003, cited in Milton 2009: 186), score better in the X-Lex test, in other words, have a bigger receptive vocabulary. To discover how much vocabulary size truly varied among the learners of the same age and between the age groups, the mean scores and standard deviations as well as minimum and maximum scores were scrutinised. At this point, it is good to emphasise the fact that the X-Lex scores do not tell the absolute number of words the participants knew receptively because the test aims to measure the knowledge of the most frequent words only. It is very probable that the learners also knew other, less frequent words, and thus, the actual size of their vocabularies was somewhat bigger.

The ninth graders, who formed the younger group of the study, achieved a mean score of 3,613.56 in the X-Lex. This refers to the number of words the ninth graders knew receptively out of the 5,000 most frequent words on average. As expected, there was a lot of variation in the scores, however. The lowest score was 1,750 and

the top score gained in the test was 4,900, while the standard deviation of the variable was 877.095. All values are presented in Table 5 along with the corresponding values of the group of USS pupils.

TABLE 5 MEANS AND STANDARD DEVIATIONS

	N	Mean	Standard deviation	Minimum	Maximum
Ninth graders	59	3,613.56	877.095	1,750	4,900
USS pupils	71	4,145.77	630.858	2,000	5,000

When the ninth graders' mean score was checked against Milton and Meara's table that links the CEFR levels to X-Lex scores (see Table 2), it was noticed that the ninth graders' vocabulary size, which was some 3,600, matched the CEFR level B2. If the range of the ninth graders' vocabulary size was considered with relation to the CEFR, their language levels varied from A2 to C2. However, as discussed in the previous chapter, ninth graders are usually expected to achieve the B1.1 level, which would require a vocabulary of at least 2,500 words. Based on the X-Lex results, only ten out of 59 had not yet reached that level, but it was still possible that they could so before the end of the ninth grade. Thus, the ninth graders' average vocabulary size was substantially larger than what could have been anticipated based on their expected language level. It is difficult to say, however, whether the pupils' vocabulary just did not correspond with their proficiency level, or whether they in deed were more advanced language learners than the national core curriculum expected them to be. It is conceivable that some of the pupils had already reached the C2 level of the CEFR, but the proportion of those learners was so high that it raises some doubts about the accuracy of the relation of the X-Lex scores and the CEFR suggested by Meara and Milton.

Thirteen ninth graders, in other words, 22% of the group, got the score of 4,500 or higher, which means that these pupils could have been considered C2 level learners based on their X-Lex scores. I was interested in finding out whether there were factors in these pupils' backgrounds that could have explained their success in the

size test. I took three factors into consideration; the pupils' L1, when they had started learning EFL and whether or not they had lived or otherwise spent a longer period in an English-speaking country or area. I compared the figures of the top scorers with those of the entire group the top scorers included. First language did not play a role because Finnish was the L1 of all the top scorers. None of them had lived or spent time in the English-speaking world, and hence that factor did not explain their good results. Three people (23%) reported they had started learning English before the third grade, while ten people (77%) had learnt it since the third or fourth grade. These figures slightly differ from those of the entire group of which 90% had started in the third grade or later and only 10% before the third grade (see Table 4).

Although a somewhat larger proportion of the top scorers had learnt English longer, the background of the top scorers did not remarkably differ from that of the entire group. It can therefore be concluded that the three factors did not explain the highest scores of the group but there are other factors, such as learning skills and hobbies, which explain why some scored better than the others did. Learning English longer (than others) and spending time among native speakers definitely contribute to a person's vocabulary size. The ninth graders who had such a background had surely benefitted from it and it had probably helped them extend their vocabulary beyond what it would have been without such experiences. As the present study shows, however, it is not necessary to have such a background to possess a large vocabulary.

Takala (1984), who conducted a study of ninth graders' EFL vocabulary size at the end of the school year, estimated that the learners in the advanced group had a receptive vocabulary of around 2,200 words after seven years or 450 hours of EFL tuition. The ninth graders studied by Takala had had a few more hours of tuition by the time of the study than the ninth graders of the present study, who had received some 410 hours of EFL instruction. Takala's results and the results of the present study differ significantly from each other, part of which is certainly due to the dissimilar instruments used to measure the ninth graders' vocabulary size (see Takala's study in Section 4.3.1). Nevertheless, it seems very much likely that the average vocabulary size of ninth graders has risen in the past years. Such a development would not be surprising considering that the English language has a

much greater role in Finnish society than some thirty years ago when Takala conducted his study. Now, EFL learners have more chances to encounter and use English outside the classroom than they did in the 1980s. Even the pupils who could have been considered weaker learners in this study and got the lowest scores in the X-Lex had a considerably larger vocabulary than the so-called slow learners in Takala's test, who knew an average of 450 words only. When compared to studies conducted in other countries with EFL learners, who were approximately of the same age, the mean score of the ninth graders in the present study was relatively high, as can be seen in Table 6. This issue will be returned to in the discussion.

TABLE 6 VOCABULARY SIZE MEASURED IN VARIOUS STUDIES

	First language	Age	Hours of tuition	Vocabulary size
Lahtikallio (2016)	Finnish	15-16	410	3,600
Takala (1984)	Finnish	15-16	450	2,200
Milton and Meara (1998)	German	14-15	400	1,200
Milton (2006)	Greek	?	650	3,500
Canga Alonso (2013)	Spanish	15-16	1,050	935

The mean vocabulary size of the ninth graders was also contrasted with the estimates of vocabulary size needed to comprehend certain types of texts. The mean score of the pupils was around 3,600 lemmas, which was still quite far away from the number of words needed to reach the threshold of 98% suggested by Nation (2006), and Laufer and Ravenhorst-Kalovski (2010). As explained in Section 3.1, understanding 98% of any written texts, which would enable a learner to be functionally independent, requires a knowledge of some 8,000 to 9,000 words families, which equals 12,000 to 14,400 lemmas. However, one has to take into consideration that the X-Lex does not measure the knowledge of all words but the knowledge of the 5,000 most frequent ones. The ninth graders also knew less frequent words, which means that their total vocabulary size was somewhat larger than what the X-Lex scores

suggested. It is rather unlikely, though, that the ninth graders would have had a vocabulary of 12,000 lemmas or more, but not entirely impossible. Furthermore, it has been suggested that to know the 2,000 most frequent lemmas would enable one to comprehend 80% of any written texts. In relation to this, the ninth graders' average score looks very good: with a vocabulary of 3,600 words, a learner should be able to understand the main points of any text type. On the other hand, even a smaller vocabulary size is sufficient with spoken texts. A vocabulary of about 1,600 lemmas should cover 80% of context-governed texts, and understanding 80% of conversational speech only requires roughly 320 lemmas. In conclusion, it seems that the ninth graders' receptive vocabulary size gives them a good basis for understanding both written and spoken texts. The same applies to the USS pupils whose scores will be introduced next.

When the values associated with the vocabulary size test were calculated for the upper secondary school pupils, who comprised the older age group of the study, the mean score was 4,145.77 out of 5,000. Naturally, the scores varied a lot also within this group; the standard deviation of the variable was 630.858 and the pupils' scores ranged from 2,000 to 5,000 (see Table 5). As with the ninth graders, attention was paid to the background of those USS pupils who scored 4,500 or more to find out whether there were factors that would have explained their high scores. The number of top scorers was 28 (39% of the group) and their L1 was Finnish. All of them had received English instruction since third or fourth grade, and only three people (11%) had spent some time in the English-speaking world. That is to say, the top scorers did not differ from the entire group in this respect (see Table 4). In this case, as well, the top scorers' success in the X-Lex test was a result of factors other than the ones covered in the present study.

Again, if adapted to the CEFR scale, the mean value suggested that the USS pupils were already at the C1 level, even though the curriculum only expected them to achieve the B2.1 level next in interacting with other people and in reading and writing texts (Finnish National Board of Education 2015a: 351-352). Only six people had not yet reached the B2 level based on their scores. Considering the variation of

the scores, the USS pupils' language levels also ranged between A2 and C2. Once again, the proportion of the pupils whose X-Lex scores had indicated that they were at the C2 level was high. Naturally, it could be expected that a greater number of the older pupils would have reached the C2 level than of the younger pupils, but it is still difficult to believe that even 39% of the USS pupils would already have been so advanced language learners. That is to say, it appears that the older pupils, as well as the younger ones, had a wider receptive vocabulary than what could have been supposed based on the CEFR levels they were expected to be or reach soon, but it is somewhat implausible that their vocabulary size scores and CEFR levels would correspond the way Meara and Milton's table claims (see Table 2).

The hypothesis regarding the vocabulary sizes of the two groups was that the older group would have more words in their vocabulary. As the figures show, (see Table 5), this hypothesis was confirmed since the USS pupils achieved a higher mean score than the ninth graders. The difference between the means was 532.215, in other words, the older pupils knew approximately 530 words more of the 5,000 most frequent words than the younger pupils. This difference was found statistically highly significant, $p < 0.001$ (see Table 7). Furthermore, both the lowest and the highest scores of the USS pupils were higher than those of the ninth graders. The older group also had a lower standard deviation meaning that the group was somewhat more consistent as far as the individual pupils' vocabulary sizes were concerned. There were no noteworthy differences in the backgrounds of the two groups so no such factors influenced the results.

TABLE 7 SIGNIFICANCE OF MEAN DIFFERENCE

	t	Sig. (2-tailed)	Mean difference
Vocabulary size	-3.898	0.000***	-532.215

*** Statistically significant at the 0.001 level.

Although the results show that the vocabulary sizes of the two different age groups deviated from each other so that the older group had a larger vocabulary, it is not possible to draw definite conclusions on the development of vocabulary size or the

pace of such a development. Learners' vocabulary should be investigated in a longitudinal study to truly gain information about the development of word knowledge. As it was hypothesised, however, the older pupils, who had received approximately 570 hours of EFL instruction, had on average more words in their receptive vocabulary than the two years younger pupils, who had had 413 hours of tuition in EFL. If the size scores were divided by the hours of EFL instruction the groups had received, the ninth graders would have learnt approximately 8.7 words per hour of instruction and the USS pupils 7.3 words per hour of instruction. That is to say, the average pace of acquiring new words had been a bit slower among the older pupils, which might support the view that more advanced learners develop their depth at the expense of size (see Section 3.3). Nonetheless, it must be underlined that the scores of single pupils varied a lot, which means that there were several ninth graders who scored better in the size test than many of the USS pupils in spite of the number of EFL lessons attended. This conforms to the common view that individual differences and interests have a great impact on language learning.

6.2 Vocabulary depth

Apart from vocabulary size, the present study aimed to explore the vocabulary depth of Finnish EFL learners by using the WAT. All participants were given three scores: a total depth score, a synonymy and a collocation score. The means of the scores and other relevant values were calculated for both age groups and they are represented in Table 8.

The ninth graders' mean depth score was 36.02 out of 80, which equals 45%. This is the average number of right answers the pupils gave, or rather selected, in the WAT test. Overall, the scores varied between 13 and 64 points, while the standard deviation was 12.707. Judging from these figures, there were great differences between individual pupils and their vocabulary depth. Since the WAT measures two different types of word knowledge included in the concept of vocabulary depth, the depth test was further divided into two parts, that is, into synonyms and collocations, when the scores were calculated. The synonymy score of course tells the

number of synonyms the participants were able to link to the stimulus words, whereas the collocation score matches the number of rightly chosen collocates. The ninth graders' mean scores of synonyms and collocations were 16.90 points out of 39 and 19.12 out of the possible 41, respectively. In other words, they recognised 43% of the synonyms and 47% of the right collocates the stimulus words had. The synonyms therefore formed 47% and the collocates 53% of the total depth score, which indicates that the ninth graders' knowledge of collocations was somewhat better than their knowledge of synonyms.

TABLE 8 MEANS AND STANDARD DEVIATIONS

		Depth score	Synonymy score	Collocation score
Ninth graders (N = 59)	Mean	36.02	16.90	19.12
	Std. deviation	12.707	6.583	6.856
	Minimum	13	5	6
	Maximum	64	33	33
USS pupils (N = 71)	Mean	42.75	20.76	21.99
	Std. deviation	11.494	6.018	6.015
	Minimum	14	4	10
	Maximum	63	32	33

That is to say, the pupils recognised more syntagmatic associations (that is, collocates) between words than paradigmatic ones (i.e. synonyms). The paired-samples t-test showed that the difference between the synonymy and collocation scores was extremely significant, $p < 0.001$ (see Table 9), which means that syntagmatic responses can be seen to have been more common among the group than paradigmatic responses. As it was discussed in Section 3.3, language learners' responses to stimulus words can vary depending on how advanced the learners are and what they already know about the particular words. Syntagmatic responses are typical of older L1 children, whereas L1 adults, who are more advanced language users, tend to produce paradigmatic responses (Meara 1983: 1-2). The same can be

applied to L2 learners; the more advanced a learner the greater the proportion of paradigmatic responses (Singleton 2000: 182). The ninth graders recognised more syntagmatic associations on average, as discovered. Nevertheless, not much can be deduced from the ratio of the synonymy score to the collocation score, because it is not known what the ratio would be if an L1 children or adult did the test and the same scores were calculated. To get some perspective, the ratio must be compared to that of the other group, which will be done later in this section.

In total, 68% of the ninth graders had a higher collocation score than a synonymy score, while 29% had a better synonymy score. Three percent scored equally well in both categories. Too much attention should not be paid to these figures, however. Since it is possible, for example, to score high in both categories, but have a slightly better collocation score, while someone else has much lower scores overall but happens to gain more synonymy points. Thus, the pupils with a higher synonymy score were not necessarily more advanced learners compared to the rest of the group. The figures only tell what a pupil's score is compared to his/her other score. For this reason, it is better to concentrate on the mean scores and their difference.

As for the USS pupils, they scored an average of 42.75 out of 80 in the depth test (see Table 8). That is, they identified 53% of the correct answers. The standard deviation was 11.494, while the individual scores ranged from 14 to 63 points. On this account, it is clear that there were significant differences in vocabulary depth among the older pupils as well. Furthermore, the mean synonymy score of the older age group was 20.76 out of 39, which is equivalent to 53% of the maximum score, and their collocation score 21.99 out of 41, in other words, 54%. Additionally, the number of synonyms correctly linked to the stimulus words corresponds to 49% of the total depth score and the collocates are responsible for the remaining 51%. The USS pupils' collocational knowledge was therefore slightly better than their knowledge of synonyms, in other words, the proportion of syntagmatic associations was a bit bigger than that of paradigmatic responses. According to the paired-samples t-test, the difference of the two sub-scores was statistically significant, $p < 0.01$ (see Table 9).

All in all, 58% of the older pupils knew more collocates than synonyms, 35% scored vice versa and 7% recognised equally many synonyms as collocates.

TABLE 9 SIGNIFICANCE OF THE DIFFERENCE BETWEEN SYNONYMY AND COLLOCATION SCORES

	Mean	t	Sig. (2-tailed)
ninth graders	-2.220	-3.891	0.000***
USS pupils	-1.225	-2.898	0.005**

** Statistically significant at the 0.01 level.

*** Statistically significant at the 0.001 level.

It was hypothesised that the older age group was likely to score better not only in the X-Lex test but in the WAT as well, because they had received more EFL instruction than the younger group and because vocabulary size and depth tend to grow the longer the language is learnt. The figures (see Table 8) reveal that there were differences between the two groups so that the older pupils had higher average scores. The USS pupils' mean depth score was 6.73 points better than what the ninth graders got, and the standard deviation was somewhat smaller than that of the younger pupils. It should be noted, nonetheless, that the ninth graders gained both the lowest (13) and the highest (64) depth score of all, although they differed from the USS pupils' lowest (14) and highest (63) scores only by one point. This supports the idea that there are other factors contributing to individuals' vocabulary learning than the number of hours of EFL tuition. Younger pupils can perform in the WAT test as good as or even better than older pupils, which many of the ninth graders did in the present study. Despite the variation between individual pupils, the older pupils had, nevertheless, a better average depth score, and the difference between the mean scores of the two age groups was statistically very significant, $p < 0.01$. The relevant numbers are presented in Table 10.

TABLE 10 SIGNIFICANCE OF MEAN DIFFERENCES BETWEEN GROUPS

	t	Sig. (2-tailed)	Mean difference
depth score	-3.168	0.002**	-6.730
synonymy score	-3.491	0.001**	-3.862
collocation score	-2.539	0.012*	-2.867

* Statistically significant at the 0.05 level.

** Statistically significant at the 0.01 level.

To get more precise information about the dissimilarity of the two groups' vocabulary depth, the sub-scores were also compared (see Table 10). The USS pupils were ahead in both word knowledge categories: the mean difference of the synonymy scores was 3.862 points and the mean collocation scores differed by 2.867 points. The former difference was statistically significant at the level $p < 0.01$, and the latter at the level $p < 0.05$. If the percentage figures are compared, a rather clear distinction between the groups can be noticed: when the ninth graders recognised 43% of the right synonyms, the USS pupils recognised 53%, and when the younger pupils found 47% of the collocates, the older pupils found 54% of them. That is, the older group knew 10% more of the synonyms and 7% more of the collocates than the younger group.

The proportions of the sub-scores out of the total depth score were also compared and there was a slight disparity between the groups. 47% of the ninth graders' total depth score was synonymy points while 53% were collocation points. The corresponding percentage figures of the USS pupils were 49% and 51%, which means there was a two-percent difference in both categories between the groups. In other words, the USS pupils had more knowledge of both synonyms and collocates than the ninth graders but the relative proportion of the collocation score, or syntagmatic responses, in relation to the synonymy score, i.e., paradigmatic responses, was smaller than among the younger pupils. It therefore seems that, as presupposed, the older learners who had studied the language longer had more knowledge of words but the vocabulary knowledge they possessed was also somewhat different.

Although it is impossible to reach conclusions about the development of vocabulary depth based on the present study, it appears possible that the more advanced EFL learners might have had more paradigmatic associations in their vocabulary compared to younger and less-advanced learners. This is in line with the suggestion that paradigmatic associations become more common when a learner advances with the language and learns more about words. In the present study, the dissimilarity of the proportions of the sub-scores was, of course, rather small.

In order to compare the WAT results of the present study with results gained from other studies using the same test, two aspects need to be taken into account. First, the depth test conducted in this study consisted of 20 items instead of 40. For this reason, the results are not directly comparable with other results. Turning the points into percentages enables contrasting results of different studies to some extent. Secondly, it must be remembered that the instructions used in this study differed from the common instructions as regards the number of words participants are told to tick and the possible ways the right answers might be divided between the two sides (see Section 5.2). The participants were told to tick four words at most, while generally participants have to tick exactly four words. They were also told that the right answers could have been divided between the two sides whichever way so that they had even more options to choose from. It is therefore likely that the participants of the present study gained somewhat lower scores than what they might have had if the test had followed the original and generally used instructions.

In the present study, the ninth graders identified 45% of the correct answers and the USS pupils 53%. Since the studies, which have utilised the WAT, have not had participants of the same age, the ninth graders and the USS pupils must be compared with older language learners and speakers. For instance, Iranian university students learning English for specific purposes in Akbarian's (2010, see Section 4.3.2) study scored 32% on average. When the group was divided into high and low proficient learners, based on their vocabulary size, the scores were 53% and 26%, respectively. Compared to the Iranian university students who had also learnt English as a foreign language, the figures of the Finnish EFL learners were very good considering that all

the Finnish participants were comprehensive and upper secondary school pupils and not university students. However, the Finnish pupils had a larger vocabulary compared to the Iranian students, out of whom only two had a vocabulary of 3,000 lemmas or more (Akbarian 2010: 399). In Hellmann's (2011, see Section 4.3.2) study, on the other hand, the adult participants were very proficient ESL or L1 speakers. The ESL speakers scored 92%, while monolingual L1 speakers scored 94% on average. These scores are much higher than those of the Finnish EFL learners are, which is understandable considering the differing status and the importance of the language in EFL, ESL and L1 speakers' lives. It is good to emphasise that, for the reasons already mentioned in Section 4.3.2, even native speakers do not score 100% in the WAT, which is why such results should not be expected from EFL learners either.

6.3 Correlations

Vocabulary size and depth are traditionally seen as two separate dimensions of vocabulary knowledge. Nevertheless, many researchers (e.g. Vermeer 2001; Qian 1999; Nurweni and Read 1999) have studied their relationship and discovered a strong correlation between the two aspects (Section 3.3). I therefore wanted to find out whether the participants' size and depth test scores correlated, and used the Pearson correlation coefficient as a measure. The correlation test revealed a positive linear correlation between vocabulary size and depth, $r = 0.767$ (see Figure 6). This indicates that the participants' vocabulary size and depth of vocabulary knowledge were closely related. The larger the vocabulary, the more knowledge of words the learners had. Nevertheless, the linear correlation was not perfect, since $r < 1$. In other words, there was some variation in the scores and not everyone conformed to the rule. A good score in the size test did not automatically mean a good score in the depth test or vice versa, although it was rather common. This is understandable because learners differ from each other in how they acquire word knowledge and how their lexicon is organised (see Section 3.3).

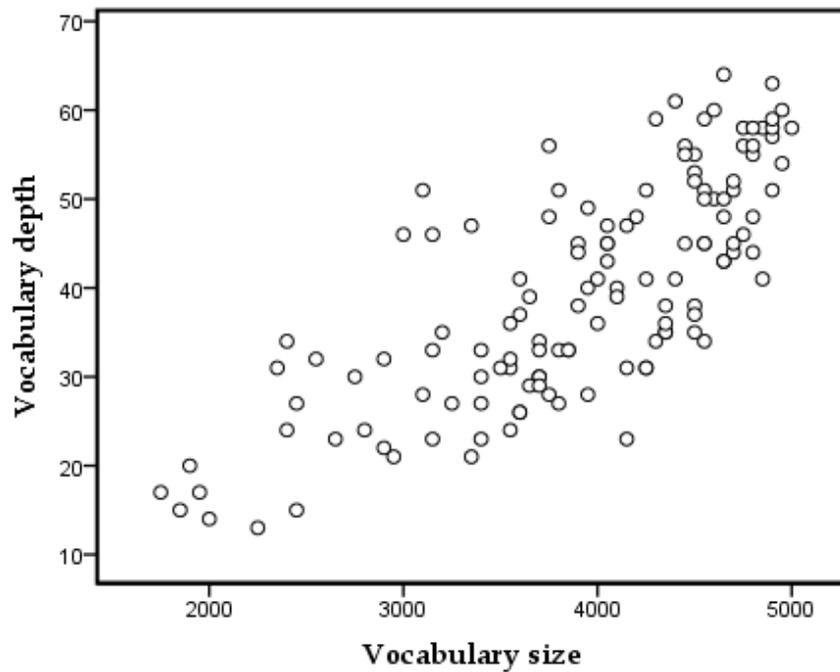


FIGURE 6 LINEAR CORRELATION BETWEEN SIZE AND DEPTH

Correlation coefficients were also calculated for the two sub-scores, that is, the synonymy and collocation scores, to see if these two dissimilar types of word knowledge correlated differently with vocabulary size. There was a minor difference between the coefficients: the synonymy score correlated a bit stronger with the size score than the collocation score the correlation coefficients being 0.741 and 0.721, respectively. The difference was not significant, however. All values concerning correlations of the entire group of participants can be found in Table 11.

TABLE 11 CORRELATION COEFFICIENTS BETWEEN SIZE AND THREE OTHER SCORES

	(N = 130)	Vocabulary depth	Synonymy	Collocation
Vocabulary size	Pearson correlation	0.767**	0.741**	0.721**
	Sig. (2-tailed)	0.000	0.000	0.000

** Correlation is significant at the 0.01 level (2-tailed).

As discussed in Section 3.3, it has been suggested that the correlation between size and depth might be stronger the more advanced the learners are. In other words, advanced learners' vocabulary size and depth might be more closely related than those of less proficient learners. Thus, to find out if this applies to the participants of

the present study, separate correlation coefficients were calculated for both age groups. As can be seen from Table 12, the ninth graders' all three correlation coefficients were somewhat higher than the coefficients introduced above concerning the entire group. By contrast, the correlations of the USS pupils' scores were lower than those of the whole group, and therefore lower than the coefficients of the ninth graders. For instance, the ninth graders' correlation coefficient for size and depth was 0.803, while the equivalent figure of the USS pupils was 0.684. Judging from these results, the ninth graders' vocabulary size and depth correlated more strongly than the corresponding vocabulary dimensions of the USS pupils, although the opposite was expected based on previous research findings. Nevertheless, the difference was not statistically significant. The figures also indicated a small difference between the groups in relation to the correlations of size and synonym and collocational knowledge. The ninth graders' vocabulary size correlated more strongly with their collocation score (0.770) than with their synonymy score (0.748), while the correlation between the USS pupils' vocabulary size and synonymy score (0.681) was stronger than the correlation between their vocabulary size and collocation score (0.625), but once again, the differences were not significant.

TABLE 12 CORRELATION COEFFICIENTS FOR BOTH AGE GROUPS

		Vocabulary depth	Synonymy	Collocation
Vocabulary size (ninth graders, N = 59)	Pearson correlation	0.803**	0.748**	0.770**
	Sig. (2-tailed)	0.000	0.000	0.000
Vocabulary size (USS pupils, N = 71)	Pearson correlation	0.684**	0.681**	0.625**
	Sig. (2-tailed)	0.000	0.000	0.000

** Correlation is significant at the 0.01 level (2-tailed).

6.4 Pupils with a longer or shorter history of EFL instruction

Because of the national core curricula and the policy of schools, pupils have conventionally started learning EFL in the third grade. This applied to the majority of the participants, i.e., 87%, which is why the third grade start was considered the norm in the present study. However, there were some exceptions within both age groups yet divergent starting times were more common among the ninth graders. I decided to take a closer look at those learners who had received more or fewer hours of EFL tuition than the average pupils and contrasted their test scores with the mean scores of their age group. All these cases are listed in Table 13. The number of cases covered here is very small, and thus, the aim is not to reach a conclusion about the effects of instruction hours on word knowledge, but rather to observe individual pupils and their scores with relation to the EFL tuition they had received.

TABLE 13 SIZE AND DEPTH SCORES OF PUPILS WITH DIFFERING HISTORY OF EFL TUITION

	Ninth graders		USS pupils	
	size	depth	size	depth
Kindergarten	4,900	51	3,150	46
First grade	4,550	45	4,200	48
	4,000	36	3,600	26
	1,850	15		
Second grade	4,750	58		
	4,100	40		
Fourth grade	3,400	33	3,800	33
	3,600	41	4,950	60
	4,100	39		
	3,000	46		
	3,200	35		
	4,500	55		

Six ninth graders had started learning English before the third grade. Five of these early learners had a size score much better than the average score and four scored higher in the depth test as well, when the ninth graders' mean scores were 3,613.56 and 36.02. Three early learners also belonged to the so-called top scorers, which means they scored 4,500 or more in the size test. The early learner who had already had English instruction in kindergarten scored the highest points of all the ninth graders in the size test, that is, 4,900 points. The same person also scored 51 points in

the depth test, which is a very high score but still fairly far away from the top score 64. On the other hand, there was a participant who had learnt EFL since the first grade but whose scores were substantially below the average in both tests, the participant's X-Lex score being only 1,850 and the WAT score 15.

There were also six ninth graders who had only received EFL tuition since the fourth grade. As far as the X-Lex test was concerned, two out of six late learners had a better score than the mean, and only one was a top scorer with exactly 4,500 points. Nevertheless, four persons topped the average score of the WAT. As a remark, the top scorer of the group of late learners got 55 points in the depth test, which is more than what the early learner with the highest size score had. This is just one example of how the size and depth of vocabulary knowledge do not always grow systematically at the same pace.

Among the USS pupils, there were only three persons who had received English tuition before the third grade. One of them scarcely managed to top the average size score by scoring 4,200, while the average was 4,145.77. The X-Lex scores of the other two were clearly below the mean. Additionally, two out of three got more points, i.e., 46 and 48, in the depth test compared to the mean, which was 42.75, whereas the third person scored considerably lower points, 26. Apart from the early learners, two USS pupils had started learning the language in the fourth grade. One scored 3,800 and 33 points, which were rather clearly below the mean scores of the age group. The other had a size score of 4,950 and a depth score of 60, both of which were significantly above the means and quite near the ultimate scores.

It is noteworthy that among the special cases introduced in this section, the ninth grader with the best vocabulary knowledge according to the tests had learnt English since kindergarten, while the highest points of the USS pupils were scored by a late learner. It is therefore clear that size and depth of word knowledge are not commensurate with the amount of instruction received although learners' vocabulary knowledge tend to grow the more tuition they get. Individual differences affect word knowledge remarkably, however, and learners learn differently and at a varying pace because of differing cognitive skills and spare time activities. These

factors were not subjects of this study, but asking learners about their hobbies, for instance, could provide interesting and useful information about where and how they use and encounter English in their free time.

6.5 Summary

The present study aimed at answering the following questions, *What is the size of Finnish EFL learners' receptive vocabulary and what can be said about their depth of vocabulary knowledge?*, *How do two age groups of Finnish learners differ from each other in their vocabulary size and depth?*, and *How do the two aspects of vocabulary knowledge correlate with each other?* In addition, the impact of the number EFL lessons on vocabulary size and depth was briefly pondered by paying attention to individual pupils' scores and the time they had started learning English. This section provides a summary of the results.

Based on the scores of the X-Lex test, which measures the knowledge of the 5,000 most frequent words, i.e., lemmas, the average vocabulary size of the ninth graders was 3,613.56 lemmas. The corresponding figure of the USS pupils, who were the older age group, was 4,145.77 lemmas, which means that the older pupils knew an average of around 530 lemmas more than the younger pupils did. This difference was found statistically highly significant, $p < 0.001$. There was, of course, a lot of variation in the scores within the age groups: the ninth graders' scores ranged from 1,750 to 4,900 lemmas and the USS pupils' scores varied between 2,000 and 5,000 lemmas. Overall, the ninth graders pace of learning new words had been slightly faster than that of the USS pupils, 8.7 and 7.3 words per hour of EFL tuition, respectively.

The mean X-Lex scores were compared with the CEFR scale as well as with the estimates of vocabulary size needed to understand a certain amount of certain types of texts. In addition, the ninth graders' mean vocabulary size was contrasted with results from previous studies with participants of the same age. In the grounds of their mean size score, the ninth graders were already at the B2 level of the CEFR scale, although the national core curriculum only expected them to reach the B1.1

level next. The USS pupils were expected to be at the B2.1 level soon, but their mean score proposed they were already at the CEFR level C1. In both groups, the suggested CEFR levels varied from A2 to C2 because of the variation in individual scores. It is noteworthy, however, that only 10 ninth graders scored below 2,500 and 6 USS pupils below 3,250, which were the minimums needed in both groups to reach the expected levels.

The comparison of the two age groups' vocabulary sizes and the number of words needed to comprehend the language in various situations indicated that the Finnish EFL learners' word knowledge gives them a good basis for dealing with various texts. The learners' vocabulary size should enable them to comprehend 80% of any written texts, which corresponds to the 2,000 most frequent lemmas, and 80% of any spoken texts, which is about 1,600 lemmas. It is said that 80% is enough for understanding the main points of a text. Nonetheless, on average the learners were still quite far away from the optimal threshold of 98%, which refers to a wider independence as a reader and would require a vocabulary of 12,000 lemmas or more, or some 10,000 lemmas as regards spoken texts. If all less frequent words that the learners had in their lexicon had also been counted, the vocabulary sizes had certainly been somewhat bigger than what was suggested by the X-Lex test, but I do not believe that the pupils who did not identify some of the frequent words would have recognised a much greater number of less frequent words. It is nevertheless possible that some pupils who, according to the test, had a wide vocabulary could have even reached the 98% thresholds.

Because several studies had assessed the vocabulary size of EFL learners of roughly the same age, the ninth graders' mean score was compared to the results of previous studies. The ninth graders did very well when compared with EFL learners in other countries (see Table 6). An even more interesting discovery was, however that the ninth graders' vocabulary size was considerably larger than what Takala (1984) had estimated in the 1980s when he conducted a similar study with Finnish ninth graders. This means that the average vocabulary size of Finnish ninth graders has significantly risen in the past thirty years.

The depth of vocabulary knowledge was measured with the Word Associates Test, the maximum points of which were 80. The mean score of the ninth graders was 36.02, which means they identified 45% of the right answer, and the USS pupils scored 42.75, which equals 53%. That is to say, the older pupils scored better in the depth test, and the difference was statistically very significant, $p < 0.01$. The comparison of the results to those of other studies that have utilised the WAT was rather difficult, because of the issues mentioned in Section 6.2. It was nonetheless discovered that compared to Iranian university students, for instance, the much younger Finnish EFL learners did quite well, but in comparison with the average results of native speakers and very advanced ESL learners, the pupils' scores were rather low, which is understandable.

To discover more precise differences between the age groups, as far as vocabulary depth was concerned, the depth scores were divided into synonymy and collocation scores. The ninth graders' mean synonymy and collocation scores were 16.90 out of 39 (43%) and 19.12 out of 41 (47%) respectively, whereas the corresponding scores of the USS pupils were 20.76 (53%) and 21.99 (54%). The mean synonymy scores differed very significantly from each other, $p < 0.01$, and the difference between the average collocation scores was also significant at the 0.05 level. In other words, the older pupils had more knowledge of both synonyms and collocates. Furthermore, it was noticed that the ratio of the synonymy score to the collocation score was slightly different between the groups. The ninth graders' vocabulary depth was 47% synonym knowledge and 53% collocational knowledge, while percentage values of the USS pupils were 49% and 51%, respectively. This indicates that paradigmatic associations played a bigger role in the older pupils' lexicon. The finding is in line with the view according to which paradigmatic links between words become more common when a learner advances with the language.

As it was hypothesised, the group of older EFL learners scored higher in both vocabulary tests on average and in other words, had a wider vocabulary knowledge than the younger learners. Even though there was a lot of variation in the results, which was also anticipated, the scores were consistent in a way that a positive linear

correlation between the size and depth scores could be noticed. The Pearson correlation coefficient for the two aspects of vocabulary knowledge was 0.767 and it was significant at the 0.01 level (2-tailed). The correlation between the size score and the synonymy and collocation scores were a bit lower, 0.741 and 0.721, respectively. Separate correlation coefficients were also calculated for both age groups. The ninth graders' size and depth correlated more strongly (0.803) than those of the USS pupils (0.684). The opposite was expected based on previous research findings, which have suggested that the correlation would be stronger when the learners are more advanced. Additionally, it was noticed that the younger pupils' vocabulary size correlated slightly more with their collocation score (0.770) than with their synonymy score (0.748), while the correlation of the older pupils' vocabulary size and synonymy score (0.681) was stronger than that of their vocabulary size and collocation score (0.625). Nevertheless, these minor differences between the correlation coefficients were not statistically significant.

Lastly, attention was paid to the pupils who were different in a sense that they had started learning English either earlier or later than the majority of the participants. Because there were 17 early or late learners overall, a statistical comparison was not possible. Instead, the idea was to observe individual learners to see what their vocabulary size and depth were like compared to the amount of EFL tuition they had received. This sort of examination of the test scores only strengthened the view that individual factors have a great impact on a learner's vocabulary knowledge. For instance, a ninth grader who had learnt English since the first grade had very weak scores, i.e., 1,850 in the X-Lex and 15 in the WAT, while the other two ninth graders who had learnt the language since the first grade scored considerably better. What is more, the person with the best vocabulary knowledge of the early or late learners in the ninth grade was the one who had learnt English since kindergarten, whereas the USS pupil who had started learning the language in the fourth grade was the best scorer among the older group of early and late learners.

7 DISCUSSION

The object of the present study was to assess Finnish EFL learners' receptive vocabulary size and depth. This was done by conducting two vocabulary tests, the X-Lex, which measures size, and the WAT, which assesses depth. The two tests were chosen because they have a simple format and they consist of decontextualized items, which makes it possible to cover a larger number of items in a relatively short time. The tests suited well the purposes and could be conducted within the limits of the study. The X-Lex and WAT have often been used in vocabulary studies to measure size and depth and are generally perceived as suitable measures for these two aspects of vocabulary knowledge. In other words, the validity of the tests can be considered good.

All participants were given the same instructions and the same time to complete the exact same tests. Moreover, the tests of all participants were treated equally and objectively. These add to the reliability of the study. (Martin 2011: 162.) However, because the tests had to be conducted within a limited time, that is, within a lesson, only one X-Lex test was used instead of three, which is the recommendation. For the same reason, the WAT was reduced to 20 items. These changes might have had an effect on the reliability of the results in terms of accuracy. If more words were tested in the size test, the role of chance would have been smaller and the results would have represented the participants' vocabulary sizes even more accurately. The same goes for the depth test. In addition, if the entire 40-item WAT were completed, it would have enabled a better comparison with the WAT results gained from other studies.

There were a total of 130 participants who represented two different age groups, ninth graders (N = 59) and second year upper secondary school pupils (N = 71). The groups were large enough for a statistical analysis and the test results were analysed using the SPSS software. However, the samples were quite small compared to the entire population, which in this case refers to all ninth graders and second year upper secondary school pupils in Finland. What is more, all participants were pupils

in the same comprehensive and upper secondary schools. For these reasons, generalising the results across the population is not possible, although the results can be considered suggestive of the vocabulary knowledge of Finnish ninth graders and USS pupils.

The results of the study confirmed what could be expected: the older pupils, who had received more EFL tuition, had a larger receptive vocabulary and possessed more knowledge of words than the two years younger pupils. However, the ninth graders had learnt more words in relation to the amount of EFL instruction they had received, which could indicate that the pace of acquiring new words might slow down the more advanced the learners are. This supports the idea that more advanced learners might develop their vocabulary depth at the expense of size. This is only speculative, however, since such developmental aspects of vocabulary learning can only be assessed with longitudinal studies. Additionally, a strong positive correlation was found between size and depth of vocabulary knowledge, as it was hypothesised.

What was surprising, however, were the high scores the participants gained in the X-Lex test in comparison with the CEFR levels they were expected to reach next. The goal for ninth graders set in the national core curriculum is to reach the B1.1 level by the end of grade 9 as regards three different skill categories: “the ability to interact with other people”, “the ability to read or interpret texts” and “the ability to write or produce texts” (National Board of Education 2015a: 351-352). As for the USS pupils, they should be near the B2.1 level in the same categories after completing all compulsory EFL courses. If Milton and Meara’s (2003: cited in Milton 2009: 186) comparison of the X-Lex scores with the CEFR levels can be trusted, the ninth graders might actually already be at the B2 level and the USS pupils at the C1 level. Only 10 out of 59 ninth graders and 6 out of 71 USS pupils had not achieved their expected level yet. It is therefore clear that, on average, the learners’ vocabulary size did not match the language level they were expected reach soon. Thus, the question is, was the learners’ vocabulary simply larger than what their expected skill levels suggest, or, had the learners, indeed, reached the higher CEFR levels proposed by the size scores. If the latter were true, it might be time to redefine the goals of EFL

instruction set in the curricula, since the current goals are too low compared to the learners' skills and potential.

Nevertheless, regardless of the right answer to the question raised above, the fact is that the Finnish EFL learners studied in the present study had a large vocabulary size. The ninth graders were contrasted with EFL learners of the same age from other countries and they outdid their foreign peers. The substantial differences between nationalities were somewhat unexpected. For instance, the German EFL learners studied by Milton and Meara (1998, see Table 6) had a vocabulary of 1,200 words after they had received 400 hours of tuition, which is only a third of the Finnish ninth graders' vocabulary size. German EFL learners could be expected to score higher compared to Finnish learners because German and English are cognate languages unlike Finnish and English. Based on the results, however, the reality is very different. What is more, Canga Alonso's (2013, see Table 6) Spanish EFL learners had had over 1,000 hours of instruction but their average vocabulary was only some 900 words. The size is remarkably small when it is set against the amount of EFL instruction. Thus, studies have revealed that there can be considerable variation in the pace of vocabulary learning between nationalities.

Although the Finnish school system is often praised, I believe that good EFL tuition is not the only reason for why the learners in Finland have such a wide vocabulary compared to foreign EFL learners. Even if English does not have an official status in Finland, it can be heard and seen everywhere in all media, and it figures more prominently in Finland than in many other countries, including Germany and Spain. The Finnish youth, in particular, are very much used to using English in social media and when playing computer or online games, for instance. That is to say, the youth learn a lot of English in their spare time as well and, I would say, this applies especially to new vocabulary. The chances to use English have clearly risen since the 1980s when Takala (1984) conducted his study with the ninth graders. It is therefore also no wonder that the ninth graders of today have a wider vocabulary than the ninth graders thirty years ago.

The X-Lex results proved that the Finnish EFL learners master the most frequent words fairly well, but there is of course always room for improvement. The results of the WAT, which also consists of words that frequently appear in the language, were surprising to some extent. Since there was no baseline to which the WAT results could have been compared, such as the CEFR scale in relation to the X-Lex scores, and previous studies had concentrated on older language learners, the evaluation of the results was somewhat difficult. In other words, stating whether the pupils scored well or not was not that straightforward. Finding synonyms and collocates for adjectives seemed to be challenging, however, and I was maybe anticipating higher scores based on the size scores. For instance, in comparison to the high proficient group of Iranian students, whose vocabulary size was less than 3,000 lemmas and who scored 53% in WAT, the Finnish pupils in both age groups could have been assumed to score better. Maybe the Finnish pupils just did not have as versatile knowledge of words as their vocabulary size suggested. One possible explanation for this might be the way words are typically presented in Finnish EFL books, that is, as lists of discrete decontextualized words with one or two meanings in Finnish. In other words, pupils may learn to recognise a lot of words by reading the word lists, but since the lists only provide one or two meanings and even fewer examples of how to use the words and how they collocate with other words, it is no wonder that the depth of their vocabulary knowledge does not correspond to their vocabulary size. I would therefore suggest that vocabulary learning and teaching in Finnish schools should not be so bound by the vocabulary lists but focus even more on texts and contextualised words.

Despite their knowledge of words not being as versatile as the assumption was, the learners' vocabulary depth followed the pattern which was expected based on previous research: both age groups had relatively more knowledge of collocates than synonyms, but the difference between collocational and synonym knowledge was narrower among the older pupils. In a sense, the bigger proportion of collocational knowledge also contradicts the typical vocabulary lists, which do not usually include any collocations. It could have been anticipated that the learners would have had better synonymy scores because the lists tend to provide one or two Finnish

equivalents for the words, but that was not the case. Evidently, providing learners with such word lists does not mean they would acquire the words and the given meanings. Instead, the higher collocation scores could be considered to embrace the idea that words are learnt and stored in the memory as chunks. It would therefore be profitable to offer learners chunks and sequences of words because that could make learning easier. When a specific word appears in several word chunks that a learner is familiar with, it helps him/her to differentiate and identify the various meanings of that word and thus, the knowledge of synonyms improves as well. I am therefore of the opinion that more attention should be paid to how words are presented in Finnish EFL books and in the classroom so that the presentation would better correspond with the way language learners tend to acquire and memorise word meanings. In relation to this, the usefulness of the most frequent words should be underlined even more and special attention should be drawn to the most frequent words and to the various word chunks they tend to appear in.

Vocabulary size and depth have not been very common subjects of study in Finland, while other areas of language learning and teaching have attracted more attention. There is evidently a void in this field of research, which the present study alone cannot fill, and thus, more research is needed. To get more reliable information about Finnish EFL learners' word knowledge, sample sizes need to be bigger and learners from across the country should be represented in the samples. Additionally, the study of vocabulary size and depth should be extended to cover learners of all ages to get a better impression of EFL learners' word knowledge overall. It would also be very important to conduct longitudinal studies which would provide information on the development of word knowledge in the course of time. Of course, attention should be drawn to both receptive and productive vocabulary.

Research on vocabulary size and depth could provide a lot of useful information about learners' language skills and give researchers, teachers and other people involved new insights into learning and teaching word knowledge. It would be very important for teachers to have more precise knowledge about what the EFL learners in Finland know about words and what kind of knowledge they lack. Learners

themselves might also be interested in finding out how wide a vocabulary they have and how much they know about words. Thus, by conducting vocabulary tests at regular intervals, learners would see concretely how their word knowledge develops, while teachers would get valuable data that could help them plan their instruction. Furthermore, EFL books and other teaching materials could be improved so that they would better match learners' needs. For instance, as discussed above, test results might give a reason to rethink the presentation of words in EFL books and in the classroom so that it would better conform to the way learners acquire word knowledge.

It is actually rather strange that so little quantitative research has been done on word knowledge in Finland, even though vocabulary is such an important part of language learning. The lack of quantitative information might be due to the fact that testing does not quite suit the Finnish school culture. Testing is often seen as something negative and tests can cause competition between pupils and schools, which is by no means the aim of the school system. However, the more knowledge of the vocabulary learning process and the development of different aspects of vocabulary knowledge there is, the better. That is to say, Finnish EFL learners' vocabulary knowledge needs to be tested and researched despite the fear that testing might entail competition. I think that the significance and usefulness of word tests should be underlined already in teacher training. If it were emphasised that tests can be used for collecting data about learners' knowledge and not only for ranking pupils' performances, testing would become a more natural part of language learning and teaching, and it would not cause so much anxiety among teachers and pupils.

8 BIBLIOGRAPHY

- Adolphs, S. and Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics* 24 (4), 425-438. doi: 10.1093/applin/24.4.425.
- Akbarian, I. (2010). The relationship between vocabulary size and depth for ESP/EAP learners. *System* 38 (3), 391-401. doi: 10.1016/j.system.2010.06.013.
- Anderson, R. C. and Freebody, P. (1981). In J. T. Guthrie (ed.), *Comprehension and teaching: Research reviews*. Newark, DE: International Reading Association, 77-117. <http://files.eric.ed.gov/fulltext/ED203299.pdf>.
- Benson, M., Ilson, R. F. and Benson, E. (2010). *The BBI combinatory dictionary of English: Your guide to collocations and grammar*. 3rd ed. Amsterdam: John Benjamins Publishing Company. <http://lexikos.journals.ac.za/pub/article/view/962/479>.
- Bogaards, P. (2000). Testing L2 vocabulary knowledge at the high level: The case of the Euralex French tests. *Applied Linguistics* 21 (4), 490. doi: 10.1093/applin/21.4.490.
- Bogaards, P. (2001). Lexical units and the learning of foreign language vocabulary. *Studies in Second Language Acquisition* 23 (3), 321-343. <http://search.proquest.com/docview/224054032?accountid=11774>.
- Cambridge Dictionaries Online. <http://dictionary.cambridge.org/dictionary/english/collocation>. (27 January, 2016).
- Canga Alonso, A. (2013). Receptive vocabulary size of secondary Spanish EFL learners. *Revista de Lingüística y Lenguas Aplicadas* 8 (1), 66-75. doi: 10.4995/rlyla.2013.1180.
- Carter, R. (2012). *Vocabulary: Applied linguistics perspective*. 2nd ed. London: Routledge.
- Cobb, T. (2015). Word associates test. <http://www.lex tutor.ca/tests/associates/>. (16 November, 2015).
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. www.coe.int/t/dg4/education/elp/elp-reg/Source/Key_reference/CEFR_EN.pdf. (1 December, 2015).
- Daller, H., Milton, J. and Treffers-Daller, J. (Eds.) (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Eyckmans, J., Van de Velde, H., van Hout, R. and Boers, F. (2007). In H. Daller, J. Milton and J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press, 59-76.
- Farvardin, M. T. and Koosha, M. (2011). The role of vocabulary knowledge in Iranian EFL students' reading comprehension performance: breadth or depth? *Theory and Practice in Language Studies* 1 (11), 1575-1580. doi: 10.4304/tpls.1.11.1575-1580.
- Finnish National Board of Education. (2015a). *Perusopetuksen opetussuunnitelman perusteet 2014*. http://www.oph.fi/download/163777_perusopetuksen_opetussuunnite

- lman_perusteet_2014.pdf. (2 February, 2016).
- Finnish National Board of Education. (2015b). *Lukion opetussuunnitelman perusteet 2015*.
http://www.oph.fi/download/172124_lukion_opetussuunnitelman_perusteet_2015.pdf. (2 February, 2016).
- Goulden, R., Nation, P. and Read, J. (1990). How large can a receptive vocabulary be? *Applied linguistics* 11 (4), 341-363.
<http://applij.oxfordjournals.org/content/11/4/341.abstract>.
- Greidanus, T. and Nienhuis, L. (2001). Testing the quality of word knowledge in a second language by means of word associations: Types of distractors and types of associations. *The Modern Language Journal* 85 (4), 567-577.
<http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=0f087dae-90ed-4667-b1bf-76983fa286b0%40sessionmgr120&vid=1&hid=105>.
- Hellman, A. (2011). Vocabulary size and depth of word knowledge in adult-onset second language acquisition. *International Journal of Applied Linguistics* 21 (2), 162-182. doi: 10.1111/j.1473-4192.2010.00265.x.
- Hu, M. and Nation, P. (2000). *Unknown vocabulary density and reading comprehension*.
nflrc.hawaii.edu/rfl/PastIssues/rfl131hsuehchao.pdf.
 (26 November, 2015).
- Huhta, A. and Tarnanen, M. (2011). Kielitaidon arviointi tutkimusvälineenä ja tutkimuksen kohteena. In P. Kalaja, R. Alanen and H. Dufva (Eds.), *Kieltä tutkimassa: tutkielman laatijan opas*. Tampere: Finn Lectura, 201-220.
- Johnson, K. (2008). *Quantitative methods in linguistics*. Malden, MA: Blackwell.
- Kallela, O. (2014). *The productive knowledge of English vocabulary of Finnish upper secondary school students: a corpus-based study*. Unpublished Pro Gradu thesis. Helsinki: Helsingin yliopisto.
- Laufer, B. and Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language* 22(1), 15-30.
<http://search.proquest.com/docview/1705667242/fulltextPDF/135B80AA4C25452EPQ/1?accountid=11774>.
- Li, M. and Kirby, J. R. (2012). Breadth and depth of vocabulary knowledge in second language reading. *Journal of Communications Research* 4 (4), 335-359.
<http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=af08369c-af8b-411b-93ac-dad7767e3136%40sessionmgr4002&vid=1&hid=4109>.
- Makkonen, H-L. (2008). *Measuring the vocabulary knowledge of advanced and intermediate English students at university level*. Unpublished Pro Gradu thesis. Turku: Turun yliopisto.
- Martin, M. (2011). Kokeilua kielellä: kokeelliset menetelmät kielenoppimisen ja -opetuksen tutkimuksessa. In P. Kalaja, R. Alanen and H. Dufva (Eds.), *Kieltä tutkimassa: tutkijan laatijan opas*. Tampere: Finn Lectura, 162-179.
- Meara, P. (1983). Word associations in a foreign language.
www.lognostics.co.uk/vlibrary/meara1983.pdf. (26 November, 2015).
- Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam: John Benjamins Publishing Company.
- Milton, J. (2007). Lexical profiles, learning styles and the construct validity of lexical

- size tests. In H. Daller, J. Milton and J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press, 47-58.
- Milton, J. (2009). *Second language acquisition: Measuring second language vocabulary acquisition*. Clevedon, GBR: Multilingual Matters.
- Milton, J. and Meara, P. (1998). Are the British really bad at learning foreign languages? *The Language Learning Journal* 18 (1), 68-76.
doi: 10.1080/09571739885200291.
- Ministry of Education and Culture. (2012). *Perusopetuksen tuntijako*.
http://www.minedu.fi/export/sites/default/OPM/Koulutus/koulutuspolitiikka/vireilla_koulutus/perusopetus/liitteet/asetusehdotus_1_2.pdf. (2 February, 2016).
- Ministry of Education and Culture. (2014). *Lukion tuntijako*.
http://minedu.fi/export/sites/default/OPM/Koulutus/koulutuspolitiikka/vireilla_koulutus/lukio/1311_fi_Liite__Lukion_tuntijako.pdf. (2 February, 2016).
- Nassaji, H. (2004). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Canadian Modern Language Review* 61 (1), 107-134.
<http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=94d6966d-dc51-455e-9824-db71d5d021e4%40sessionmgr4001&vid=1&hid=4109>.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle, Cengage Learning.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review* 63 (1), 59-81.
<http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=248bed1a-9d6b-4647-83f8-cbf30c93fd60%40sessionmgr4004&vid=1&hid=4109>.
- Nation, I. S. P. (2007). Fundamental issues in modelling and assessing vocabulary knowledge. In H. Daller, J. Milton and J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press, 35-43.
- Nation, I. S. P. (2009). *Learning vocabulary in another language*. 11th ed. Cambridge: Cambridge University Press.
- Nation, I. S. P. and Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle, Cengage Learning.
- Nurweni, A. and Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes* 18 (2), 161-175.
doi: 10.1016/S0889-4906(98)00005-2.
- Paribakht, T. M. and Wesche, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal* 11 (1), 9-29.
<http://www.teslcanadajournal.ca/index.php/tesl/article/view/623/454>.
- Pietilä, P. and Lintunen P. (Eds.) (2014). *Kuinka kieltä opitaan: opas vieraan kielen opettajalle ja opiskelijalle*. Helsinki: Gaudeamus.
- Pirilä, A. (2012). *Knowledge of English vocabulary at advanced levels: vocabulary size and*

- academic vocabulary knowledge of Finnish 1st university students and MA students of English*. Unpublished Pro Gradu thesis. Turku: Turun yliopisto.
- Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review* 56 (2), 282-308.
<http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=d09232a9-8939-444a-b4e7-5539bad977b3%40sessionmgr198&vid=1&hid=105>.
- Rasinger, S. M. (2008). *Quantitative research in linguistics: An introduction*. London: Continuum International Publishing Group.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing* 10 (3), 355- 371. doi: 10.1177/026553229301000308.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards and B. Laufer (Eds.), *Vocabulary in a second language. Selection, acquisition, and testing*. Philadelphia, PA: John Benjamins Publishing Company, 209-227.
<http://site.ebrary.com/lib/jyvaskyla/detail.action?docID=10064637>.
- Ringbom, H. (1987). *The role of the first language in foreign language learning*. Clevedon, Avon, England: Multilingual Matters.
<http://web.ebscohost.com/ehost/ebookviewer/ebook/bmxlYmtfXzIzNDIwX19BTg2?sid=a80342ad-ee70-4b08-a20c-cb04916c067e@sessionmgr198&vid=0&format=EB&rid=1>.
- Schmitt, N., Schmitt, D., and Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing* 18 (1), 55-88.
<http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=d683391a-ed62-4d7c-90dc-41470ba1fca7%40sessionmgr111&vid=1&hid=105>.
- Schmitt, N., Ng, J. W. C. and Garras, J. (2011). The word associates format: Validation evidence. *Language Testing* 28 (1), 105-126.
 doi: 10.1177/0265532210373605.
- Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.
- Singleton, D. (2000). *Language and the lexicon: An introduction*. London: Arnold.
- Takala, S. (1984). *Evaluation of students' knowledge of English vocabulary in the Finnish comprehensive school*. Jyväskylä: Kasvatustieteiden tutkimuslaitos.
- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics* 22 (1), 217-234.
<http://search.proquest.com/docview/200938686?accountid=11774>.
- Vilka, H. (2007). *Tutki ja mittaa: määrällisen tutkimuksen perusteet*. Helsinki: Tammi.

9 APPENDIX: The vocabulary tests

Sanastotutkimus

tammikuu 2016 luokka-aste: _____

Täytä esitiedot!

sukupuoli: mies nainen

äidinkieli: _____

Millä luokalla olet aloittanut englannin opiskelun? _____

Oletko asunut/opiskellut/oleskellut pidemmän aikaa englanninkielisessä maassa? Jos olet, kuinka kauan ja missä.

TESTI 1

Alla olevassa taulukossa on oikeita englanninkielisiä sanoja sekä keksittyjä sanoja, jotka on tehty muistuttamaan oikeita sanoja. Käy kaikki taulukon sanat yksitellen läpi ja rastita ne sanat, joita osaat käyttää englanniksi lauseessa tai joille tiedät jonkin suomenkielisen merkityksen. Esimerkiksi:

dog	X	vengeance	happy	X
-----	----------	-----------	-------	----------

system	interval	had	warboy	govern	mosquito
proceed	which	industry	overcome	position	cordons
skemp	rot	frequent	border	little	knowledge
relation	just	trudgeon	manly	dozen	grass
turn	perform	rabbit	pat	opponent	stillhard
sneeze	steady	word	plenty	astell	style
wire	inertible	drag	reference	overlook	open
steam	gallimore	enclose	previous	wife	worry
miserable	take	climb	screen	surman	manager
chicorate	combine	squeeze	main	enigmatic	serve
wedge	upset	vain	bring	eckett	thick
wet	educate	varney	network	odd	meet

person	collar	leadership	widgery	prepare	simplicity
you	dial	callisthemia	liner	cap	castle
postherent	sleeve	low	display	dip	excite
instant	cord	moffant	faint	recommend	wrong
goat	qualify	native	troake	prepare	artificial
audience	troublesome	boy	sum	flag	waggett
interest	nurse	property	frank	gorman	forwards
girl	murrow	sake	envelope	pan	fog

TESTI 2

20 **tummennetun** adjektiivin (= *millainen jokin on?*) alla on rivi sanoja. Rastita rivin vasemmalta puolelta kaikki ne sanat, joilla on sama tai ainakin osittain sama merkitys kuin tummennetulla adjektiivilla. Lisäksi, rastita rivin oikealta puolelta ne sanat, joita käytetään hyvin usein yhdessä tummennetun adjektiivin kanssa. Oikeita vastauksia on korkeintaan 4 per rivi ja ne voivat jakaantua kahden puolen välillä miten tahansa. Etsi oikeat vastaukset eli rastita korkeintaan 4 sanaa joka kohdassa. Esimerkiksi:

little samaa tai osittain samaa tarkoittavat | usein yhdessä käytettävät

small X	strange	young X	short X	wall	language	boy X	weather
---------	---------	---------	---------	------	----------	-------	---------

1. beautiful

enjoyable	expensive	free	loud	education	face	music	weather
-----------	-----------	------	------	-----------	------	-------	---------

2. bright

clever	famous	happy	shining	colour	hand	poem	taste
--------	--------	-------	---------	--------	------	------	-------

3. calm

open	quiet	smooth	tired	cloth	day	light	person
------	-------	--------	-------	-------	-----	-------	--------

4. natural

expected	helpful	real	short	foods	neighbours	parents	songs
----------	---------	------	-------	-------	------------	---------	-------

5. fresh

another	cool	easy	raw	cotton	heat	language	water
---------	------	------	-----	--------	------	----------	-------

6. general

closed	different	usual	whole	country	idea	reader	street
--------	-----------	-------	-------	---------	------	--------	--------

7. bare

empty	heavy	uncovered	useful	cupboard	feet	school	tool
-------	-------	-----------	--------	----------	------	--------	------

8. acute

hidden	often	rich	sharp	angle	hearing	illness	stones
--------	-------	------	-------	-------	---------	---------	--------

9. common

complete	light	ordinary	shared	boundary	circle	name	party
----------	-------	----------	--------	----------	--------	------	-------

10. complex

angry	difficult	necessary	sudden	argument	passengers	patterns	problem
-------	-----------	-----------	--------	----------	------------	----------	---------

11. broad

full	moving	quiet	wide	night	river	shoulders	smile
------	--------	-------	------	-------	-------	-----------	-------

12. conscious

awake	healthy	knowing	laughing	face	decision	effort	student
-------	---------	---------	----------	------	----------	--------	---------

13. convenient

easy	fresh	near	suitable	experience	sound	time	vegetable
------	-------	------	----------	------------	-------	------	-----------

14. dense

crowded	hot	noisy	thick	forest	handle	smoke	weather
---------	-----	-------	-------	--------	--------	-------	---------

15. curious

helpful	interested	missing	strange	accident	child	computer	steel
---------	------------	---------	---------	----------	-------	----------	-------

16. distinct

clear	famous	separate	true	advantage	meanings	news	parents
-------	--------	----------	------	-----------	----------	------	---------

17. dull

cloudy	loud	nice	secret	colour	knife	place	rock
--------	------	------	--------	--------	-------	-------	------

18. direct

honest	main	straight	wide	fence	flight	heat	river
--------	------	----------	------	-------	--------	------	-------

19. favourable

helpful	legal	possible	positive	habit	response	teacher	weather
---------	-------	----------	----------	-------	----------	---------	---------

20. secure

confident	enjoyable	fixed	safe	game	job	meal	visitor
-----------	-----------	-------	------	------	-----	------	---------