

**Assessing the Role of Vocabulary Breadth and Depth
in Reading Comprehension: A Quantitative Study of
Finnish EFL Learners**

Noora Harkio

MA Thesis

University of Turku

School of Languages and Translation Studies

Department of English

April 2016

Turun yliopiston laatu järjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

UNIVERSITY OF TURKU

School of Languages and Translation Studies / Faculty of Humanities

HARKIO, NOORA: Assessing the Role of Vocabulary Breadth and Depth in
Reading Comprehension: A Quantitative Study of Finnish
EFL Learners

MA Thesis, 81 p., 6 app.

Department of English

April 2016

In this MA thesis, Finnish learners of English were studied in order to examine the relationship between second language vocabulary size, vocabulary depth, and reading comprehension. In addition, given the well-established connection between vocabulary size and reading comprehension, the second aim of the study was to see whether assessing vocabulary depth could add another dimension in predicting and explaining reading comprehension proficiency.

Two groups were studied: the first group consisted of 39 Finnish upper secondary school students (the TOKA group) whereas the second group consisted of 19 university students of English at the University of Turku (the YLI group). Thus, comparisons were made between the results of a less advanced and a very advanced group of English learners, which was the third aim of the study. The participants in both groups filled in a background information form and took three tests: a multiple-choice reading comprehension test, a multiple-choice vocabulary size test, and a test designed to elicit information on learners' depth of vocabulary knowledge of certain English words. The data were analysed using statistical methods.

The results of the study show that the scores on the three tests were positively correlated in both study groups as well as in the two groups together. However, the correlations were higher in the TOKA group and in the two groups in total than in the YLI group. When examining the variance in reading comprehension test scores explained by vocabulary size and vocabulary depth, the figures of explained variance were again higher in the TOKA group and in the two groups in total than in the YLI group. When it comes to the results of the YLI group, vocabulary depth did not indeed seem to add any explained variance into the explanation of reading comprehension test scores.

Based on the results of the study, it seems that vocabulary size and depth have a less significant role in the reading comprehension skills of more advanced learners of English. When looking at the less advanced TOKA group, on the other hand, vocabulary size and depth seem to be clear indicators of reading proficiency. In addition, the test results of the YLI group were clearly more uniform than those of the TOKA group. The variance in the test results of the TOKA group was large.

Key words: English as a Foreign Language, Reading Comprehension, Vocabulary Depth, Vocabulary Learning, Vocabulary Size

Table of Contents

1 Introduction	1
2 Vocabulary as a multi-dimensional construct	4
2.1 Definition of <i>a word</i>	4
2.2 Definition of <i>vocabulary</i>	5
2.3 Different perspectives on defining vocabulary knowledge	7
3 Role of vocabulary knowledge in L2 learning in general	12
4 Reading as one of four skills in L2	15
4.1 What is involved in reading.....	15
4.2 Vocabulary knowledge affects reading comprehension.....	17
4.3 Previous studies	21
5 Methodology	25
5.1 Assessing L2 skills	25
5.1.1 Assessing L2 vocabulary size	25
5.1.1.1 Vocabulary Levels Test.....	27
5.1.1.2 Eurocentres Vocabulary Size Test.....	28
5.1.1.3 Measure used in this study: Vocabulary Size Test.....	29
5.1.2 Assessing L2 vocabulary depth	32
5.1.2.1 Word Associates Format	33
5.1.2.2 Measure used in this study: Vocabulary Knowledge Scale ...	34
5.1.3 Assessing reading comprehension	40
5.2 Participants.....	43
5.3 Test administration.....	46
5.4 Statistical methods	47
6 Results.....	49
6.1 Test scores	49
6.1.1 Vocabulary Size Test.....	50

6.1.2 Vocabulary Knowledge Scale	51
6.1.3 Reading comprehension test	52
6.2 Examination of the relationships between the test results	52
6.3 Examination of the impact of vocabulary size and depth in reading comprehension test results.....	54
7 Discussion.....	58
7.1 Discussing the test results.....	58
7.2 Relationships between vocabulary breadth, depth and reading comprehension: comparison with earlier studies.....	61
7.3 Role of vocabulary depth in reading comprehension: comparison with earlier studies	66
7.4 Comparison of the results found in the TOKA and YLI groups	68
8 Conclusion	71
References.....	75
Appendices	

List of Tables

Table 1. Aspects of vocabulary knowledge (Nation 2001: 27)	9
Table 2. VKS elicitation scale self-report categories (Wesche & Paribakht 1996: 30)	36
Table 3. VKS scoring categories: Meaning of scores. Adapted from Wesche and Paribakht (1996: 30), Paribakht and Wesche (1997: 180) and Joe (1998: 363)	36
Table 4. Background information about the informants in the TOKA group	43
Table 5. Background information about the informants in the YLI group	45
Table 6. Scores on the Vocabulary Size Test	50
Table 7. Scores on the Vocabulary Knowledge Scale	51
Table 8. Scores on the reading comprehension test	52
Table 9. Spearman's rho correlations in the TOKA group	53
Table 10. Spearman's rho correlations in the YLI group	54
Table 11. Spearman's rho correlations in both groups in total	54
Table 12. Multiple regression results of the TOKA group with vocabulary size and the scores on the VKS as independent variables	56
Table 13. Multiple regression results of the YLI group with vocabulary size and the scores on the VKS as independent variables	56
Table 14. Multiple regression results of both groups together with vocabulary size and the scores on the VKS as independent variables	57

List of Abbreviations (in alphabetical order)

BNC	British National Corpus
COCA	Corpus of Contemporary American English
EFL	English as a Foreign Language
EURO	Eurocentres Vocabulary Test
L1	First Language
L2	Second Language
SPSS	Predictive analysis software: <i>IBM SPSS Statistics 23.0.0.2</i>
TOEFL	Test of English as a Foreign Language
TOKA	Informant(s) who are on their fifth upper secondary school English course
VKS	Vocabulary Knowledge Scale
VLT	Vocabulary Levels Test
VST	Vocabulary Size Test
WAF	Word Associates Format
YLI	Informant(s) who are English majors at the University of Turku

1 Introduction

Today, second language (L2) vocabulary knowledge is considered to be in the centre of L2 learning. It is indeed very difficult to imagine a successful L2 learner without a large enough vocabulary. Vocabulary plays a crucial role in all four skills of L2: writing, reading, speaking and listening. Thus, it is essential that vocabulary is taught in efficient ways and that attention is paid to learning new vocabulary. This progress has to be monitored as well. The present study will look at Finnish learners of English, their vocabulary knowledge and their reading comprehension skills. More specifically, both vocabulary size and vocabulary depth will be examined as features of vocabulary knowledge.

Before turning to the main concepts and aims of the present study, I will briefly clarify a few central terms that will be used throughout the thesis. *First language* (L1) refers to the language that a person learns during the first years of his or her life and that is used by the child's parents, siblings and caretakers (Ortega 2009: 5). In this thesis, a very broad sense of the term *second language* (L2) is applied; L2 refers to any language that is learnt after L1 has been acquired (*ibid.*). Despite the fact that some researchers insist on making a clear distinction between L2 and a foreign language, the difference is not decisive in this thesis. Hence, the common term *English as a foreign language* (EFL) is used when discussing learners of English, for example.

In the present study, both size of vocabulary and depth of vocabulary will be involved. In addition to these two, another important concept is *reading comprehension*. The aim of this study is to find out how the three variables, namely L2 vocabulary size, vocabulary depth and reading comprehension, relate to each other. Moreover, another goal of this study is to see whether L2 vocabulary depth as a variable can offer yet another factor, in addition to that provided by L2 vocabulary size, in predicting learners' performance in L2 reading comprehension. In earlier studies, it has been shown that large vocabularies lead learners to be more successful in reading comprehension.

My focus, however, is on finding out what kind of information another variable of vocabulary knowledge, L2 vocabulary depth, can offer in both explaining and predicting learners' reading comprehension skills. To the best of my knowledge, so far no studies involving all these three variables have been conducted in Finland. L2 vocabulary size has been in the centre of many studies, both in Finland and abroad; however, L2 vocabulary depth has not been considered in many studies at all. Based on the few previous studies conducted, I assume that vocabulary size and depth correlate strongly both with each other and with reading comprehension test scores, but that vocabulary depth can nevertheless serve as an independent factor in predicting and explaining performance in reading tests.

In this study, two groups of learners will be involved. Earlier studies have mainly concentrated on university level EFL learners, but in this study, two groups that are at different proficiency levels will be studied. The first group consists of Finnish upper secondary school students, whereas the second group consists of students of English at the University of Turku. This way, the present study will also look at what kind of differences there are between these two groups. I assume that the more advanced group of university students of English will perform better in all language tests than the group that consists of upper secondary school students. Moreover, my hypothesis is that both size and depth of vocabulary knowledge are more strongly related to the performance in the reading comprehension test in the upper secondary school group, whereas the results of the university group are less related to vocabulary knowledge and could possibly be explained by high English proficiency overall, including other aspects of language proficiency and the use of reading strategies, for example.

In the following chapter, I will discuss concepts that are fundamental in order to understand the starting point for this study. Both *a word* and *vocabulary knowledge* will be defined, and various perspectives on defining the construct of vocabulary knowledge will be presented. In the third chapter, I will examine the role of vocabulary knowledge in L2 learning in general. The focus of Chapter 4 will be on reading. In the subchapters, I will look into what is involved

in reading, what is the significance of vocabulary knowledge in reading, and, lastly, some previous studies will be presented. I will move on to present the methodology of the present study in Chapter 5. In this section, I will discuss different vocabulary test formats as well as introduce the tests used in this study. In addition, the groups of informants, the process of test administration and the statistical methods used will be described. The results will be presented in Chapter 6, and these findings will be further discussed in Chapter 7. Chapter 8 will conclude the study: the results will be reviewed, the limitations of the study acknowledged, and suggestions for further research made.

2 Vocabulary as a multi-dimensional construct

In this chapter, I will firstly introduce the concept of a *word*: what this concept refers to and what is meant by a *word* in this study. Next, the concept of a *word* will be further developed into the concept of *vocabulary*. I will discuss, for example, what vocabulary knowledge exactly is and what it consists of. Moreover, some frameworks related to vocabulary as a construct will be presented in the final subchapter of this chapter.

2.1 Definition of a *word*

In order to discuss vocabulary knowledge, it is first important to think about what a *word* actually means. In this subchapter, I will introduce some basic definitions related to the concept of a *word* and, lastly, come to a conclusion as to what is meant by a *word* in the present study.

When discussing words and how to count them, for example, some most commonly used terms include tokens, types and lemmas. When counting *tokens*, each and every word is counted as a token of its own (e.g. Nation 2001: 7, Schmitt 2010: 188). Thus, the sentence *It may be challenging to change it* contains seven words, or seven tokens. The same sentence may as well be analysed in terms of *types*: if the same word occurs again, it is not counted (*ibid.*). This way, the sentence above would consist of six words, or six types. Another way of analysing words in a given text is to look at *lemmas*. A lemma is a broader term than a type or a token; both the headword and some of its reduced and inflected forms are included in a lemma (e.g. Nation 2001: 7, Schmitt 2010: 189). For example, a single lemma would include the headword *arrive* and some of its inflected forms, such as *arrives* or *arrived*. Similarly, the headword *toy* and its plural form *toys* would constitute a single lemma. Some problems that relate to the use of lemmas concern irregular forms as well as closely related words that belong to different word classes (e.g. Nation 2001: 7–8). It is unclear whether forms such as *go* – *went* – *gone* or *woman* – *women* constitute a single lemma or several lemmas. Usually,

words belonging to one lemma are of the same part of speech. However, it is questionable whether noun and verb uses of for example *talk* or *walk* constitute a single lemma or not.

The term *word family* was introduced in order to overcome some of the problems encountered with lemmas. Word family is a somewhat broader term than the lemma as it includes the headword along with “its inflected forms and its closely related derived forms” (Nation 2001: 8). When it comes to the inflected forms, more affixes, such as *un-* and *-ness*, are involved than in the definition of a lemma (*ibid.*). As with lemmas, though, it is sometimes problematic to know what is included in a word family and what is not. Learners’ knowledge of word families might in fact be more like a continuum as their knowledge of affixes, for example, develops as they learn more about a given language (*ibid.*). The concept of a word family seems indeed psychologically real and the use of word families is supported by existing and growing evidence that strongly suggests that the mind processes these kinds of forms together (e.g. Nagy *et al.* 1989; Bertram, Baayen & Schreuder 2000; Bertram, Laine & Virkkala 2000). The use of word families has gained much interest, and it has become a term commonly applied in vocabulary-related studies, for example. Consequently, in the present study, unless otherwise mentioned, a word refers to a word family.

2.2 Definition of *vocabulary*

Shortly, *vocabulary* refers to the component of language that involves words and their meanings (Saville-Troike 2006: 191). Nowadays, it is commonly accepted that vocabulary is not a single dimension but rather a multi-dimensional construct. This is why it is common to make distinctions between different dimensions and features of vocabulary. Also, there are many complementary frameworks that try to explain what kind of a construct vocabulary is and what it means to know a word.

A distinction that is commonly made is the one between productive and receptive vocabulary knowledge. By productive language skills in general,

linguists refer to writing and speaking: activities that require communicating meanings to others (e.g. Saville-Troike 2006: 137, 193). Receptive language skills, on the other hand, refer to reading and listening, which are activities that involve interpreting the meanings of others (*ibid.*). Some linguists refer to the same phenomenon as the distinction between active and passive language skills (for discussion, see e.g. Nation 2001: 24). However, this is a rather controversial topic and, thus, in this study, the terms *productive* and *receptive* will be used. Productive and receptive vocabulary knowledge or use link these two features explained above with vocabulary specifically. Productive vocabulary knowledge or use occurs when a language user recalls and produces a suitable word form through writing or speaking (Nation 2001: 25). On the contrary, receptive vocabulary knowledge or use occurs when a language user is reading or listening: he or she becomes aware of a given word form and recalls its meaning (*ibid.*). Of course, the two features presented here are to some extent overlapping as well as complementary in nature.

Another common feature that is usually discussed in relation to L2 vocabulary knowledge is the distinction between the size of vocabulary and the depth of vocabulary. The notion of *size of vocabulary* is fairly self-explanatory: it simply refers to the quantity of words a learner knows or the number of words for which the learner has at least some superficial or minimum knowledge of meaning (e.g. Qian 1999: 283, Qian 2002: 515). Sometimes, the term *vocabulary breadth* is also used. In this thesis, size of vocabulary, vocabulary size, breadth of vocabulary and vocabulary breadth are used interchangeably to refer to the number of words known by the L2 learner.

As for *depth of vocabulary*, it is more complicated to give a comprehensive definition. In general, it refers to the quality of knowledge: how well words are known by an L2 learner (e.g. Qian 1999: 283, Qian 2002: 515). This may be conceptualised through two approaches: the developmental approach and the dimensions (components) approach (Read 2000, Schmitt 2010). According to the developmental approach, depth of vocabulary knowledge grows incrementally, from not knowing a word to mastering the word (Schmitt 2010: 217). When considering the dimensions approach, on the other hand,

knowledge of a word is seen as consisting of different types of word knowledge (Schmitt 2010: 224). In this study, depth of vocabulary knowledge is mainly understood as developing stages of word knowledge. Nonetheless, the definitions given here overlap to some degree, and that is why some vocabulary knowledge frameworks will be presented in the next subchapter. When addressing depth of vocabulary knowledge, it is also possible to use the term *vocabulary depth*, and, in this thesis, the two terms are used interchangeably to refer to the same phenomenon. These two dimensions of vocabulary knowledge, i.e. vocabulary size and depth, are strongly associated, yet separate entities (Qian 1999). Hence, it is important to include both dimensions when studying learners' vocabulary knowledge.

2.3 Different perspectives on defining vocabulary knowledge

Many researchers have proposed different models for explaining what kind of a construct vocabulary actually is. As mentioned in the previous subchapter, defining vocabulary size is rather straightforward but defining vocabulary depth, on the other hand, is more complex. Definitions of the concept vary and the term has become quite disarrayed. Some linguists suggest a distinction to be made between the dimensions and the developmental approach (Read 2000, Schmitt 2010), whereas others have suggested presenting vocabulary depth through three approaches: precision of meaning, comprehensive word knowledge and network knowledge (Read 2004). In this subchapter, I will clarify these concepts more and present ideas and frameworks related to these approaches.

In the *dimensions approach*, knowledge of a word is seen as consisting of different subcomponents (Schmitt 2010). Similarly, in the *comprehensive knowledge approach*, knowledge of a word involves many features, such as orthographic, phonological, morphological, semantic, syntactic and collocational as well as pragmatic features (Read 2004: 211). One of the earliest contributors, whose work can be said to approach vocabulary along these lines, is Jack C. Richards. His framework dates back to the 1970s, and he was one of the first researchers to suggest that there is actually more to

knowing a word than just simple connections between forms and meanings. According to Richards (1976), knowing a word requires at least linguistic, psycholinguistic and sociolinguistic aspects. He further goes on to introduce seven aspects that are all involved in L2 vocabulary knowledge: word frequency, collocation, register, case relations, underlying forms, word associations, and semantic structure.

Richards' work was a starting point for Paul Nation who first started to develop his idea of aspects of vocabulary knowledge in the late 1980s and in the early 1990s. A framework that consists of nine aspects of vocabulary knowledge was published in 2001 (Nation 2001). These aspects are further divided into three categories: form, meaning, and use, and each aspect also shows a receptive or a productive feature (Nation 2001: 26). Moreover, Nation proposed that these nine aspects together illustrate what it means to actually know a word (*ibid.*). The framework is summarised on the following page in Table 1.

Nation's model is certainly comprehensive, but it is difficult to design a test format that would involve all aspects of vocabulary knowledge. Testing aspects of word knowledge seems to be very time-consuming if all aspects are to be included in tests (Meara 1996: 44, Read 2004: 217). Some efforts have been made but they have not been very encouraging (e.g. Read 2000: 178–180).

Form	Spoken	R	What does the word sound like?
		P	How is the word pronounced?
	Written	R	What does the word look like?
		P	How is the word written and spelled?
	Word parts	R	What parts are recognizable in this word?
		P	What word parts are needed to express the meaning?
Meaning	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?

Table 1. Aspects of vocabulary knowledge (Nation 2001: 27). Note: R = receptive knowledge; P = productive knowledge

In the *developmental approach*, vocabulary knowledge is seen as developing stages of mastery of a word (Schmitt 2010). In the *precision of meaning approach*, vocabulary knowledge is likewise seen as developing knowledge, ranging from having no knowledge of a word, through recognition and having a vague idea of its meaning, to having elaborated knowledge of the word (Read

2004: 211). As with the dimensions approach presented in the previous paragraph, it has also been contemplated in the case of this approach whether all words need to be known to the same degree (e.g. Read 2004: 213, Brown 2010: 85). Developing tests measuring vocabulary depth as developing stages of knowledge has proven to be more successful than with the dimensions approach. Some tests require precise knowledge and elicitation of definitions whereas other tests use self-report methods (Read 2004: 213–216). In this study, vocabulary depth is understood as developing degrees of knowledge and the test used to measure this is the Vocabulary Knowledge Scale (Wesche & Paribakht 1996) that will be more thoroughly presented in Chapter 5.1.2.2.

The *network knowledge approach* proposes that words are organised in lexical networks and that new words are incorporated into this network of already known words (Read 2004: 213). Accommodating new words into the network might also require restructuring the network (Read 2004: 219). It is suggested that L2 learners' lexical networks are less developed and smaller than those of L1 users (Meara 1996: 77). Instead of concentrating on separate words, this approach underlines the significance of the links between words (Read 2004: 219), and it seems that it is these links that are simpler with L2 learners than with L1 users (Meara 1996: 77). A test that is commonly used as a means of measuring vocabulary depth is the Word Associates Format (Read 2000), and, depending on the interpretation, this test can be seen either as a measure of the dimensions approach or the network approach. The test will be discussed in more detail in Chapter 5.1.2.1.

In sum, vocabulary size and depth are commonly seen as the two main dimensions of vocabulary knowledge. Furthermore, vocabulary depth is often approached in a number of different ways that were presented in this subsection. In previous studies, depth of vocabulary knowledge has most commonly been defined as different aspects of knowledge and this has been tested through the Word Associates Format (Meara 1996, Read 2004). However, this is only one way of representing and operationalising the concept of vocabulary depth. In order to reach a fuller understanding of the nature of vocabulary depth, a broader range of assessment tools needs to be used in

future studies (Read 2004: 223). Consequently, in the present study, stages of developing knowledge will be examined. Next, I will turn to discuss the significance of vocabulary knowledge in L2 learning in general.

3 Role of vocabulary knowledge in L2 learning in general

Vocabulary was long neglected in L2 teaching, and other components of language, such as grammar, received more attention in L2 learning (Laufer & Nation 2012: 163). Nevertheless, there is an evident consensus among linguists today that L2 vocabulary knowledge is a vital building block of L2 proficiency (Henriksen 2006). Many studies have constantly shown that learners with small vocabularies are less competent in a number of L2 skills than learners who know more words (e.g. Meara 1996). Furthermore, current research shows that vocabulary skills are clearly involved in nearly all aspects of L2 competence (Schmitt 2010: 4), and, in addition, it has been suggested that it is indeed vocabulary knowledge that constitutes “the single largest obstacle to advancement” (Laufer 1992: 101). Nation and Meara (2002: 43) point out, however, that it is not only knowing many words that is significant; in order to make use of vocabulary well it has to be available for use. Making the most use of what is already known can be achieved for example through developing fluency (*ibid.*). In sum, learning vocabulary poses both a quantitative and a qualitative challenge for L2 learners (Laufer & Nation 2012: 163). It is a quantitative challenge in the sense that vocabulary is an open set of thousands of words, and a qualitative one in the sense that learning vocabulary requires mastery of various features.

Despite the central role of vocabulary in all four L2 skills (reading, writing, listening, and speaking), there are some differences between the written and spoken modes. It has been suggested that more knowledge of words is required for writing and reading than for speaking and listening (e.g. Nation 2001: 125, Nation 2006: 79, Schmitt 2008). However, in the case of speaking and writing, for example, the difference might simply result from the topic and degree of familiarity: more weighty matters are often debated in writing and more casual issues are conversed through speaking (Nation 2001: 125). In L2 writing research, strong correlations have constantly been found between L2 vocabulary breadth, depth and writing skills, and L2 vocabulary knowledge is

indeed seen as a significant predictor of L2 writing skills (Henriksen & Danelund 2015: 29, 31). Additionally, more advanced L2 learners use richer vocabulary in their compositions than less advanced learners do (Nation & Meara 2002: 51).

In addition to the significant role of L2 vocabulary knowledge in productive language skills, vocabulary knowledge is vital in the development of receptive language skills as well. Much research has been conducted in the area of vocabulary and reading comprehension (see Chapter 4), but less is known about the connection between vocabulary and listening comprehension. There is not enough research to establish a firm number on how many words a learner should know for listening comprehension to be successful. However, it has been suggested that a learner has to know about 95% of spoken input in order to reach adequate comprehension (e.g. Schmitt 2008: 331, Schmitt *et al.* 2015: 4). Taking into consideration the fact that higher-frequency words are often used more in spoken rather than in written language, the estimated numbers of word families needed for listening are lower than the numbers for reading (Nation 2006: 79). It has been suggested that somewhere between 5000 and 7000 word families are needed for listening whereas the numbers for reading are often slightly higher, ranging most commonly from 8000 to 9000 word families (e.g. Nation 2006: 59, Schmitt 2008: 359). Of course, the figures may vary considerably depending on the kind of input or the level of understanding desired (Nation 2006). More research is still necessary in this field in order to establish more reliable figures (Schmitt 2008: 331).

Nevertheless, reading and listening are both clearly linked to vocabulary knowledge. A study of Iranian learners of English (Mehrpour & Rahimi 2010) revealed that both general and specific vocabulary knowledge affect learners' performance in both reading and listening, the impact being bigger on reading than on listening, though. Similarly, in another study involving Spanish learners of English (Proctor *et al.* 2004), a significant link was found between vocabulary knowledge and both reading and listening comprehension. Moreover, a rather close relationship has been found between EFL learners' vocabulary breadth and their performance in reading, listening, and other

formal tests of English proficiency (Nation & Meara 2002: 51).

The role of vocabulary knowledge in L2 learning is particularly noteworthy when considering the English language because English is a language that has a large vocabulary and the number of words is constantly growing. In addition, there are many very specific vocabulary registers for particular areas of discourse, such as academic, legal and business-related vocabularies, and from the viewpoint of EFL learners, it is worthwhile to learn different kinds of vocabulary (Nation & Meara 2002: 50–51). As has been shown in this subchapter, learning vocabulary is indeed crucial for all four skills of L2 knowledge to develop. It has to be also noted that while learning more vocabulary leads to improved L2 proficiency, the relationship between the two is not merely one-way but rather a complementary one. Nation clarifies that “vocabulary knowledge enables language use, language use enables the increase of vocabulary knowledge, knowledge of the world enables the increase of vocabulary knowledge and language use and so on” (1993, as cited in Nation & Waring 1997: 6). As learners learn more vocabulary, they can engage themselves in more challenging situations where L2 is used, such as in reading more demanding texts or participating in discussions in L2. These situations will then lead to learning more vocabulary which again enables L2 learners to pursue even more demanding tasks and so on. It is a cycle where learning L2 vocabulary, engaging in different L2-related situations and tasks as well as world knowledge complement each other.

Of course, vocabulary knowledge is only one of numerous components which together affect an individual’s overall L2 skills. Based on the findings presented in this subchapter, though, it is not surprising that researchers constantly underline the importance of vast vocabulary instruction in L2 teaching (Laufer 1992: 101). It is indeed crucial to recognise the fundamental relevance of vocabulary knowledge in overall L2 proficiency. In the next chapter, reading as an L2 skill will be examined more thoroughly.

4 Reading as one of four skills in L2

When discussing L2 proficiency, it is common to divide it into four skills, i.e. reading, writing, listening, and speaking. In the first subchapter of this section, I will look more closely at reading as an L2 skill. Attention will first be paid to analysing reading in general and to presenting some fundamental differences between L1 and L2 reading. Then, L2 reading will be examined in more detail. In the second subchapter, I will move on to discuss the impact of vocabulary knowledge on reading comprehension. In the last subchapter of this section, some earlier research in the area of vocabulary knowledge and reading comprehension will be presented.

4.1 What is involved in reading

In the research of both L1 and L2 reading, two views are often referred to (Nassaji 2013: 262). In the more traditional, psycholinguistic models, reading is analysed in a very linear manner where lower-level textual components follow one another. The reader decodes the text into words, assembles this information into phrases, sentences and other larger units and then constructs the meaning of the text (Koda 2012: 158, Nassaji 2013: 262). In the interactive models that are currently perhaps more popular than the psycholinguistic models, both lower-level and higher-level processes are involved in a complex manner (Nassaji 2013: 261–2). Higher-level processes include analysing and interpreting syntactic and semantic properties of the text as well as integrating ideas with previous knowledge (*ibid.*). Lower-level processes explained above work together with these higher-level processes so that the reader arrives at an understanding of the text (e.g. Grabe 1991: 379, Perfetti, Landi & Oakhill 2005: 228). The level of understanding can vary, but usually the levels of understanding are assumed to be ordered: one has to understand the lines before reading between them (Alderson 2000: 8).

It is important to acknowledge that L2 reading differs fundamentally from L1 reading (Birch 2011: 497). When learning to read in L1, children already know

the spoken language and learning to read mainly deals with learning to recognize what they already know (Alderson *et al.* 2015: 6). When learning to read in L2, on the other hand, the learner usually knows how to read in one language already. The learner knows a set of rules concerning the relationship between graphemes and phonemes in L1, but often a new set of rules has to be learnt for L2 reading, which can be a slow process; automatization and fluency take time to develop (*ibid.*). Naturally, there are differences in how easy or difficult it is to learn to read in a particular language. For example, Finnish is a highly transparent language when it comes to orthography, whereas in English the relationship between graphemes and phonemes is more complex and opaque (Alderson *et al.* 2015: 69).

It is sometimes proposed that poor L2 reading performance results from poor L1 reading skills but this seems to be an inadequate statement, as available reading research strongly suggests that L2 reading is more of a language problem than an actual reading problem (Alderson *et al.* 2015: 70–71). As a matter of fact, there are some advantages that L2 readers have over L1 readers (Grabe 1991: 386–387). L2 learners are often older than L1 users learning to read and they in general know more about the world; their conceptual sense of the world is more developed. They are able to make logical inferences and use more metacognitive strategies when reading a text. However, there are still some fundamental differences between EFL learners and L1 speakers of English. Not surprisingly, EFL learners know fewer words, and, secondly, they also have less knowledge about the meaning of these words (Wallace 2007: 193). In order to develop EFL learners' reading skills, Wallace (*ibid.*) underlines the significance of vocabulary knowledge in reading and insists on more extensive vocabulary teaching.

L2 reading is clearly a complex phenomenon where the reader interacts with a given text. On top of linguistic knowledge, the learner can also resort to reading strategies. For example, he or she might make inferences and guess the content of unfamiliar words in context, identify the text type and structure, make expectations and predictions of the content, and grasp the main idea of a paragraph (Nation & Coady 1988: 102, Laufer 1997: 20). It is evident that

cognitive abilities as well as linguistic knowledge are involved in reading (Koda 2012: 158) and the learner uses not only the surrounding text but also other knowledge sources to gain information needed in order to understand a particular text (Paribakht & Wesche 1997: 175–176). Skilled L2 readers are able to use both general textual features (e.g. morphological, syntactic, and discourse-related information) and background knowledge in text processing (Nation & Coady 1988: 102). In addition, they can also tolerate a small number of unfamiliar words without comprehension being disrupted (Carver 1994). However, a crucial component of reading comprehension is vocabulary knowledge, and that is why some light will next be shed on how vocabulary knowledge relates to reading comprehension.

4.2 Vocabulary knowledge affects reading comprehension

As noticed in Chapter 3, vocabulary knowledge is vital in L2 learning in general. Moreover, it is fundamental in reading comprehension specifically, and this has been proposed and supported by many authors as well (e.g. Nation & Coady 1988, Laufer 1992, 1997, Wallace 2007). Vocabulary knowledge is actually understood to be the most obviously perceivable component of the reading skill (Nation & Coady 1988: 98). In addition, vocabulary difficulty, estimated in terms of for example word frequency, familiarity and word length, is possibly indeed the most indicative predictor of general readability (Nation & Coady 1988: 97). Therefore, it is easier for an L2 learner to read texts that involve high-frequency words and words that are preferably quite short. There are, of course, a number of other factors that make a text easy or difficult for a learner to read, but text readability is still most accurately predicted by measuring vocabulary (*ibid.*).

In the case of insufficient vocabulary knowledge, using reading comprehension strategies and grasping the text's main idea would sometimes and to some degree result in some kind of understanding; nevertheless, the comprehension would still remain partial (Laufer 1997). In addition, previous research undeniably shows that the link between reading comprehension and vocabulary knowledge is stronger than between reading comprehension and

other subcomponents of reading, such as grammar knowledge or background knowledge (Mehrpour & Rahimi 2010: 293). Of course, other skills are needed in L2 reading as well, but when considering grammar knowledge, for example, it seems that it is less significant for L2 reading than vocabulary knowledge (Alderson *et al.* 2015: 103). Furthermore, having reviewed many studies, Laufer (1997: 21) concludes that lexical problems undeniably hinder reading comprehension and that the threshold for successful reading is primarily related to vocabulary knowledge. In the case of EFL in particular, vocabulary size strongly limits the number of texts a learner can read with ease (Nation & Meara 2002: 51).

Many studies have examined the vocabulary size needed to understand a variety of texts. Laufer (1992) studied L2 lexical knowledge and the readers' general academic ability. Based on her findings, she proposes three predictions related to L2 reading comprehension and vocabulary size. Firstly, with knowledge of fewer than 3000 word families, reading comprehension will remain unsatisfactory, and no degree of general academic ability, including L1 reading proficiency, can balance this shortcoming. Secondly, with a rich vocabulary size of about 5000 word families, the learner will be a good reader regardless of the general academic ability. Thirdly, with vocabulary knowledge ranging from 3000 to 4000 word families, reading proficiency could or could not be impacted by general academic ability. However, on this lexical level also, general academic ability is of little significance as well, and, consequently, on all levels of knowledge, L2 vocabulary size is more efficient in predicting L2 reading performance than the informant's general academic ability (Laufer 1992: 101). Laufer (1992: 100) suggests, thus, that knowing about 3000 word families is enough to ensure a starting point for L2 reading comprehension. Correspondingly, Nation and Waring (1997: 10) propose that the lexical threshold for reading comprehension is somewhere between 3000 and 5000 word families.

Many researchers, however, propose somewhat larger vocabularies that are needed for L2 reading to be successful (e.g. Nation 2006, Schmitt 2008). Estimates of *text coverage* needed for successful L2 reading vary

considerably: according to many, it is somewhere between 95 and 99% (Mehrpour & Rahimi 2010: 294). Text coverage needed for reading comprehension refers to how many word families or lexical units are needed for understanding a text; in other words, it is the percentage of running words that a learner needs to know in order to comprehend a given text (Nation 2006: 61). A 95% text coverage would mean that approximately one word out of twenty is unknown, and with a 98% text coverage one word out of fifty would be unknown. In recent studies, the most commonly suggested text coverage figure is 98% (e.g. Schmitt, Jiang & Grabe 2011: 26, Schmitt *et al.* 2015: 1)

Similarly, Nation (2006), after scrutinizing numerous previous studies, suggests that a 98% text coverage is needed for satisfactory, unaided reading comprehension. In practice, this means that a learner is required to know between 8000 and 9000 word families (Nation 2006: 59). To compare, a well-educated native speaker of English is estimated to have knowledge of about 20 000 word families (Nation 2006: 60). Undoubtedly, the number of words needed for reading depends also on what type of a text is in question. For a text coverage of 98%, when reading a novel, a vocabulary of about 9000 word families is needed (Nation 2006: 71). A similar size of vocabulary, from 8000 to 9000 word families, is also required for reading a newspaper in order to reach a 98% coverage (Nation 2006: 72). Similar numbers have also been proposed by for example Schmitt (2008: 329) who suggests that somewhere between 8000 and 9000 word families are required for reading. In the case of graded readers with simplified text, on the other hand, only 3000 word families are needed to reach an adequate understanding of the text (Nation 2006: 72).

In the light of these numbers, it is valuable to look at the number of words Finnish learners of English know. In their study, Jaatinen and Mankkinen (1993) studied university students and discovered that an average English major student knows about 18 100 lexemes. Likewise, in a recent MA thesis, Pirilä (2012) studied advanced learners of English. The informants were English majors at the University of Turku, and the study revealed that their vocabularies comprised about 22 000 lexemes in general. In both of these studies, the concept of a word family was not applied. Instead, Jaatinen and

Mankkinen (1993) as well as Piriälä (2012) used the concept of a lexeme which was defined as something “between a lemma and a word family” (Piriälä 2012: 7): lexeme refers essentially to a dictionary entry that includes compounds and phrasal verbs but excludes proper nouns and acronyms, for example (*ibid.*). Both Jaatinen and Mankkinen’s (1993) as well as Piriälä’s (2012) numbers seem high, especially when considering Nation’s (2006: 60) estimate of an educated native speaker’s vocabulary size: 20 000 word families. However, it is likely that the vocabulary size figures obtained in the two studies would have been somewhat smaller if they had been expressed in terms of word families due to the more inclusive nature of word family. In another recent MA thesis, Ala-Akkala (2010) studied upper secondary school students. According to the results of her study, Finnish upper secondary school students’ average receptive vocabulary size was 3 700 word families. Groups of learners studied in these three studies presented were different, but nonetheless, the results are still surprisingly opposed. The results of these studies will be examined more thoroughly and compared to the results of the present study in Chapter 7.1.

It is clear that many studies concentrate mostly on how large a vocabulary is required for reading in an L2 and how many words a learner knows. However, it is not enough to focus merely on the number of words known; both size and depth of vocabulary are equally significant in reading comprehension (Qian 2002: 517, Wallace 2007: 189). Vocabulary is learnt incrementally and, thus, it is likely that words learnt first have more depth than words that were only recently learnt (Qian 2002: 517–518). Having a large vocabulary will undeniably provide the learner with a large database which to use when inferring meanings of unknown words, for example, but this work will most likely be still improved if the learner has deeper vocabulary knowledge to use as well (*ibid.*). This is also what Qian’s study (2005) strongly suggests: learners with more depth of vocabulary knowledge are capable of making better use of context and, hence, are more inclined to succeed in guessing the meaning of unknown vocabulary items. Given the interactive relationship between vocabulary breadth and depth, it is worthwhile to measure both dimensions when making predictions of learners’ reading comprehension skills (Qian 2002:

532). As with L2 skills in general, better vocabulary knowledge often leads to higher achievements in reading comprehension and these achievements in turn will help to acquire more vocabulary (Nation 2001: 144, Qian 2005: 48).

4.3 Previous studies

Studies concerning the relationship between vocabulary size and reading comprehension are numerous, and the connection between the two is well-established, as was seen in Chapter 4.2. Many studies report relatively high intercorrelations between the two (e.g. Laufer 1992, 1996). The number of studies concerning both vocabulary breadth and depth as well as reading comprehension is very limited; nevertheless, some studies do exist. Previous studies related to the topic of the present study will be introduced next.

Rashidi and Khosravi (2010) studied Iranian EFL learners by examining the role of the informants' vocabulary size and depth in reading proficiency. They also compared two distinct groups, namely, learners with high or low vocabulary breadth and depth. In their study, vocabulary depth was measured through the means of Word Associates Format and, thus, vocabulary depth was seen as dimensions rather than as degrees. The learners' minimum vocabulary size in the study was 3000 word families. The findings reveal high and positive correlations between vocabulary size, depth and reading comprehension. When examining reading comprehension predictions, vocabulary depth afforded a noteworthy contribution. Moreover, students with large vocabulary size and strong vocabulary depth scored better on the reading comprehension test.

Similarly, in Li's study (2015), a moderate, positive correlation was found between vocabulary breadth and reading comprehension and also between vocabulary depth and reading comprehension. According to this study, vocabulary depth can provide EFL learners with richer lexical knowledge. What this means is that readers can then integrate knowledge of individual words with sentences and, thus, arrive at reading comprehension as accurate as possible. Furthermore, another study by Carlo *et al.* (2004) suggests that when

EFL learners' vocabulary and word analysis skills are developed, their reading comprehension skills improve as well.

Perhaps the biggest contributor in this field, however, is David D. Qian. He has been one of the first researchers to acknowledge the significance of depth of vocabulary knowledge as a vital element of reading comprehension skills along with vocabulary size. He has conducted various studies, two of them having had major influence in this field. In the first study (Qian 1999), he examined 74 Chinese and Korean university-level students who were EFL learners. Qian set out to discover how results on vocabulary breadth, depth and reading comprehension tests relate to one another. Moreover, another goal of his was to describe vocabulary depth and the kind of prediction it could offer in predicting learners' reading comprehension skills, both together and separately from the prediction provided by vocabulary breadth alone. A background questionnaire and four tests were used for the purposes of Qian's study. The Vocabulary Levels Test (see Chapter 5.1.1.1) was used to assess the participants' vocabulary breadth: each participant knew at least 3000 word families. A standardized TOEFL (Test of English as a Foreign Language) multiple-choice reading comprehension test was used to measure informants' reading skills. Two tests were used to study vocabulary depth: the Word Associates Format (see Chapter 5.1.2.1) and a test designed by Qian himself in order to complement the Word Associates Format. The first test measures vocabulary depth as dimensions and the second morphological aspects of vocabulary knowledge. High and positive correlations were established between the four test scores. A high correlation was also found between vocabulary depth and vocabulary size scores, which clearly suggests that these two features are interconnected and interdependent. Moreover, both vocabulary breadth and depth provided significant contributions to the prediction of reading proficiency. What is noteworthy here is that vocabulary depth indeed added 11 percentage points of explained variance in reading test scores beyond the prediction provided by vocabulary breadth alone. The study undeniably supports the significance of vocabulary depth in reading comprehension.

Qian continued his work in yet another study (Qian 2002). The informants were 217 students whose level of English was intermediate or beyond intermediate. They were all participants in an intensive English as an L2 program at the University of Toronto and came from multiple L1 backgrounds. The aim was to study the contribution of certain aspects of vocabulary depth (synonymy, polysemy, collocation) as well as vocabulary size to reading comprehension. The tests used were similar to the ones used in Qian's previous study. Size of vocabulary was assessed with the Vocabulary Levels Test, reading comprehension with a version of the TOEFL reading comprehension test, and finally, vocabulary depth was assessed with a refined version of the Word Associates Format. In addition, a TOEFL Vocabulary Item Measure was used as quality control. The study showed that all three vocabulary measures correlated with reading comprehension test scores to a resembling extent. When looking at the reading comprehension test results, it was calculated that vocabulary depth scores alone explained about 59% of the variance in the results whereas the scores on vocabulary size alone explained about 54% of the same variance. The results of this study imply that using more variables than just one alone will yield better results in predicting performance on reading comprehension tests.

In order to develop a new measure for assessing L2 vocabulary depth, Qian and Schedl (2004) designed a new test based on the Word Associates Format. The new test would essentially measure knowledge of word meaning, polysemy, synonymy, and collocation. Together with the new test that they called the Depth of Vocabulary Knowledge Measure, they used a version of a TOEFL vocabulary test as quality control and a TOEFL reading test to assess reading comprehension. No test for vocabulary size was involved. The informants were 207 international students in Canada with different L1 backgrounds. The results of the study revealed strong correlations between all three test measures. When it comes to the significance of depth of vocabulary knowledge in explaining variance in reading comprehension scores, the scores on the vocabulary depth test explained 55% of the variance in reading comprehension. According to Qian and Schedl (2004), the results suggest that both vocabulary breadth and depth can, and should, be used to predict and

explain L2 learners' reading proficiency. Accordingly, they still underline the importance of developing measures of L2 vocabulary depth further in order to better understand the role of vocabulary depth in reading comprehension (Qian & Schedl 2004: 47).

When examining Qian's (1999, 2002) and Qian and Schedl's (2004) studies, some significant observations have to be made. The more recent studies (Qian 2002, Qian & Schedl 2004) managed to overcome one of the limitations of the earlier study; in the more recent studies, the participants were of mixed L1 backgrounds. However, there is still a limitation that is difficult to overcome: the complexity of depth of vocabulary knowledge. This variable consists of many dimensions which makes it hard to measure. The measures used in these studies represent only partially the concept of vocabulary depth and, thus, it is vital to continue research in this field. Consequently, on the basis of his work, Qian (1999, 2002) calls both for recognition of the significance of improving EFL learners' vocabulary depth in different learning processes and for more research in this area.

The previous studies presented in this subchapter strongly propose that taking depth of vocabulary knowledge also into consideration when studying the relationship between L2 vocabulary knowledge and reading comprehension skills instead of only looking at learners' vocabulary size is useful and offers new perspectives in this matter. As informative as these previous studies are, more research is, however, definitely needed in this field. Most studies use the Word Associates Format as a tool for measuring vocabulary depth. In my study, though, another tool, the Vocabulary Knowledge Scale (see Chapter 5.1.1.2), will be used. Moreover, most studies mentioned here only involve university-level students. In the present study, the focus is also on upper secondary school students who most likely are not as advanced as university-level students. In addition, comparison is made between two groups: a group of upper secondary school students and a group of advanced university students of English. The methodology and the participants of this study will be presented more carefully in the following chapter.

5 Methodology

In this chapter, the methodology of the present study will be introduced. I will start by introducing some most important principles related to the assessment of L2 vocabulary breadth and depth as well as reading comprehension. Some most common test formats will also be introduced as well as the tests used in the present study. It would have been possible to discuss some of these matters already in the theoretical section of the thesis, but I decided to present all relevant information in this chapter due to clarity. It is hopefully easier for the reader to follow the discussion and comparison of different test formats as well as understand my choice of test formats now that they are examined in one chapter. After discussing the assessment of L2 skills, I will move on to introducing the participants of the study in Chapter 5.2. The procedure of test administration will be described in Chapter 5.3, and, lastly, I will shortly explain the statistical methods used in the present study in Chapter 5.4.

5.1 Assessing L2 skills

In this subchapter I will discuss assessing and evaluating various L2-related skills. Firstly, I will look into the assessment of L2 vocabulary size. Secondly, the assessment of L2 vocabulary depth will be discussed. Thirdly, I will present some ideas related to the assessment of reading comprehension skills. The tests used in the present study will also be introduced in this chapter.

5.1.1 Assessing L2 vocabulary size

As shown in the previous chapters, there are many studies where EFL learners' vocabulary size has been measured. There are various tests that have been used for this purpose. In this subchapter, I will discuss some key issues related to the assessment of L2 vocabulary size. Moreover, I will also introduce two known and commonly used test formats: the Vocabulary Levels Test (Nation 1983, 1990; Schmitt *et al.* 2001) and the Eurocentres Vocabulary Size Test (Meara & Jones 1988). In this study, a more recent test for measuring L2 vocabulary breadth, the Vocabulary Size Test (Nation & Beglar 2007), will

be used, and I will also justify my choice in this subchapter.

Like any other language proficiency test, a vocabulary size test has to meet certain requirements regarding validity, reliability and practicality. Main issues in measuring L2 vocabulary size concern the selection of test items and formats of testing. Two main approaches in selecting test items include sampling from a dictionary and using frequency lists. If sampling from a dictionary, the dictionary has to be comprehensive enough so that a proportional sample can be drawn from it (Wesche & Paribakht 1996: 15). This method has traditionally been more popular when measuring L1 speakers' vocabulary size (Nation 2001: 363). Frequency lists, on the other hand, are based on corpora, and they demonstrate words grouped by the frequency of their occurrence (Wesche & Paribakht 1996: 15). The basic principle behind this approach is that the more often a word occurs in certain corpora, the more probable it is that a learner knows the word (*ibid.*). Vocabulary tests based on this method are often used to test non-native speakers (Nation 2001: 363).

There are many test formats that can be used in order to elicit information on learners' receptive vocabulary size. Two very common test formats are the multiple-choice format and the checklist format (Wesche & Paribakht 1996: 18–19). In the multiple-choice format, the learner has to choose the correct answer among given alternatives. In general, the tests are easy to administer, score, and analyse (*ibid.*). Nevertheless, there are still problems related to this method. An obvious shortcoming of the method is the possibility of guessing the right answer or arriving at the right answer using elimination strategies (Gyllstad, Vilkaite & Schmitt 2015: 278). Moreover, it is possible that the informant knows another meaning for the word, but not the one sought, or that he or she is not able to choose the right answer due to not understanding the given alternatives (Wesche & Paribakht 1996: 19). Constructing a multiple-choice test is thus laborious, but the format is still one of the most popular language test formats (*ibid.*). In the checklist format, on the other hand, the learners simply mark whether or not they know a given word. Sometimes, this format is also called the Yes/No method. This format also has its advantages and disadvantages, which will be discussed in more detail in 5.1.1.2. Next, a

type of multiple-choice test, the Vocabulary Levels Test, will be presented.

5.1.1.1 Vocabulary Levels Test

Vocabulary Levels Test (VLT) is a prevalent test format with many strengths: it is simple and quick to take, mark and interpret (Nation 2001: 21). It has to be noticed, though, that the test was not initially designed as a test of vocabulary size, but it became popular as a means of estimating learners' vocabulary size anyway, most likely due to the absence of a standardized vocabulary size measure (Read 2000: 118, Schmitt 2010: 198). The VLT was originally designed by Nation (1983, 1990), but in this paragraph, I will refer to the versions updated by Schmitt *et al.* (2001). This is also supported by Nation himself, who suggests replacing the original version with Schmitt *et al.*'s versions and sees these new versions as "a major improvement" to the original (Nation 2001: 416). Schmitt *et al.* (2001: 58) explain that the updated VLT provides a profile of language users' vocabulary knowledge. The VLT tests four frequency levels (2000, 3000, 5000 and 10 000 levels), and, furthermore, the final section tests academic vocabulary knowledge (Schmitt *et al.* 2001: 58). The authors designed two versions of the test, Version A and Version B, that produce similar results but are not completely equivalent (Schmitt *et al.* 2001: 78).

The test uses a type of form-meaning matching where the learner has to pick the correct word to go with each definition given. Each cluster includes six options and three definitions (Schmitt 2010: 197). Each level, then, consists of ten clusters (*ibid.*). An example from the 3000 word level is given below.

1 assemble	
2 attach	___ look closely
3 peer	___ stop doing something
4 quit	___ cry out loudly in fear
5 scream	
6 toss	

This format was initially developed for two reasons: to minimise guessing and to involve as little reading as possible (Read 2000: 119). Despite high

practicality of the test, the new versions that were at the time only under development as well as the original version ought to go through more validation and analysis if they are to be used as a reliable method of making educated estimates on learners' vocabulary knowledge (Read 2000: 124).

5.1.1.2 Eurocentres Vocabulary Size Test

Another well-known measure of vocabulary size was developed by Meara and Jones (1988). Since the development of this tool, many updates and versions have been made mainly by Meara and his colleagues (Read 2000: 127). The Eurocentres Vocabulary Size Test (EURO) differs from the VLT in many ways. It was designed to measure learners' vocabulary size in an effective and practical way, and the basic idea behind the test is that the learners are simply requested to indicate whether or not they know a given word (Meara & Jones 1988: 80–81). What makes this type of checklist format special is that it involves both real and imaginary words; usually two real words per every imaginary non-word (Meara & Jones 1988: 81–82). The initial versions were designed to be administered on a computer but there are also versions that can be done through the more traditional pen-and-paper method. The computerized versions are very easy to use: they are quickly done and automatically scored (Read 2000: 127). The pen-and-paper versions are also very practical as numerous words can be covered within a reasonable time frame and the task of the informants is simple (Schmitt 2010: 200). An example taken from Level 1 of the test (*ibid.*) looks like this:

- | | | | |
|--------------|-----------------|-----------------|---------------|
| 1. ___ obey | 2. ___ thirsty | 3. ___ nonagate | 4. ___ expect |
| 5. ___ large | 6. ___ accident | 7. ___ common | 8. ___ shine |
| 9. ___ sadly | 10. ___ balfour | | |

Even if the format looks quite convincing, there are still serious problems related to this kind of testing method. Most importantly, it can be contemplated what 'knowing' a word actually means for an individual informant (Schmitt 2010: 200). An informant might tick a word even if they have only a faint idea of its meaning or simply just have seen it somewhere with no idea of the meaning; another informant might want to be sure of the meaning before

ticking a word. Despite imaginary non-words used to overcome this problem, the problem nonetheless exists and is indeed very hard to overcome. EURO does not require the learner to actually demonstrate any knowledge of test items, and this is perhaps the most serious shortcoming of the test. EURO has, nevertheless, gained popularity as a vocabulary size test format (Read 2000: 132). It certainly has its advantages but there are serious shortcomings as well. Hence, I decided to opt for another test format that is used for the purposes of the present study.

5.1.1.3 Measure used in this study: Vocabulary Size Test

In order to overcome some shortcomings of many previously designed vocabulary size tests, Nation and Beglar (2007) created another test that they simply named the Vocabulary Size Test (VST). The test is freely available for teaching and research purposes and it is possible to take the test either online or through the more traditional pen-and-paper format. In this study, the pen-and-paper format was used. The test is not included in the section of Appendices as it would take up much space and as the whole test is available online (http://www.lex Tutor.ca/tests/levels/recognition/1_14k/). A PDF-version can also be found through the same address. As opposed to the VLT, the VST is not a diagnostic test but a test used to elicit knowledge on how much vocabulary a person knows (Nation & Beglar 2007: 10).

The VST is based on frequency levels that were sampled according to word families' occurrence in the British National Corpus (BNC) (Nation 2012). Taking into account the very formal, written nature of the BNC, however, some changes had to be made into the fourteen 1000 BNC word lists compiled earlier by Nation (2006). In these earlier lists, words like *hello* and *sun* appeared only in the 4th 1000 level, and, on the other hand, words like *civil* and *commission* appeared among the most frequent 1000 words (Nation & Beglar 2007: 10). Accordingly, the spoken section of BNC was used to revise the first twelve 1000 frequency levels. The changes in frequency levels were rather small, but, nevertheless, the ordering of the words was now more sensible and, above all, more suitable for compiling the VST (*ibid.*).

There are currently at least three available versions of the VST. The original test is based on the most frequent 14 000 English words and the test itself contains 140 items (ten items per each frequency level). Two other versions are somewhat different: they are based on the most frequent 20 000 English words but they only contain 100 test items (Nation 2012). For the purposes of this study, the original version of 140 test items was chosen as it has been through more evaluation and validation processes than the two more recent versions (Nation 2012). Moreover, it was considered more reliable as more items are used to test a single frequency level: 100 word families are represented by one test item whereas in the other two versions, one test item actually represents 200 word families. The format of the VST is a monolingual multiple-choice format where the learner has to circle the alternative that is closest in meaning to the item being defined. Each test item is presented in a “short, non-defining context” (Nation & Beglar 2007: 12). Furthermore, whenever feasible, the words used in the alternatives were of higher frequency than the test item in question (Nation & Beglar 2007: 11–12). An example is provided below:

- STONE: He sat on a **stone**.
- a. hard thing
 - b. kind of chair
 - c. soft thing on the floor
 - d. part of a tree

Scoring the test is very straightforward. For each test item, there is only one correct answer, and each test item is worth one point. No points are subtracted for incorrect answers. The maximum score for the test is, thus, 140 points. This number has to be then multiplied by 100 to get an estimate of the informant’s total vocabulary size (Nation & Beglar 2007: 12). For the purposes of this study, however, the test was shortened. This was done mostly because of time restraints. The number of test items was reduced to 100: the last forty questions testing the 11th to the 15th 1000 frequency levels were deleted. According to the authors of the test, this is perfectly appropriate, as long as the informants do a few levels beyond their expected proficiency level (Nation &

Beglar 2007: 11). This way, it was possible to do this part of the test session within the time limit of twenty minutes.

The main differences that separate the VST from the widely used VLT and EURO relate both to the selection of the test items and to the format. All of these tests are based on different sets of frequency levels, and it can always be discussed which frequency level count is the most suitable for measuring L2 vocabulary size. However, EURO is based on Thorndike and Lorge's frequency count, which, firstly, is quite outdated as it was compiled already in the 1940s, and, secondly, is based on lemmas and not word families. Bearing in mind the more inclusive nature of word family, it is questionable if frequency counts based on lemmas are as suitable for assessing receptive vocabulary size as frequency counts based on word families. Learners with some L2 vocabulary knowledge also have understanding of some word building strategies; in addition, they can also comprehend various relationships between "regularly affixed members of a word family" (Nation & Beglar 2007: 10).

Furthermore, the format of the VST differs radically from the format used in EURO. Both the multiple-choice format used in the VST and the checklist format used in EURO are easy to administer and score. However, as explained earlier, EURO does not provide any actual demonstration of knowledge and it relies on self-report only. Furthermore, the VST is a recently developed test that has already gained popularity as a tool of assessing EFL learners' size of vocabulary: it is convenient to use and it does not seem to be very affected by blind guessing either (Gyllstad, Vilkaite & Schmitt 2015: 281, 292).

The VST and VLT both use a type of multiple-choice format. However, in the VST, the distractors and the correct answer are usually somehow related or they share some elements of meaning whereas in the VLT, the distractors may be totally unrelated; thus, informants can choose the correct answer even with only a little knowledge of a given word (Nation & Beglar 2007: 11). In the VST, on the other hand, informants are usually required to have a somewhat more developed understanding of a particular word meaning (*ibid.*). Hence, the VST

is a little more challenging than the VLT. Another important difference between these two tests lies in the form of the questions. In the VLT, the words are simply given without any context whereas in the VST, the words are given in a slightly different format: the word in question is presented both alone and in a short sentence where the particular word is used in context. According to Nation (2001: 353–354), sentence contexts should be used where possible because this will give the learners the greatest chance of showing what vocabulary knowledge they have.

The three tests presented all elicit information on learners' receptive vocabulary size. Subsequently, they provide little to no information on learners' productive vocabulary skills, that is, how well learners could use the tested words in writing and speaking. There are, nevertheless, tests designed to measure learners' productive vocabulary size, such as Laufer and Nation's Productive Vocabulary Levels Test, Meara's P_Lex as well as a number of other methods: for example, different type-token-based methods and error recognition (Schmitt 2010: 203, 208, 212; Wesche & Paribakht 1996: 23–24). In the present study, however, in addition to L2 vocabulary size and depth measures, reading comprehension skills were assessed as well. Considering the nature of reading as a receptive skill, it was decided that, for the purposes of this study, a receptive vocabulary size test would be the most suitable, and for the reasons stated above, the VST was chosen as the vocabulary size test used in the present study.

5.1.2 Assessing L2 vocabulary depth

Contrary to the assessment of L2 vocabulary size, ways to measure L2 vocabulary depth seem limited. Due to the complex nature of this feature of vocabulary knowledge, there are only a few test formats that measure vocabulary depth. Earlier studies mostly seem to consider this feature as aspects of knowledge (e.g. Qian 199, 2002; Read 2000) or as degrees of developing knowledge (e.g. Wesche & Paribakht 1996). A major difficulty lies indeed in operationalising aspects of vocabulary knowledge. Different frameworks propose numerous different aspects; for example, Nation's

framework (2001) involves nine aspects. Testing all aspects would be highly time-consuming both for the researcher and the informants (Read 2000: 178–180).

Sometimes, interviews are used in order to elicit information on learners' depth of vocabulary knowledge (Nation 2001: 356). Interviews are hardly a suitable method for testing large samples of informants, and this method was thus not chosen for this study. Two test formats that have been used in a number of studies are the Word Associates Format (Read 2000) and the Vocabulary Knowledge Scale (Wesche & Paribakht 1996). The two formats look at vocabulary depth from different viewpoints; both formats will be presented in more detail in the following subsections. For the purposes of this study, Vocabulary Knowledge Scale was chosen to measure the learners' developing depth of vocabulary knowledge.

5.1.2.1 Word Associates Format

The Word Associates Format (WAF) was initially created by Read in the 1990s but, since then, many improvements and updates have been made to the test both by Read and other linguists (Schmitt 2010: 226). I will refer here to Read (2000) for a comprehensive description of the WAF. The key concept in WAF lies in word associations. Each cluster is built of a stimulus word and eight alternatives, and the learners' task is to circle four correct associates, or in other words, words that are related in meaning to the test item (Read 2000: 181). Most commonly, the options are divided into two groups of four possible associates, as can be seen below:

common

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

The two groups consist of different kinds of words. Adjectival forms are placed on the left side, and they are either synonyms of the test item or they illustrate one of its meaning aspects (Read 2000: 184). Nouns are placed on the right side, and they are words that often occur with the test item, i.e. collocates

(*ibid.*). With the aim of decreasing the likelihood of guessing, the number of correct associates on each side is not fixed, and, thus, the right pattern may be one of three options: 2-2, 3-1, or 1-3 (*ibid.*).

The WAF has been used in many studies in order to obtain information on informants' depth of vocabulary knowledge. Some previous studies were presented in Chapter 4.3. A more comprehensive listing of studies involving the WAF can be found in Schmitt (2014). The task of the informants is straightforward and it looks like scoring of the test would be quite simple as well. Unfortunately, scoring the test has proved to be complicated, mostly due to guessing, which, despite the efforts made to reduce it, still occurs (Read 2000: 185). Furthermore, there is no clear answer as to what to do with 'split' scores (Schmitt 2010: 227). If a learner manages to choose two correct associates together with two distractors, it is not clear how this should be interpreted into scores. Responses like this may indeed be the result of guessing, and, hence, it is rather dubious to merely score the number of right associates and ignore the number of distractors marked (*ibid.*). In some studies, the researchers have only accepted such items as correct where all four associates have been correctly marked (Schmitt 2010: 228). Another limitation of WAF is that there is no single generally accepted version available, but, rather, researchers are forced to create their own versions (*ibid.*). The format itself is available but creating a test based on the format is very demanding, as it requires careful consideration as well as piloting before the test could actually be used.

5.1.2.2 Measure used in this study: Vocabulary Knowledge Scale

Another commonly used measure for L2 vocabulary depth is the Vocabulary Knowledge Scale (VKS), which was initially designed by Wesche and Paribakht in 1993. However, the tool is very comprehensively presented in an article that appeared a few years later (Wesche & Paribakht 1996), which is why I will refer mostly to this article here in the description of the tool. The VKS was designed to make up for the lack of appropriate assessment tools for measuring the development of L2 vocabulary knowledge, whether through

different kinds of knowledge that learners can have about particular words, or through stages that characterise how well learners know particular words (Wesche & Paribakht 1996: 13). Moreover, the kind of approach commonly applied in vocabulary size tests that looks at vocabulary knowledge simply as a yes or no condition implies, for example, that each word would only have a single meaning (Wesche & Paribakht 1996: 14). The problem is that this kind of approach does not consider any variation in either aspects or levels of vocabulary knowledge, and even at their best, vocabulary size measures based on this approach can only distinguish between no or only a little knowledge of a single word or some knowledge of it (*ibid.*). As a result, the VKS was originally created for assessing levels of familiarity with particular words, and this kind of approach involving degrees of knowledge is what the authors perceive as depth of vocabulary knowledge (Wesche & Paribakht 1996: 13). The definition and different perspectives on depth of vocabulary knowledge were discussed earlier (see Chapters 2.2 and 2.3), but as a short reminder, in the present study, depth of knowledge is defined as developing degrees of knowledge.

The VKS is a scale that elicits both self-perceived and demonstrated knowledge of given words in written form. It combines both self-report and performance items, ratings varying from utter unfamiliarity, through word recognition and having a vague idea of the meaning, to the learner being able to construct a sentence where the word is correctly used (Wesche & Paribakht 1996: 29). A somewhat similar idea of degrees of word knowledge was proposed earlier by Dale (1965), not including any tool for language assessment, though, but as Wesche and Paribakht (1996: 29) themselves report, they were unaware of Dale's work when designing the VKS. The original VKS involves five self-report categories, which can be seen in Table 2.

Self-report categories	
I	I don't remember having seen this word before.
II	I have seen this word before, but I don't know what it means.
III	I have seen this word before, and I <i>think</i> it means _____ (synonym or translation).
IV	I <i>know</i> this word. It means _____ (synonym or translation).
V	I can use this word in a sentence: _____ (if you do this section, please also do Section IV).

Table 2. VKS elicitation scale self-report categories (Wesche & Paribakht 1996: 30)

In the original VKS, informants' answers are scored on a scale from 1 to 5. In an adaptation made by Joe (1998), the answers are scored on a scale from 1 to 6. After considering both Wesche and Paribakht's (1996) as well as Joe's (1998) approach to VKS scoring, I decided to adapt a model for scoring based on earlier suggestions (Wesche & Paribakht 1996: 30, Paribakht & Wesche 1997: 180, Joe 1998: 363). The scoring categories can be seen in Table 3.

Self-report categories	Possible scores	Meaning of scores
I. 	0	The word is not familiar at all.
II. 	1	The word is familiar but its meaning is not known.
III. 	2	The word is familiar: an association or general meaning is given.
IV. 	3	A correct synonym, paraphrase or translation is given.
V. 	4	The word is used with semantic appropriateness in a sentence.
	5	The word is used with semantic appropriateness and grammatical accuracy in a sentence.

Table 3. VKS scoring categories: Meaning of scores. Adapted from Wesche and Paribakht (1996: 30), Paribakht and Wesche (1997: 180) and Joe (1998: 363)

Scoring self-report categories I and II is very straightforward: marks in these categories result in scores 0 and 1, respectively. In the following self-report categories III, IV and V, on the other hand, the scoring becomes more complex.

Wrong answers in these categories will lead to a score of 1. In category IV, an answer might also result in a score of 2 if the answer given is partly correct: the answer involves an association or a general meaning, but is not exactly a correct synonym, paraphrase or translation. A correct answer will result in a score of 3. In category V, there are many possible scoring options. If the sentence given is totally nonsense, the answer given in Category IV will be scored according to the principles given above. If the informant has written a sentence where the word in question is applied in a semantically appropriate way, a score of 4 will be awarded. Small errors are accepted, for example if an adjective is used as a noun (e.g. *You have good self-confident*) or if a form is mistakenly conjugated (e.g. *eated* for *ate*). The most important factor here is that the sentence demonstrates the informant's knowledge of the word. In order to reach a score of 5, the word has to be used in a both semantically appropriate and grammatically correct way in a sentence. However, small errors in other parts of the sentence are not decisive.

The reliability of the VKS has been examined for example through a test-retest administration, and the results of the process showed that the VKS can elicit reliable responses (Wesche & Paribakht 1997: 180). Moreover, the VKS has been used in many studies in order to elicit knowledge of L2 learners' depth of vocabulary knowledge. In her study, Joe (1998) made some minor alterations to the scale, and, consequently, in Pirilä's (2012) MA thesis, the scale was used following Joe's (1998) modifications. It has to be noted here, though, that despite Pirilä's initial plans, she did not assess EFL learners' depth of vocabulary knowledge; instead, she ended up using the information afforded by the VKS only as a means of verifying some responses that the informants had given in another part of her language test that was a Yes/No format test designed to measure L2 vocabulary size (Pirilä 2012: 53). Culligan (2015) successfully used the VKS in assessing L2 learners' vocabulary depth. In addition, some researchers have used the VKS to trail the acquisition of new words (e.g. Ehsanzadeh 2012). Nevertheless, the primary purpose of the tool is to use it for assessing vocabulary depth (Wesche & Paribakht 1996).

As shown in this subchapter, however, there are difficulties in operationalising

the concept of L2 vocabulary depth so that it could be accurately measured even when large samples are needed. A critical analysis of the VKS has been provided by, for example, Bruton (2009). Bruton (2009: 290, 295) argues that the VKS cannot be used for studying a large number of participants due to the nature of the test that requires hand-scoring; nevertheless, he admits that it seems like the test is applicable at almost any L2 proficiency level, as long as the sample of learners is limited. A shortcoming that the authors of the test themselves also acknowledge (Wesche & Paribakht 1996: 33) is that the test actually excludes multiple meanings. Hence, the VKS can only be used with decontextualized words (Bruton 2009: 292). Furthermore, the VKS does not tap sophisticated knowledge or knowledge of aspects of vocabulary knowledge, such as collocations or derivative forms, nor is it capable of depicting lexical networks (Wesche & Paribakht 1996: 29, 33).

Another problematic point concerns stages I and II of the test, where the learners have to report whether or not they have seen a given word before. As always with self-report, a question of reliability rises: at stage I, it might be, for example, that a learner has indeed seen a given word before but simply does not recall seeing it before. A case like this is not, however, a major problem because it probably does not make much of a difference whether a learner has never seen a given word before or if he or she has simply forgotten ever seeing that word. In a test-retest administration conducted by Wesche and Paribakht (1996: 32), it was indeed revealed that many learners may not remember having seen words that they have in fact seen. This provides evidence for the claim that vocabulary acquisition only begins when learners notice the new word for the first time (*ibid.*). A bigger problem would be consciously reporting wrong answers. In the case of this study, though, it is unlikely for this to occur because it was carefully explained to the learners that their answers would not have any effect on their overall course assessment; the answers of the test would solely be used for the purposes of this study, and, furthermore, all answers would be anonymous. Moreover, the authors also examined the relationship between learners' self-perceived knowledge and demonstrated knowledge by looking at their answers on the self-report categories and their actual scores (Wesche & Paribakht 1996: 31). The self-report and actual

scores were strongly connected, which strongly suggests that self-report data is adequate for this kind of research purpose (*ibid.*). This is a clear advantage of the VKS: it is possible to verify actual knowledge against perceived knowledge (Wesche & Paribakht 1996: 33).

Another shortcoming is related to scoring. No exact criteria are given as to determine correctness of the informants' answers. The researcher is, therefore, responsible for treating each informant's answers equally and being consistent throughout the whole assessment process. In the present study, I considered this matter closely, striving for equality and consistency at all stages of assessment. Moreover, as was explained earlier in this subchapter, I adapted the scoring categories from three different sources so that the scoring would serve the purposes of this study as well as possible. Counting the total score for each participant is also possibly problematic, because two noticeably different combinations of response scores might result in two similar total scores (Bruton 2009: 294). For example, an informant might respond having seen many words before but not remembering their meaning, and arrive at a score similar to another informant who might know the meaning of a few words very accurately but who might not have seen other words used in the test ever before.

As explained in this subchapter, it is undeniably challenging to design a test that would measure depth of vocabulary knowledge accurately, that would be easy to score, and that would be doable in a reasonable amount of time. Currently, there is no vocabulary test that could tap more than only a few aspects of what is considered as vocabulary depth. The WAF focuses more on associations and the links between words (Read 2004: 220–221), whereas the VKS traces developmental stages in informants' word knowledge (Paribakht & Wesche 1997: 179). As explained in subchapters 2.2 and 2.3, vocabulary depth is understood as developing degrees of knowledge in the present study, and that is why the VKS was considered more suitable for the purposes of this study, as it captures certain stages in the developing word knowledge (Wesche & Paribakht 1996: 29).

In order to assess the knowledge of different kinds of words through the VKS, I decided to choose words that represent different frequency levels. As the VST is based on the first fourteen 1000 BNC word lists, I wanted to use something similar for this purpose as well. However, as frequency lists based on the BNC are exclusively British, I decided to use the BNC-COCA 1-25K frequency lists. This is a list that integrates both BNC and COCA (*Corpus of Contemporary American English*) into a comprehensive set of frequency lists. For the lower frequency levels, I checked some texts appearing in the textbooks that are used in the particular upper secondary school where the tests were administered. The book series used in this school is *Open Road* and from this series I randomly selected three books (*Open Road: Course 2, Open Road: Course 5, and Open Road: Course 7*) and from each book I randomly selected a few texts that I ran through VocabProfile, an online program (<http://www.lex Tutor.ca/vp/comp/>) that can be used to assess word frequency levels found in a given text. In VocabProfile, the user can choose which frequency lists are used as the basis for the analysis. I chose BNC-COCA 1-25k for the analysis of the selected texts. I then came up with a list of fourteen words that range from the first 1K level to the 14K level. This is to ensure that each informant has the chance of knowing at least some words in the test and, on the other hand, to ensure that even very skilled informants would have the chance to show that they know some low-frequency words as well. Also, by choosing words that represent quite a range of frequency levels, I hoped to be able to tap differences in the informants' depth of vocabulary knowledge as well. The complete list of words used for this test as well as the test format with its instructions can be found in Appendix 4.

5.1.3 Assessing reading comprehension

In this subchapter, I will review common test formats and principles related to the assessment of reading comprehension. I will also discuss the process of choosing the reading comprehension test to be used in this study. Lastly, I will describe the test shortly.

L2 reading tests usually follow one of the three prevalent test formats: cloze,

short answer questions or multiple-choice (Weir 1997). However, the cloze format seems to provide more information on knowledge of syntax than of reading comprehension (Weir 1997: 40–41). A similar problem occurs with short answer questions as well. Moreover, when the informants have to produce answers themselves either in L1 or in L2, their task becomes clearly more demanding than in the case of multiple-choice questions, for example (*ibid.*). Nevertheless, if an informant produces a correct answer, it is highly unlikely that he or she arrived at the correct answer through guessing. Problems experienced with the multiple-choice test format are similar to those that were already addressed when examining vocabulary size tests: there is a possibility for guessing or for using test-taking strategies. Using test-taking strategies, however, seems to be a universal phenomenon concerning many test types, not just multiple-choice tests alone (Gyllstad, Vilkaite & Schmitt 2015). Multiple-choice tests are widely used due to their efficiency: the informants' task is simple, and they are easy to score and administer. In the present study, too, a multiple-choice test format was chosen.

As for the test used in this study, I decided to opt for reading comprehension sections used in previous matriculation exams. These exams are designed and created every year by the Finnish Matriculation Examination Board (*Ylioppilastutkintolautakunta*). The exams are taken simultaneously every autumn and spring by Finnish upper secondary school students usually in their third and final year of their studies. In English, the learner can choose between an exam based on basic or advanced syllabus. Most upper secondary school students have started learning English in the third grade; thus, they choose the advanced syllabus exam. The reading comprehension section most commonly consists of a number of texts that are accompanied by multiple choice questions in English. There are usually also some texts accompanied by open questions either in English or in Finnish, but there are nonetheless more multiple choice questions than open questions. The tests from previous years along with the correct answers and comments from the Matriculation Examination Board are found online (<http://yle.fi/aihe/artikkeli/2015/12/15/yo-kokeet-englanti>), but in order to use the tests for research purposes, a consent has to be acquired from the Board. I applied for a research permission in order

to be able to use the reading comprehension tests for the purposes of the present study, and the permission was granted.

I looked at a number of reading comprehension sections in previous English matriculation exams, concentrating mainly on the last ten years. I also asked the English teachers of my informants if they had done any of these tests as a reading comprehension exercise, in order to avoid giving the students familiar texts to read. Considering the university level group of learners, it was more difficult to make sure that they would not be familiar with the texts. I decided to choose somewhat recent texts, so that the university students would have already graduated by the time of the publication of the test. Moreover, for them it had already been years since they were in upper secondary school, which is why it is unlikely that they would still remember texts and correct answers from that time. Having pondered on the difficulty level of these tests myself and with other university students of English, it was decided that the reading comprehension test would not be too easy for the university level students, either, as long as the texts and questions were selected carefully. It has to be also remembered that the university-level informants are indeed very advanced users of English, and so it is expected that their reading comprehension skills are on a high level, too. Based on these reasons, I chose two reading comprehension sections, one from autumn 2012 and one from spring 2013, read carefully through all texts, questions and responses, and came up with a reading comprehension test that is comprised of three texts accompanied by respective multiple choice questions. I chose texts that represent different fields, the first text being an extract from a novel, the second an editor's note from *National Geographic*, and the third an article from a magazine. For each text, there are three to five multiple choice questions each with three alternatives in English. The total number of questions in this reading comprehension test is 12, and this is also the maximum score in this test. The test can be found in Appendix 3.

5.2 Participants

In this study, there were two groups of informants that together comprised a group of 58 informants. The first group, that will from now on be called the TOKA group, comprised of 39 Finnish upper secondary school students from southern Finland. Originally, there were 41 students, but due to coming late to the class, two of them could not participate in all parts of the test. The participants were all taking their fifth English course and most of them were second-year students. Some basic information about the informants in the TOKA group is presented below in Table 4.

Basic information on the TOKA group. N=39.		
Gender	Male	13
	Female	26
Age	17 years	30
	18 years	7
	19 years	2
L1	Finnish	32
	Other	7
Language most commonly used in everyday life	Finnish only	30
	Finnish + some other language	9
When started studying English	3 rd grade	30
	Earlier	1
	Later	8

Table 4. Background information about the informants in the TOKA group

As can be seen in Table 4, the majority of informants in the TOKA group were girls. Most participants in this group were 17 years old. The majority also reported Finnish as their L1; seven participants reported some other language as their L1. When asked what language the participants most commonly used in their everyday life, all participants reported using only Finnish (30 participants) or Finnish and some other language (9 participants). Due to this

and to the fact that they all attended an upper secondary school where Finnish is the primary language of instruction, I decided not to exclude participants with a different L1 from this study. The majority had started studying English in the third grade. One informant reported having started earlier, in the second grade, whereas eight informants reported having started later: six informants in the fourth grade, one informant in the sixth grade and one informant in the seventh grade. The background information form also included a question about possible longer stays in an English-speaking area or country. One informant had been born in the UK, but had nevertheless lived her whole life in Finland and reported Finnish both as her L1 and as the only language used in everyday life. Other informants did not report any longer stays in an English-speaking environment.

The second group, that will from now on be called the YLI group, was a group of 19 very advanced learners of English. All of them were English majors completing their teacher training at the University of Turku at the time of the study. Basic information about the participants in this group is presented in Table 5.

Basic information on the YLI group. N=19.		
Gender	Male	2
	Female	17
Age	23 years	5
	24 years	3
	25 years	1
	26 years	2
	Younger	2
	Older	6
L1	Finnish	19
	Other	0
Language most commonly used in everyday life	Finnish only	17
	Finnish + some other language	2
When started studying English	3 rd grade	14
	Earlier	2
	Later	3
First year at the English department at the University of Turku	2012	8
	Earlier	9
	Later	2

Table 5. Background information about the informants in the YLI group

As in the TOKA group, the majority of the participants in the YLI group were females. The age range in this group was more versatile: the youngest participant was 22, whereas the oldest participant was 43 years old. The mean age for the YLI group was 27. All of the informants reported Finnish as their L1; however, two of them reported using most commonly both Finnish and some other language in their everyday life. Similarly to the TOKA group, the majority in the YLI group had started studying English in the third grade. Two informants reported having started already in the second grade, whereas three reported having started later: two in the fifth grade and one in the seventh grade. The year of starting studying at the English department at the University of Turku varied considerably. The majority had started their studies in 2012. In

total, nine informants had started their studies earlier; most of them in either 2010 or 2011, but three informants even earlier. Two informants started their studies at the English department later, in 2013. Nine informants in the YLI group reported one or several longer stays in an English-speaking area or country for different reasons. The length of the stays ranged from working for a couple of months to completing a three-year Bachelor's degree abroad. Ten of them did not report any longer stays in such area or country. As all the informants in the YLI group were English majors who were already on a very advanced level in their studies, the length or number of stays in an English-speaking environment was not considered a decisive factor in the present study.

Apparently, the selection of the texts for the reading comprehension test was successful: only one of the university students graduated from upper secondary school in autumn 2012 or in spring 2013, but this participant did not report being familiar with the text passages. For other informants in the YLI group, it is very unlikely that they would have been familiar with the reading comprehension tests.

5.3 Test administration

The tests used in the present study were thoroughly introduced in the previous subchapters. These tests can be found in their full form in the section of Appendices, except for the VST which can be found online (http://www.lex Tutor.ca/tests/levels/recognition/1_14k/). In total, the test comprised of a background information form (Appendices 1 and 2) and three other parts, a reading comprehension test (Appendix 3), a vocabulary size test (online) and a vocabulary depth test (Appendix 4).

Firstly, each informant filled in a background information form. This form included questions concerning for example the gender, age and L1 of the informant. There were also some questions related to their knowledge of English, such as when they had started studying English or if they had, for example, lived or studied in an English-speaking country. As for the YLI group,

there were also some questions related to their studies at university. The answers on this form provided basic information about the participants, which was already presented in the previous subchapter 5.2. After the background information form had been completed, the actual test began. The informants did not know what parts of language skills were assessed; they were merely informed that the study would focus on EFL learners and their language skills.

Some general guidelines for taking part in this study were told to the informants in the beginning of each testing session. As the test had several parts, the participants were required to write their name on each paper. It was carefully explained to them that this was only done so that it would be possible to combine their answers on different tests, and that they would still be treated as anonymous informants in the study. It was also clarified to them that the results of these tests would have no effect on their English course grades and that their answers would only be used for the purposes of this study. It was explained to them that for each part of the test, they would have twenty minutes. Instructions for each part were explained in the beginning of each part. After completing a test, they could return the paper but they could not start the following test before everyone was finished. Also, nobody was allowed to leave the classroom during the tests. The tests were administered in one session so that each testing session then took about 60 to 70 minutes, depending on the group.

5.4 Statistical methods

The data were analysed using the *IBM SPSS Statistics 23.0.0.2* (SPSS) computer program. As handbooks, I used Muijs (2004) and Larson-Hall (2016), and all statistical methods used in the present study are more thoroughly explained in these manuals, which is why the reader is advised to consult these handbooks when in need of deeper knowledge of the methods. In order to examine the relationships between different test results and between the performances of the two study groups, I used the *correlation coefficient* (r). It has to be acknowledged, though, that correlation coefficients do not reveal any causal relationships; instead, calculating the correlation coefficient will tell how

strongly two variables are associated with each other (Muijs 2004: 142–143). Normally, a parametric test, *Pearson's r*, is used for this purpose in conjunction with continuous variables. However, the results in one of the tests were so uniform that the scores actually behaved more like a categorical variable rather than a continuous one, and, in addition, the scores were not normally distributed. This is why a non-parametric test, *Spearman's rho*, was chosen instead, as it does not presuppose normal distribution nor all variables to be continuous. When describing the strength of the relationship between two variables, I followed Muijs' guidelines (2004: 145): $<0.\pm 1$ being *weak*, $<0.\pm 3$ *modest*, $<0.\pm 5$ *moderate*, $<0.\pm 8$ *strong*, and $>0.\pm 8$ *very strong*.

As a means of describing the variance in reading comprehension scores explained by vocabulary breadth and depth, I used the *multiple linear regression*. This method of analysis makes it possible to examine the relationship between one dependent variable and one or more independent variables (Muijs 2004: 160). Multiple linear regression also reveals the amount of variance in the dependent variable explained by one or more independent variables, and this is expressed in R square (R^2) values that vary from 0 to 1 (Muijs 2004: 163, 165).

When reporting *probability values* (p), I have mostly used 0.05 as the cut-off point to decide whether a result is statistically significant or not, which is a figure commonly used in L2 research (Larson-Hall 2016: 64). When a result was particularly significant statistically, I have reported 0.01 or 0.001 cut-off points. This is encouraged by, for example, Muijs (2004: 78).

6 Results

This study set out to answer the following research questions:

- 1) How do L2 vocabulary size, L2 vocabulary depth, and reading comprehension relate to each other?
- 2) What is the role of L2 vocabulary depth in reading comprehension? More accurately, what kind of a contribution can L2 vocabulary depth possibly bring into explaining and predicting L2 reading comprehension skills, in addition to the contribution provided by L2 vocabulary breadth alone?
- 3) Regarding the first two research questions presented above, what kind of differences are there between EFL learners at upper secondary school and English majors at the University of Turku, if any?

In this chapter, the results of the present study will be presented. I will first shortly present the results of each test separately, following the order in which the tests were introduced in Chapter 5.1: the VST, the VKS, and the reading comprehension test. As the aim of the present study was to look at the relationships between L2 vocabulary breadth, depth and reading comprehension, more attention will be paid to examining these relationships in the following subchapters. Attention will also be paid to discovering the role of L2 vocabulary depth in reading comprehension. In all subchapters of this section, I will treat the study groups both separately and together. Thus, any differences found between the TOKA and YLI groups can be observed throughout the chapter.

6.1 Test scores

In this subchapter, the results of each test will be presented. The most central figures for each test will be given (minimum score, maximum score, standard deviation, mean, median): both for each group separately and for two groups as a total. Standard deviation refers to “the extent to which the values in a distribution cluster around the mean” (Muijs 2004: 107), and median is “the middle category of a distribution” (Muijs 2004: 100). In order to verify if the

means of the two groups differ in a statistically significant way, Mann-Whitney U-tests were conducted for each test. T-tests are often used for this purpose, but because all test results for all groups were not normally distributed, I decided to use the Mann-Whitney U-test instead across all tests for clarity and reader-friendliness. Both the independent samples t-test and Mann-Whitney U-test essentially measure the same phenomenon, but the latter is often used with smaller samples and when the results are not normally distributed (Larson-Hall 2016: 74, 478).

6.1.1 Vocabulary Size Test

The informants' vocabulary size was measured with the VST (see Chapter 5.1.1.3). As explained in Chapter 5.1.1.3, the maximum score for the VST was 100 points. In order to arrive at an estimation of a learner's vocabulary size, the score had to be then multiplied by 100. Table 6 below presents the central findings of informants' estimated vocabulary sizes; for each informant group separately and also as a total. All vocabulary measures here were rounded to integers.

	Min	Max	St. Dev.	Mean	Median
TOKA (n=39)	1500	8500	1761	5715	5600
YLI (n=19)	7300	9800	637	8816	8900
Total (n=58)	1500	9800	2086	6731	7300

Table 6. Scores on the Vocabulary Size Test. Note: Min = The least successful score; Max = The most successful score; St. Dev. = Standard Deviation

As can be seen in Table 6, the least successful informant in the TOKA group had a vocabulary of 1500 words, whereas the most successful informant in this group had a vocabulary of 8500 words. Hence, the standard deviation in the TOKA group was large: 1761. The mean vocabulary size of the TOKA informants was 5715 words.

As a whole, the YLI group performed better in the VST than the TOKA group. In the YLI group, the least successful participant had a vocabulary of 7300

words, and the most successful participant a vocabulary of 9800 words. The standard deviation in the YLI group was considerably lower than in the TOKA group: only 637 compared to 1761 reported in the TOKA group. The mean vocabulary size of the YLI informants was 8816 words. A Mann-Whitney U-test was conducted in order to further verify the statistical significance of the differences in results of the two study groups ($U = 23.0, p < 0.01$).

6.1.2 Vocabulary Knowledge Scale

The informants' depth of vocabulary knowledge was assessed with the VKS (see Chapter 5.1.2.2). The maximum score for this test was 70 points. Similarly to the previous subchapter, the measures were rounded to integers (except for standard deviation, which is presented to one decimal place), and the most important findings are presented in Table 7.

	Min	Max	St. Dev.	Mean	Median
TOKA (n=39)	16	58	11.3	36	36
YLI (n=19)	52	66	3.5	60	60
Total (n=58)	16	66	14.7	44	45

Table 7. Scores on the Vocabulary Knowledge Scale. Note: Min = The least successful score; Max = The most successful score; St. Dev. = Standard Deviation

In this test, the mean score for the TOKA group was 36 points. The least successful participant scored 16 points, whereas the most successful participant scored 58 points. Similarly to the scores in the VST, the most successful participants in the TOKA group were close to the mean score of the more advanced YLI group. Again, the standard deviation of the TOKA group's scores was large: 11.3. In the YLI group, on the other hand, the differences between the participants were less dramatic. The mean score was 60 points and the standard deviation only 3.5. In this group, the least successful informant scored 52 points compared to the most successful informant's 66 points. A Mann-Whitney U-test indicated that the differences in the means between the two groups were statistically significant ($U = 11.5, p < 0.01$).

6.1.3 Reading comprehension test

The reading comprehension test used in this study comprised of three different texts accompanied by multiple-choice questions (see Chapter 5.1.3). The maximum score in this section was 12 points. The most central findings of this test can be seen in Table 8. Again, the measures were rounded to integers for clarity (except for standard deviation that was rounded to one decimal place).

	Min	Max	St. Dev.	Mean	Median
TOKA (n=39)	0	11	2.8	6	7
YLI (n=19)	9	12	1.1	11	11
Total (n=58)	0	12	3.2	8	8

Table 8. Scores on the reading comprehension test. Note: Min = The least successful score; Max = The most successful score; St. Dev. = Standard Deviation

In the TOKA group, the differences between the learners were again noteworthy. The least successful informant did not manage to provide any correct answers to the multiple-choice questions, whereas the most successful informant scored 11 out of 12 points. The standard deviation was 2.8. The mean score for this section was 6 points in the TOKA group. Similarly to the scores in the other two tests reported in the previous two subchapters, the differences between the participants' results in the YLI group were less notable. In this group, the lowest score was 9 points and the highest 12 points. The standard deviation was 1.1. The mean score of the informants in the YLI group was 11 points. As with the other two test results presented above, a Mann-Whitney U-test was conducted, and the differences in the means of the TOKA and YLI groups were statistically significant ($U = 40.0, p < 0.01$).

6.2 Examination of the relationships between the test results

In the previous subchapter, the test results were reviewed separately. In this subchapter, I will move on to looking at the relationships between the test results. The relationships were studied through correlations by using the

Spearman's rho. The correlation coefficients were rounded to two decimal places.

The correlations in the TOKA group were rather high and positive across all test results (see Table 9). Vocabulary size and the scores on the VKS correlated very strongly and positively ($r = 0.89$, $n = 39$, $p < 0.001$). Vocabulary size also correlated strongly with the reading comprehension test results ($r = 0.65$, $n = 39$, $p < 0.001$). The correlation between the scores on the VKS and on the reading comprehension test were also strongly related ($r = 0.59$, $n = 39$, $p < 0.001$).

	Vocabulary size	VKS	RC
Vocabulary size		0.89	0.65
VKS	0.89		0.59
RC	0.65	0.59	

Table 9. Spearman's rho correlations in the TOKA group. Note: RC = reading comprehension

In the YLI group, on the other hand, the correlations were not as high as in the TOKA group (see Table 10). Similarly to the TOKA group, vocabulary size and the scores on the VKS were strongly and positively correlated ($r = 0.72$, $n = 19$, $p = 0.001$) in the YLI group as well. The other two correlations were clearly weaker. There was a moderate positive correlation between vocabulary size and reading comprehension scores ($r = 0.31$, $n = 19$, $p = 0.205$). A modest positive correlation was also found between the scores on the VKS and on the reading comprehension test ($r = 0.29$, $n = 19$, $p = 0.224$). As can be seen from the p -values, the correlations between vocabulary size and reading comprehension as well as between vocabulary depth and reading comprehension scores were not statistically significant ($p > 0.05$). This may be at least partially explained by the small sample size ($n = 19$).

	Vocabulary size	VKS	RC
Vocabulary size		0.72	0.31
VKS	0.72		0.29
RC	0.31	0.29	

Table 10. Spearman’s rho correlations in the YLI group. Note: RC = reading comprehension

In Table 11 below, the correlations for both groups together are presented. The correlations were strong and positive across all variables. A very strong, positive correlation was found between vocabulary size and the scores on the VKS ($r = 0.95$, $n = 58$, $p < 0.001$). Vocabulary size was also very strongly and positively related to the scores on the reading comprehension test ($r = 0.83$, $n = 58$, $p < 0.001$). Another very strong, positive correlation was found between the scores on the VKS and on the reading comprehension test ($r = 0.95$, $n = 58$, $p < 0.001$).

	Vocabulary size	VKS	RC
Vocabulary size		0.95	0.83
VKS	0.95		0.95
RC	0.83	0.95	

Table 11. Spearman’s rho correlations in both groups in total. Note: RC = reading comprehension

Based on the findings presented in this subchapter, the differences between the results of the two study groups are evident. These differences will be further discussed in Chapter 7. Next, I will move on to presenting the results related to the effect of vocabulary size and depth in reading comprehension test scores.

6.3 Examination of the impact of vocabulary size and depth in reading comprehension test results

As shown in Chapter 4.2, L2 learners’ vocabulary size strongly affects L2 reading comprehension. The role of L2 learners’ depth of vocabulary knowledge in reading comprehension, on the other hand, has not been in the

focus of many studies. One of the aims of this study was to examine the role of L2 vocabulary depth in reading comprehension more thoroughly and to try to find out if vocabulary depth could bring something more into the prediction and explanation of reading comprehension skills, in addition to the information afforded by vocabulary size alone. As for the method, multiple linear regression was chosen in order to answer this research question. Multiple linear regression is not usually applied when the independent variables (in this case, vocabulary size and the scores on the VKS) are rather strongly correlated with each other and when all variables are not normally distributed (Muijs 2004: 176, Larson-Hall 2016: 229). It is rather questionable, though, what counts as a strong connection between two independent variables, as this seems to vary from one researcher to another (Larson-Hall 2016: 244). In addition, the use of multiple linear regression was further encouraged by existing research (e.g. Qian 1999, 2002, Qian & Schedl 2004, Rashidi & Khosravi 2010) where this method has successfully been used together with independent variables that correlate with each other. Contrary to these studies, however, I decided to report adjusted R^2 values instead of mere R^2 values. R^2 refers to the amount of variance in a dependent variable (in this case, reading comprehension) that is explained by independent variables (vocabulary size and depth) (Muijs 2004: 165). An adjusted R^2 is a correction to R^2 that “takes into account that we are looking at a sample rather than at the population” (*ibid.*). Adjusted R^2 values are often slightly lower than mere R^2 values (Larson-Hall 2016: 246), but as explained above, they represent the population better than R^2 values. The adjusted R^2 values were rounded to two decimal places.

In the TOKA group, vocabulary size alone explained about 38% of the variance in reading comprehension scores (adjusted $R^2 = 0.38$, $F(1, 37) = 24.68$, $p < 0.001$). When looking at the explanation afforded by vocabulary size and the scores on the VKS together, the percentage was 39% (adjusted $R^2 = 0.39$, $F(2, 36) = 12.88$, $p < 0.001$). Adjusted R^2 change was 0.01, which suggests that vocabulary depth added 1 percentage point of explained variance in reading comprehension scores. The findings are summarised in Table 12. When looking at the explained variance provided by vocabulary depth alone, the percentage was about 38% (adjusted $R^2 = 0.38$, $F(1, 37) = 23.77$, $p < 0.001$).

Step	Variable	Adjusted R ²	R ² change	p-value
1	Vocabulary size	0.38		<0.001
2	Scores on the VKS	0.39	0.01	<0.001

Table 12. Multiple regression results of the TOKA group with vocabulary size and the scores on the VKS as independent variables

The multiple regression analysis results in the YLI group were rather different from the TOKA group. Vocabulary size explained only about 19% of the variance in reading comprehension test scores (adjusted R² = 0.19, F(1, 17) = 5.32, p<0.05). Furthermore, when scores on the VKS were added to the equation, it looks as if the two independent variables together are not very good at predicting or explaining reading comprehension test scores: the percentage was only 15% (adjusted R² = 0.15, F(2, 16) = 2.62, p>0.05). I also ran a regression analysis for the scores on the VKS alone, and the predictive power of vocabulary depth alone was less than a percent (adjusted R² = 0.03, F(1, 17) = 1.06, p<0.05). As can be seen in the p-values reported, only the variance explained by vocabulary size alone is statistically significant. This may be partly explained by the small sample involved so the results may not represent the whole population very well. Nevertheless, the results suggest, firstly, that neither vocabulary size nor vocabulary depth are very good at predicting or explaining reading comprehension proficiency, and secondly, that scores on the VKS do not actually add anything into the prediction of reading comprehension in the YLI group. A summary of the findings is found in Table 13 below.

Step	Variable	Adjusted R ²	R ² change	p-value
1	Vocabulary size	0.19		<0.05
2	Scores on the VKS	0.15	-0.04	>0.05

Table 13. Multiple regression results of the YLI group with vocabulary size and the scores on the VKS as independent variables

Lastly, I looked at the two groups as a uniform group to see what the effects of vocabulary size and vocabulary depth are like in reading comprehension in general. Vocabulary size alone explained about 64% of the variance in reading comprehension test results (adjusted R² = 0.64, F(1, 56) = 102.17, p<0.001).

When both vocabulary size and depth of vocabulary knowledge scores were added to the equation, they together explained about 66% of the variance in reading comprehension (adjusted $R^2 = 0.66$, $F(2, 55) = 55.07$, $p < 0.001$). A modest addition, about 2 percentage points, was afforded by scores on the VKS. When looking at the percentage of explained variance in reading comprehension afforded by vocabulary depth alone, the figure was 64% (adjusted $R^2 = 0.64$, $F(1, 56) = 103.95$, $p < 0.001$). As with all reported explained variance figures presented in this subchapter, the percentages and changes in R^2 would have been slightly higher had I decided to report unadjusted R^2 values. For reliability, though, I decided to stick with reporting adjusted R^2 values as they represent the whole population better than mere R^2 values. Table 14 below sums up the findings

Step	Variable	Adjusted R^2	R^2 change	p-value
1	Vocabulary size	0.64		<0.001
2	Scores on the VKS	0.66	0.02	<0.001

Table 14. Multiple regression results of both groups together with vocabulary size and the scores on the VKS as independent variables

As with the correlation coefficients reported in the previous subchapter, the differences found between the two study groups are again obvious. There are many possible explanations for this; hence, the findings will be further discussed in the following chapter. In addition, despite the small number of previous studies conducted in this field, the results of the present study will be compared with existing research findings in the next chapter.

7 Discussion

The results of the present study were presented in the previous chapter, and in this chapter, I will move on to discuss the findings. In the following subchapter 7.1, I will discuss the test results and compare especially the results of the vocabulary size test to previous research conducted in Finland. Next, in subchapter 7.2, I will shed some more light on the correlations found between the size and depth of vocabulary and reading comprehension. The results of my study will also be compared to existing research findings. In subchapter 7.3, the role of vocabulary depth in reading comprehension from the perspective of the results found in the present study will be discussed in more detail. As in the previous subchapters, the results obtained in the present study will be compared to earlier studies as well. Finally, in subchapter 7.4, the results of the TOKA and YLI groups will be compared and discussed.

7.1 Discussing the test results

In this section, the test results will be discussed. Considering the lack of comparable research in this area, however, the focus will mostly be on discussing vocabulary size and comparing the results obtained in this study with earlier studies conducted in Finland.

In Chapter 4.2, I reviewed a few studies concerning Finnish EFL learners' vocabulary size and now I will discuss how the results of the earlier studies relate to the results of the present study. It has to be firstly noted, though, that the previous studies are not directly comparable with each other nor with the present study due to the use of different vocabulary test formats and different study groups; nevertheless, it is interesting to see what kind of results have been found in different studies and how testing methods, for example, can affect the results. Jaatinen and Mankkinen (1993) studied university students of English and the results of their study suggest that an average English major has a vocabulary size of about 18 000 lexemes: 17 100 on the Bachelor level and 19 500 on the Master level. As a test method, they used a Yes/No test that was built on dictionary entries instead of frequency levels. This might have

influenced the seemingly high results; when compiling a test based on dictionary entries, the test measures the knowledge of lexemes, and not of word families. In addition, as was discussed in Chapter 5.1.1.2, Yes/No test formats do not elicit information on any perceived word knowledge, and guessing plays a role in this test format as well. A similar Yes/No format was used in Piriälä's (2012) study as well, and the results of her study suggest that Finnish university students of English have even larger vocabularies than what was recorded in Jaatinen and Mankkinen's (1993) study. On average, a university student of English knew 22 000 lexemes: first-year students knew 21 100 and Master-level students 23 200 lexemes. No pseudowords were included in Piriälä's test format and the test items were sampled from a dictionary.

In the present study, the YLI group seemed to have somewhat smaller vocabularies than is suggested by the studies of Jaatinen and Mankkinen (1993) and Piriälä (2012). On average, a major student of English knew 8816 words. The student with the smallest vocabulary size knew 7300 words whereas the student with the largest vocabulary size knew 9800 words. In the case of the present study, however, it is noteworthy to remark that the VST only tested the knowledge of the 10 000 most frequent word families of English. In order to tap the whole scope of the learners' vocabulary breadth, another test involving further frequency levels as well, ought to be used. It is likely that the results obtained would have been somewhat higher had the test been longer, i.e. tested the knowledge of word families on even lower frequency levels. Moreover, the definition of a word in the present study differed from that applied in the previous studies. Lexeme is a somewhat narrower definition (see Chapter 4.2), and it is likely that the figures obtained in the earlier studies would have been smaller if the concept of a word family had been applied in these studies as well.

In her study, Ala-Akkala (2010) studied Finnish upper secondary school students and their vocabulary size of English. She used a type of Yes/No test that was based on word families and included pseudowords (that were nonetheless ignored in scoring) as well. Two groups of students were involved

in the study: the first group was on their first year and the second on their third and final year of upper secondary school. According to the results of her study, the first-year students knew 3400 word families, whereas the third-year students knew 4000 word families. In the present study, the TOKA group consisted of upper secondary school students most of whom were second-year students. Thus, it is very interesting to compare Ala-Akkala's results to the present study. According to the test scores obtained in this study, the TOKA group had a vocabulary size of 5715 word families. The informant with the smallest vocabulary knew 1500 words, whereas the informant with the largest vocabulary knew 8500 words. In Ala-Akkala's study, when considering both groups of informants, the smallest vocabulary size reported was less than 1500 word families and the largest almost 7000 word families. Even if the average vocabulary sizes reported by Ala-Akkala and myself seem somewhat different, the differences are not dramatically opposed. Moreover, the results of both studies seem to suggest that the differences between individual upper secondary school students are large.

A major difference between the earlier studies described here and the present study lies in the test format. The Yes/No test format does not require the informant to actually show any knowledge of word meaning; the method is more like self-assessing one's knowledge. In the multiple-choice format that was applied in the present study, however, the informant is required to demonstrate his or her knowledge. On the other hand, the multiple-choice format mostly requires the informant to know one particular meaning of the test item in question whereas in the Yes/No format, the informant can report any meaning out of all the meanings of a particular word. Thus, it is likely that the use of multiple-choice formats results in somewhat smaller estimates of learners' vocabulary size than the use of tests based on the Yes/No format. In addition, as mentioned earlier, with the VST applied in this study, the maximum vocabulary size that the test could report was 10 000 word families. A longer test or a test compiled differently (for example, five test items instead of ten to represent one frequency level) might be useful in order to better estimate the vocabulary size of advanced learners of English especially. Even if the authors of the VST, Nation and Beglar (2007: 11), accept shortening the test, they

suggest nonetheless that the best is to strive for a test where the last frequency levels to be tested are a few levels beyond the learners' expected vocabulary size. This may explain why the results of the YLI group differ so drastically from other studies conducted with Finnish university students of English (Jaatinen & Mankkinen 1993, Pirilä 2012) and why the results of the TOKA group seem rather similar to the results reported by Ala-Akkala (2010). It may simply be that the VST functioned better with the TOKA group than with the YLI group, and that more frequency levels ought to have been tested with the YLI group.

The results of the other two tests, the VKS and the reading comprehension test, will not be discussed here due to the lack of comparable previous studies. However, they will be discussed in the following subchapters in conjunction with other test scores. As the focus of this study was primarily on describing the relationships found between vocabulary size, depth and reading comprehension, on discovering the role of vocabulary depth in reading comprehension and on presenting any differences found between the two study groups, the findings related to these issues will be further discussed in the next subsections.

7.2 Relationships between vocabulary breadth, depth and reading comprehension: comparison with earlier studies

In the present study, positive correlations were found between L2 vocabulary size, depth and reading comprehension (see Chapter 6.2). The correlations were especially high in the TOKA group and in the two groups together. In the YLI group, on the other hand, the correlations were somewhat lower. In this subchapter, I will compare the results of my study to earlier studies and discuss possible reasons for any differences found. With the intention of providing the reader with a clear discussion of these relationships, I will examine each relationship one by one. In order to avoid unnecessary repetition of figures, only the correlation coefficients (r) will be mentioned in conjunction with the results of the present study. Other associated figures related to correlations can be found in Chapter 6.

Many researchers have pointed out that the relationship between the size and depth of vocabulary knowledge is a close one (e.g. Qian 1999). The findings of the present study support this view. The correlation between vocabulary size and the scores on the VKS was high and positive in all groups: 0.89 in the TOKA group, 0.72 in the YLI group, and 0.95 in the two groups together. Similar results have been reported in earlier studies as well. Rashidi and Khosravi (2010) studied 38 Iranian university-level EFL learners whose vocabulary size was at least 3000 word families. They measured vocabulary size with the VLT (see Chapter 5.1.1.1) and vocabulary depth with the WAF (see Chapter 5.1.2.1). Thus, depth was understood mainly as aspects of collocation and meaning. A very strong, positive correlation was found between vocabulary size and depth ($r = 0.81, n = 38, p < 0.01$). Similar results were found in Li's (2015) study where the test formats used were similar to those used in Rashidi and Khosravi's (2010) study: the VLT measured vocabulary size and the WAF vocabulary depth. Li studied Chinese EFL learners, all of whom had been learning English for at least six years. The correlation between vocabulary size and depth was strong and positive ($r = 0.66, n = 30, p < 0.01$).

Positive correlations between vocabulary size and vocabulary depth have also been reported in Qian's studies (e.g. Qian 1999, 2002). In the first study, Qian (1999) studied 74 Korean and Chinese EFL learners. Again, the VLT and the WAF were used as measures of vocabulary breadth and depth, but in addition, Qian compiled another test for vocabulary depth that assessed the learners' morphological knowledge. The correlations found between vocabulary size and both measures of vocabulary depth were high and positive: 0.82 ($n = 74, p < 0.05$) between vocabulary size and the WAF, and 0.69 ($n = 74, p < 0.05$) between vocabulary size and the morphological knowledge. Similar correlations were reported in the second study of Qian (2002). The number of informants was higher in this study, 217, and the participants came from various L1 backgrounds; yet all were EFL learners. Again, Qian used the VLT and the WAF to measure vocabulary breadth and depth, respectively. On top of these measures, he used a part of TOEFL vocabulary test to measure the knowledge of synonyms. This was done in order to gain more information on the informants' vocabulary depth. The results showed clearly the positive

correlation that there is between vocabulary breadth and depth: 0.70 ($n = 217$, $p < 0.01$) between vocabulary size and the WAF, and 0.80 ($n = 217$, $p < 0.01$) between vocabulary size and the synonym test.

From the viewpoint of the present study, it is interesting to look at the correlations found between different measures of vocabulary knowledge and scores on reading comprehension tests. As reading comprehension measures used in the previous studies mentioned in the paragraphs above differ from one study to another, I will not describe them in detail here. In addition, as the studies were already presented in Chapter 4.3, and shortly also in the previous paragraphs of this subchapter, the studies will not be thoroughly described in the following paragraphs in order to avoid unnecessary repetition. I will first compare the correlations found between vocabulary size and reading comprehension, and then move on to the relationship between vocabulary depth and reading comprehension.

In this study, the correlation between vocabulary size and reading comprehension was positive in all groups: 0.65 in the TOKA group, 0.31 in the YLI group, and 0.83 in total. The results of the present study are similar to those obtained in earlier studies. Rashidi and Khosravi (2010) reported a correlation of 0.75 ($n = 38$, $p < 0.01$) and Li (2015) a correlation of 0.60 ($n = 30$, $p < 0.01$) between vocabulary size and reading comprehension. Resembling findings were also reported in Qian's studies (1999, 2002). The correlation was 0.78 ($n = 74$, $p < 0.05$) in the first study (Qian 1999), and 0.74 ($n = 217$, $p < 0.01$) in the second study (Qian 2002).

Vocabulary depth and reading comprehension were also positively correlated in the present study: the correlation was 0.59 in the TOKA group, 0.29 in the YLI group, and 0.95 in the two groups in total. This is in line with the results reported in previous research. Rashidi and Khosravi (2010) reported a very strong and positive correlation between vocabulary depth and reading comprehension ($r = 0.87$, $n = 38$, $p < 0.01$) whereas Li (2015) reported slightly lower, yet positive correlation between the two variables ($r = 0.61$, $n = 30$, $p < 0.01$). In both of Qian's studies (1999, 2002), two measures were used to

assess the learners' depth of vocabulary knowledge. Nonetheless, the correlations still resemble the ones already mentioned here. In the first study (Qian 1999), a positive correlation was found between the WAF and reading comprehension ($r = 0.82$, $n = 74$, $p < 0.05$) as well as between the morphological knowledge test and reading comprehension ($r = 0.64$, $n = 74$, $p < 0.05$). Similarly, in the second study (Qian 2002), a positive relationship was reported between the WAF and reading comprehension ($r = 0.77$, $n = 217$, $p < 0.01$), and between the synonym test and reading comprehension ($r = 0.73$, $n = 217$, $p < 0.01$).

In Qian and Schedl's (2004) study, in addition to the development of a new tool for measuring vocabulary depth, the relationship between vocabulary depth and reading comprehension was studied as well. Vocabulary size was not involved in the study. The participants were 207 international students from various L1 backgrounds; most of them university-level students or beyond. The tests used differed a little from those used in the previous studies reported here. A TOEFL vocabulary test was used as quality control so that the new measure could be tested thoroughly. The new test was close to the WAF in design and it tested the knowledge of word meaning, polysemy, synonymy, and collocations. For assessing reading comprehension, a TOEFL reading test was used. Similarly to the present study, a positive correlation was found between the new vocabulary depth test and the reading comprehension test ($r = 0.74$, $n = 207$, $p < 0.01$).

It is clear that the comparisons made in this subchapter strongly suggest that vocabulary breadth, depth and reading comprehension are profoundly connected with each other. Moreover, the results of the present study are in line with previous research findings. This is especially interesting from the viewpoint of Finnish EFL learners. To the best of my knowledge, this kind of research has not been conducted with Finnish EFL learners. Finnish is not a cognate of English, nor of any of the languages that were any of the informants' L1 mentioned in the previous studies of the field (e.g. Qian 1999, 2002, Qian & Schedl 2004, Nashidi & Khosravi 2010, Li 2015). The results of the present as well as the earlier studies support the views introduced in subchapters 4.1

and 4.2: L2 reading is fundamentally a language problem that cannot be explained only by the learner's L1 or L1-related skills, and L2 vocabulary knowledge strongly affects reading comprehension.

It has to be remembered here, though, that the measures used in the studies were varied, so the findings may not be directly comparable with each other. In addition, given the close relationship between vocabulary breadth and depth, it is likely that there is possible overlap in the measures used to assess both size and depth of vocabulary knowledge. In order to further investigate the relationships between the two variables and reading comprehension, it might be useful to include more tests of different dimensions of vocabulary knowledge. This is usually rather difficult to do in practice due to the fact that the testing situation might then become overly long and demanding. A fundamental difference between the earlier studies and the present study lies in the definition of depth of vocabulary; in all earlier studies mentioned, vocabulary depth was seen as dimensions whereas in the present study, it was understood as developing stages of knowledge. However, as far as I am aware of, there are no studies where depth of vocabulary knowledge would have been studied as degrees of knowledge, and where vocabulary size and reading comprehension would have been involved as well.

In addition, the groups studied were varied as well. In many of the previous studies, only the educational background and age of the informants were given, with the sole mention that they were EFL learners. There was rarely any indication on the competence level or information on how long the participants had been learning English. Of course, the scores obtained on different tests offer some suggestions on the competence level of the informants; nevertheless, the studies are not directly comparable. Despite the limitations concerning the comparability of the studies reported here, the results of the present study are still similar to those of earlier studies.

7.3 Role of vocabulary depth in reading comprehension: comparison with earlier studies

In this subchapter, the results of the present study will be further discussed in relation to the significance of vocabulary depth in reading comprehension specifically. The findings will be discussed and compared to previous studies. As the number of studies in this field is limited (see Chapter 4.3), the comparison is restricted to the studies of Qian (1999, 2002), Qian and Schedl (2004), and Rashidi and Khosravi (2010). All of these were already mentioned and shortly described in Chapters 4.3 and 7.2, which is why the reader is advised to consult these sections when in need of more information. Similarly to the previous subchapter and in order to avoid unnecessary repetition, only the adjusted R^2 values will be presented when discussing the results of the present study. For other related figures, the reader should turn to Chapter 6.

The results of the present study suggest that vocabulary depth can indeed add another significant variable in the prediction and explanation of scores in a reading comprehension test. In the TOKA group, vocabulary breadth alone explained 38% of the variance in reading comprehension. Together with vocabulary depth, the two variables explained 39% of the variance. When examining the TOKA and YLI groups together, the corresponding figures were 64% and 66%, respectively. These results suggest that vocabulary depth added 1 percentage point and 2 percentage points of explained variance in reading comprehension scores when considering the TOKA group and the TOKA and YLI groups together. Similarly to the figures of variance explained by vocabulary size alone, vocabulary depth alone could explain 38% and 64% of variance in reading comprehension scores in the TOKA group and in the two groups in total, respectively. It seems that vocabulary depth does not add much to the explanation and prediction of L2 reading, in addition to that afforded by vocabulary size alone. Nevertheless, the results suggest that both vocabulary size and depth could be used separately as equivalent predictors of reading proficiency. However, when looking at the results of the YLI group, the results were somewhat different. In the YLI group, vocabulary size alone explained 19% of the variance in reading comprehension. Together with vocabulary

depth, the two variables explained only 15% of the variance. However, this number was not statistically significant ($p>0.05$). Within the scope of the present study, nevertheless, the numbers suggest that vocabulary depth does not necessarily provide any additional information, which would not be provided by vocabulary size alone, in the prediction of reading comprehension skills of the YLI group. Moreover, when looking at the predictive and explanatory power of vocabulary depth alone, it could only explain less than a percent of the variance in reading comprehension test scores. For the purposes of this subchapter, I will refer to the total figures of the two groups together.

A similar method of analysis was applied in all previous studies that will be discussed here. Multiple linear regression was used to examine the possible addition of explanatory power of vocabulary depth in reading comprehension, in addition to the explanatory power already provided by vocabulary breadth alone. Nonetheless, there is a minor difference that is related to reporting the R^2 values. The previous studies merely report R^2 values, whereas I decided to report the adjusted R^2 values in the present study. The choice was explained in more detail in Chapter 6.3, but as a reminder, adjusted R^2 values are lower, but they represent the whole population better than mere R^2 values. Thus, it is likely that some of the differences found between my study and the previous studies are explained by this difference in reporting R^2 values.

In their study, Rashidi and Khosravi (2010) only looked at the separate contributions afforded by vocabulary breadth and depth in reading comprehension. According to them, vocabulary size accounted for 55% and vocabulary depth for 69% of the variance in reading comprehension test scores. The results are rather similar to those obtained in the present study where vocabulary size alone explained 64%, and, similarly, vocabulary depth alone explained 64% of variance in reading comprehension test results, when looking at the two groups of informants in total. In Qian and Schedl's (2004) study, only vocabulary depth and reading comprehension study were involved. The scores on their new vocabulary depth test explained 55% of the variance in reading comprehension test scores. Together with another vocabulary test

that was used as quality control, the two tests explained 61% of the variance.

The changes in R^2 values were reported in the studies of Qian (1999, 2002). In the first study (Qian 1999), vocabulary size alone explained 60% of the variance in reading comprehension scores. Together with vocabulary depth measured with the WAF, the two variables could explain 71% of the variance. This suggests that vocabulary depth added significant 11 percentage points of explained variance in reading comprehension test results. However, when the second measure of vocabulary depth, a morphological knowledge test, was further included, it only added another 1 percentage point of explained variance. In the second study (Qian 2002), vocabulary size accounted for 54% of the variance. In this study, two measures of vocabulary depth were used: the WAF and a synonym test. In addition to the explained variance afforded by vocabulary size alone, the WAF and the synonym test added 13 and 6 percentage points, respectively, of explained variance.

When examining the results of earlier studies and the present study, it seems that measuring vocabulary depth can indeed serve as a tool for predicting and explaining reading comprehension test scores. Both vocabulary size and depth alone were able to explain a considerable and similar amount of variance in reading comprehension test scores. However, the percentage of added explained variance provided by vocabulary depth (in addition to vocabulary size alone) was not very high, and, in the present study, it was indeed lower than in previous studies (Qian 1999, 2002). It might be that the test format played a role here because the VKS used in the present study is fundamentally different from the WAF used in the majority of earlier research. Nevertheless, the results of the present study suggest that measuring vocabulary depth as a means of predicting or explaining reading comprehension test scores is quite as efficient as measuring vocabulary size.

7.4 Comparison of the results found in the TOKA and YLI groups

Previous studies related to the present study were thoroughly discussed in the

previous subchapters, and next, the focus will be turned to comparing the TOKA and YLI groups. I will present some key differences found between the two groups of informants and try to explain them. Firstly, I would like to acknowledge an issue related to the tests used in this study. Because the original vocabulary size test VST was shortened from 140 questions to 100 mostly due to time limitations, the VST used in the present study only tested knowledge of the most frequent 10 000 English words. It is likely that the participants in both groups would have known at least some words in the higher frequency levels as well and this would have then resulted in somewhat higher estimates of their vocabulary size. Nevertheless, the VST was exactly the same for both groups, and thus, it is possible to compare the two groups. The same applies to the other two tests as well: the tests administered were exactly the same in the both study groups, which makes it possible to directly compare the results obtained.

An obvious difference between the groups is of course in the test scores across all three tests conducted. This is not a surprising finding taken into account that the two groups were of different competence levels: the TOKA group consisting of upper secondary school students and the YLI group of advanced university students of English. As expected, the YLI group performed better overall than the TOKA group in all three tests. Moreover, the differences within the two groups varied considerably. The more advanced YLI group was clearly more homogeneous whereas the differences within the less advanced TOKA group were very large. In addition to the proficiency level of the participants as an explanative factor, the possible roles of the testing situation and the participants' motivation have to be acknowledged as well. In the TOKA group, the tests were administered during an ordinary English lesson, but as the informants were told that their scores would not affect their course grade, it may be that the informants were less motivated to fill in the tests as they perhaps would have been if their scores had had an impact on their course grade. All participants in the YLI group, on the other hand, were English majors doing their teacher training, and it is thus likely that they were more motivated to participate in a study where English language skills were assessed. Moreover, it might be that university-level participants in general are more

willing to participate in studies like this, as they probably know more about the challenges of finding informants and conducting research on the whole than upper secondary school students, who most likely do not yet have much experience or knowledge about university studies.

When it comes to the correlations found between the three variables (vocabulary size, vocabulary depth, and reading comprehension), they were stronger in the TOKA group than in the more advanced YLI group. The correlations seem to suggest that despite the lower correlations reported between vocabulary size and reading comprehension as well as between vocabulary depth and reading comprehension, the informants in the YLI group still performed very well on the reading comprehension test. Considering their very high competence of English and their level of studies, it is likely that they are able to use more effective reading strategies, for example. In addition, it might be that the tests conducted were simply at least partly too easy for the participants in the YLI group. The reading comprehension test was compiled of old matriculation exam tests and the VST only had items from the 10 most frequent 1000 word bands of English. However, this was taken into account when designing the methodology of the present study: it was expected that the very advanced YLI group would do well on all tests.

The significant finding here is that the results suggest that vocabulary knowledge, understood both as vocabulary size and vocabulary depth, plays a fundamental role in reading comprehension proficiency of learners who are not very advanced. When learners become more advanced, the significance of vocabulary knowledge seems to decrease to some extent; the very advanced learners seem to reach high reading comprehension scores despite their scores on the VST and the VKS. It is likely that on the more advanced levels of competence, the learners also possess a large variety of other skills, such as more advanced linguistic knowledge and reading strategies, which help them when encountering possible problems in L2 reading. It indeed seems that in the developing, less advanced stages of proficiency, vocabulary size and depth affect reading comprehension more strongly than in the more advanced stages of proficiency.

8 Conclusion

Vocabulary knowledge has a central role in L2 learning in general and in reading comprehension specifically. This thesis set out to examine how vocabulary size, vocabulary depth and reading comprehension relate to each other. Furthermore, special emphasis was placed on depth of vocabulary knowledge, as this dimension is clearly a fundamental part of vocabulary knowledge, but, unfortunately, has not been included in many studies. The second aim of the study was thus to examine the kind of information that L2 vocabulary depth can possibly bring into the prediction and explanation of L2 learners' reading comprehension, in addition to that provided by vocabulary breadth alone. The third aim of the study was to compare two groups at different stages of proficiency and see if there are any differences regarding the first two research questions between the groups. The TOKA group consisted of 39 upper secondary school students and the YLI group of 19 university students of English. All participants filled in a background information form and sat three tests: a multiple-choice reading comprehension test, the VST to measure vocabulary size, and the VKS to measure vocabulary depth.

The test scores revealed, as expected, that the more advanced YLI group performed better than the less advanced TOKA group in all tests. Moreover, the YLI informants reached similar scores with each other and the variance in the group's results was not large. In the TOKA group, on the other hand, the variation was large: the most successful informants obtained scores that were close to the scores of the YLI group whereas the weakest informants' scores were very low. The YLI group's results were clearly more uniform, whereas the TOKA group's scores were scattered.

The measures of vocabulary size and vocabulary depth were strongly correlated both in the TOKA and YLI groups and in the two groups together. The correlations between vocabulary size and reading comprehension scores as well as vocabulary depth and reading comprehension scores were strong in the TOKA group and in the two groups in total. However, the correlations

were clearly lower in the YLI group. When looking at the explained variance in reading comprehension test scores afforded by vocabulary size and depth, the results were in line with the correlations. In the TOKA group, vocabulary size alone explained 38% of the variance, and vocabulary depth added one percentage point of explained variance to this figure. Vocabulary depth alone explained a similar 38% of the variance. In the two groups together, vocabulary size explained 64% of the variance to which vocabulary depth added another 2 percentage points of explained variance. Again, vocabulary depth alone could explain a similar 64% of the variance in reading comprehension scores. As with the correlation coefficients, it was the results of the YLI group that differed: vocabulary size could only explain 19% of the variance in the reading comprehension test results, and vocabulary depth did not indeed add anything to this figure. Moreover, vocabulary depth alone explained less than 1% of the variance in the reading comprehension scores.

The results of the present study suggest that both vocabulary breadth and depth can be used as predictors of reading comprehension proficiency, as they seem to provide a similar factor in the explanation of reading comprehension test scores. However, caution has to be applied when considering advanced learners. It seems that at the more advanced proficiency levels, the learners will succeed in reading comprehension despite their vocabulary knowledge. It is likely that advanced learners have a wide range of other skills, such as the use of other linguistic knowledge or reading strategies, that they can resort to when encountering problems in reading. In the beginning and developing stages of competence, on the other hand, both vocabulary size and depth are good predictive and explanative factors of reading comprehension. In both of the study groups as well as in the two groups together, vocabulary depth did not seem to add much into explaining success in reading comprehension. However, the results suggest that when examined separately, vocabulary breadth and depth both have equal predictive and explanative power.

When examining and interpreting the results of the present study, some limitations have to be acknowledged. As explained in the theoretical section of the thesis, vocabulary depth was understood as developing degrees of

knowledge, and this affected the choice of the test format. In previous studies that have had similar topics of interest, vocabulary depth has often been considered as different dimensions of knowledge and the tests used to elicit knowledge on learners' depth of vocabulary knowledge have mainly concentrated on meaning, polysemy, synonymy, and collocations. As a construct, vocabulary depth is difficult to define thoroughly and this affects measuring it as well. There is no standardized vocabulary depth test format and the few existing tests each come with its own limitations. Furthermore, the relationship between vocabulary breadth and depth is a close one, and the two are overlapping and complementing in nature. It is challenging to test them separately and it is questionable if they could even be tested totally isolated from each other. This was shown in the present study as well: the scores on vocabulary size and depth tests correlated strongly. Another limitation is related to the VST: strictly speaking, it only tested the knowledge of 10 000 most frequent English words. In the case of the YLI group especially, it may have been useful to test the knowledge of less frequent items as well to gain a fuller picture of the YLI participants' vocabulary size. Nevertheless, it was hypothesised already in the beginning of the study that the YLI group would score well on all tests. When discussing the limitations of the study, the small sample size involved has to be acknowledged as well, especially in the case of the YLI group. Due to difficulties in finding informants, there were only 19 informants in the YLI group, and, at least partly because of this, some results obtained in the YLI group were not statistically significant.

Considering the limitations of the present study and the gap that there is in research in this field, some suggestions for further research are made next. Given the many-sided nature of vocabulary depth, it would be useful to examine its role in reading comprehension from different perspectives, perhaps by including different tests of vocabulary depth into a study. For example, there could be different tests for measuring vocabulary depth separately as dimensions and degrees of knowledge. Vocabulary depth is still a rather obscure feature of vocabulary knowledge and it ought to be examined more. As explained in the theoretical section of the present study, there is more to learning a word than mere form-meaning connections. In this study, the

number of informants in total was a little restricted, which is why bigger sample sizes are suggested for future research. Furthermore, out of four L2 skills, only reading was involved in the present study. The number of previous studies in this field is very limited, but it is even more so in the case of other L2 skills, i.e. listening, writing, and speaking. In order to reach a fuller understanding of the nature of vocabulary depth and its role in overall L2 proficiency, more research is needed in this field.

References

- Ala-Akkala Minna 2010. *The Development of English Vocabulary During Upper Secondary School*. Unpublished Master's Thesis, University of Turku.
- Alderson Charles J. 2000. *Assessing Reading*. Cambridge: Cambridge University Press.
- Alderson Charles J., Haapakangas Eeva-Leena, Huhta Ari, Nieminen Lea & Ullakonoja Riikka 2015. *The Diagnosis of Reading in a Second or Foreign Language*. New York: Routledge.
- August Diane, Carlo Maria, Dressler Cheryl & Snow Catherine 2005. The critical role of vocabulary development for English language learners. *Learning Disabilities Research and Practice* 20 (1): 50–57.
- Bertram Raymond, Baayen Harald R. & Schreuder Robert 2000. Effects of family size for complex words. *Journal of Memory and Language* 42: 390–405.
- Bertram Raymond, Laine Matti & Virkkala Minna Maria 2000. The role of derivational morphology in vocabulary acquisition: Get by with a little help from my morpheme friends. *Scandinavian Journal of Psychology* 41 (4): 287–296.
- Birch Barbara 2011. Out of my orthographic depth: Second language reading. In Hinkel (ed.): 488–506.
- Bogaards Paul & Laufer Batia (eds) 2004. *Vocabulary in a Second Language: Selection, Acquisition, and Testing*. Amsterdam: John Benjamins Publishing Company.
- Brown Dale 2010. What aspects of vocabulary knowledge do textbooks give attention to? *Language Teaching Research* 15 (1): 83–97.
- Brown Gillian, Malmkjaer Kirsten & Williams John (eds) 1996. *Performance and Competence in Second Language Acquisition*. Cambridge: Cambridge University Press.
- Bruton Anthony 2009. The vocabulary knowledge scale: A critical analysis. *Language Assessment Quarterly* 6: 288–297.
- Carlo M., August D., McLaughlin B., Snow C., Dressler C., Lippman D., Lively T. & White C. 2004. Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream

- classrooms. *Reading Research Quarterly* 39 (2): 188–215.
- Carter Ronald & McCarthy Michael (eds) 1988. *Vocabulary and Language Teaching*. London: Longman.
- Carver Ronald P. 1994. Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Reading Behavior* 26 (4): 413–437.
- Clapham Caroline & Corson David (eds) 1997. *Encyclopedia of Language and Education. Volume 7: Language Testing and Assessment*. Dordrecht: Kluwer Academic Publishers.
- Coady James & Huckin Thomas (eds) 1997. *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*. Cambridge: Cambridge University Press.
- Culligan Brent 2015. A comparison of three test formats to assess word difficulty. *Language Testing* 32 (4): 503–520.
- Dale E. 1965. Vocabulary measurement: Techniques and major findings. *Elementary English* 42: 895–901, 948.
- Ehsanzadeh Seyed Jafar 2012. Depth versus breadth of lexical repertoire: Assessing their roles in EFL students' incidental vocabulary acquisition. *TESL Canada Journal* 29 (2): 24–41.
- Grabe William 1991. Current developments in second language reading research. *TESOL Quarterly* 25 (3): 375–406.
- Grunwell P. (ed.) 1988. *Applied Linguistics in Society*. London: Centre for Information on Language Teaching and Research.
- Gyllstad Henrik, Vilkaite Laura & Schmitt Norbert 2015. Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *International Journal of Applied Linguistics* 166 (2): 276–303.
- Henriksen Birgit 2006. Exploring the quality of lexical knowledge in the language learner's L1 and L2. In Pietilä, Lintunen & Järvinen (eds): 13–36.
- Henriksen Birgit & Danelund Lise 2015. Studies of Danish L2 learners' vocabulary knowledge and the lexical richness of their written production in English. In Pietilä, Doró & Pípalová (eds): 29–55.

- Hinkel Eli 2011. *Handbook of Research in Second Language Teaching and Learning: Volume II*. New York: Routledge.
- Jaatinen Sirma & Mankkinen Tarja 1993. The size of vocabulary of university students of English. In Sajavaara & Takala (eds): 147–211.
- Joe Angela 1998. What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics* 19 (3): 357–377.
- Karapalo Elina, McWhirr Jim, Mäki Jaakko, Päckilä Teijo, Riite Marjut & Silk Riitta 2009. *Open Road: Course 2*. Helsinki: Kustannusosakeyhtiö Otava.
- Karapalo Elina, McWhirr Jim, Mäki Jaakko, Päckilä Teijo, Riite Marjut & Silk Riitta 2009. *Open Road: Course 5*. Helsinki: Kustannusosakeyhtiö Otava.
- Karapalo Elina, McWhirr Jim, Mäki Jaakko, Päckilä Teijo, Sadeharju Marjut & Silk Riitta 2011. *Open Road: Course 7*. Helsinki: Kustannusosakeyhtiö Otava.
- Koda Keiko 2012. How to do research on second language reading. In Mackey & Gass (eds): 158–177.
- Laufer Batia 1992. Reading in a foreign language: How does L2 lexical knowledge interact with the reader's general academic ability? *Journal of Research in Reading* 15 (2): 95–103.
- Laufer Batia 1996. The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In Coady & Huckin (eds): 20–34.
- Larson-Hall Jenifer 2016. *A Guide to Doing Statistics in Second Language Research Using SPSS and R*. New York, NY: Routledge, Taylor & Francis Group.
- Li Kezhen 2015. A study of vocabulary knowledge and reading comprehension on EFL Chinese learners. *Studies in Literature and Language* 10 (1): 33–40.
- Mackey Alison & Gass Susan M. (eds) 2012. *Research Methods in Second Language Acquisition: A Practical Guide*. Chichester: Wiley-Blackwell.
- Meara Paul 1996. The dimensions of lexical competence. In Brown, Malmkjaer & Williams (eds): 35–53.

- Meara Paul & Jones Glyn 1988. Vocabulary size as a placement indicator. In Grunwell (ed.): 80–87.
- Mehrpour Saeed & Rahimi Mohammad 2010. The impact of general and specific vocabulary knowledge on reading and listening comprehension: A case of Iranian EFL learners. *System* 38 (2): 292–300.
- Muijs Daniel 2004. *Doing Quantitative Research in Education with SPSS*. London: SAGE Publications Ltd.
- Nagy William, Anderson Richard C., Schommer Marlene, Scott Judith, Stallman Anne 1989. Morphological families in the internal lexicon. *Reading Research Quarterly* 24 (3): 262–282.
- Nassaji Hossein 2003. Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *The Modern Language Journal* 87: 261–276.
- Nation I. S. P. 1983. Testing and teaching vocabulary. *Guidelines* 5 (1): 12-25.
- Nation I. S. P. 1990. *Teaching and Learning Vocabulary*. New York: Newbury House.
- Nation I. S. P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation I. S. P. 2006. How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review* 63 (1): 59–82.
- Nation Paul 1993. Vocabulary size, growth, and use. In Schreuder & Weltens (eds): 115–134.
- Nation Paul 2012. The vocabulary size test: Instruction and description [online]. Available: <http://www.victoria.ac.nz/lals/about/staff/paul-nation> (21 March 2016).
- Nation Paul & Beglar James 2007. A vocabulary size test. *The Language Teacher* 31 (7): 9–13.
- Nation Paul & Coady James 1988. Vocabulary and reading. In Carter & McCarthy (eds): 97–110.
- Nation Paul & Meara Paul 2002. Vocabulary. In Schmitt (ed.): 35–54.
- Nation Paul & Waring Robert 1997. Vocabulary size, text coverage and word lists. In Schmitt & McCarthy (eds): 6–19.
- Ortega Lourdes 2009. *Understanding Second Language Acquisition*. New

York: Routledge.

- Paribakht Sima T. & Wesche Marjorie 1997. Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In Coady & Huckin (eds): 174–200.
- Perfetti Charles A., Landi Nicole & Oakhill Jane 2005. The acquisition of reading comprehension skill. In Snowling & Hulme (eds): 227–247.
- Pietilä Päivi, Doró Katalin & Pípalová Renata (eds) 2015. *Lexical Issues in L2 Writing*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Pietilä Päivi, Lintunen Pekka & Järvinen Heini-Marja (eds) 2006. *Kielenoppija tänään - Language Learners of Today*. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA.
- Pirilä Amanda 2012. *Knowledge of English Vocabulary at Advanced Levels: Vocabulary Size and Academic Vocabulary Knowledge of Finnish 1st Year University Students and MA Students of English*. Unpublished Master's Thesis, University of Turku.
- Proctor C., Carlo M., August D. & Snow C. 2005. Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology* 97 (2): 246–256.
- Qian David D. 1999. Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review* 56 (2): 282–307.
- Qian David D. 2002. Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning* 52 (3): 513–536.
- Qian David D. 2005. Demystifying lexical inferencing: The role of aspects of vocabulary knowledge. *TESL Canada Journal* 22 (2): 34–53.
- Qian David D. & Schedl Mary 2004. Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing* 21 (1): 28–52.
- Rashidi N. & Khosravi N. 2010. Assessing the role of depth and breadth of vocabulary knowledge in reading comprehension of Iranian EFL learners. *Journal of Pan-Pacific Association of Applied Linguistics* 14 (1): 81–108.
- Read John 2000. *Assessing Vocabulary*. Cambridge: Cambridge University

Press.

- Read John 2004. Plumbing the depths: How should the construct of vocabulary knowledge be defined? In Bogaards & Laufers (eds): 209–227.
- Richards Jack C. 1976. The role of vocabulary teaching. *TESOL Quarterly* 10 (1): 77–89.
- Sajavaara Kari & Takala Sauli (eds) 1993. *Finns as Learners of English: Three studies*. Jyväskylä: University of Jyväskylä.
- Saville-Troike Muriel 2006. *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press.
- Schmitt Norbert (ed.) 2002. *An Introduction to Applied Linguistics*. London: Edward Arnold.
- Schmitt Norbert 2008. Review article – Instructed second language vocabulary learning. *Language Teaching Research* 12 (3): 329–363.
- Schmitt Norbert 2010. *Researching Vocabulary: A Vocabulary Research Manual*. Hampshire: Palgrave Macmillan.
- Schmitt Norbert 2014. Size and depth of vocabulary knowledge: What the research shows. *Language Learning* 64 (4): 913–951.
- Schmitt Norbert, Jiang Xiangying & Grabe William 2011. The percentage of words known in a text and reading comprehension. *The Modern Language Journal* 95: 26–43.
- Schmitt Norbert & McCarthy Michael (eds) 1997. *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.
- Schmitt Norbert, Schmitt Diane & Clapham Caroline 2001. Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing* 18 (1): 55–88.
- Schreuder Robert & Weltens Bert (eds) 1993. *The Bilingual Lexicon*. Amsterdam: John Benjamins Pub.
- Snowling Margaret J. & Hulme Charles (eds) 2005. *The Science of Reading*. Malden: Blackwell.
- The Matriculation Examination Board 2012. English: Written part. Pitkä oppimäärä / lång lärokurs (advanced level), 17.09.2012 [online]. Available: http://yle.fi/progressive/fynd/oppiminen/oppiminen.yle.fi/yo-koheet/englanti_kirjallinen_pitka_s2012.pdf (21 March 2016).

- The Matriculation Examination Board 2013. English: Written part. Pitkä oppimäärä / lång lärokurs (advanced level), 15.03.2013 [online]. Available: <http://yle.fi/progressive/fynd/oppiminen/oppiminen.yle.fi/yo-koheet/eat-k13.pdf> (21 March 2016).
- Wallace Christopher 2007. Vocabulary: The key to teaching English language learners to read. *Reading Improvement* 44 (4): 189–193.
- Weir C. J. 1997. The testing of reading in a second language. In Clapham & Corson (eds): 39–49.
- Wesche Marjorie & Paribakht Sima T. 1996. Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review* 53 (1): 13–40.

Appendix 3. Reading comprehension test.

Name: _____ Group: _____

Read each text carefully. Circle the right alternative. Note: there is only one right answer per question so please do not circle more than one option.

The Distant Hours

It started with a letter. A letter that had been lost a long time, waiting out half a century in a forgotten postal bag in the dim attic of a nondescript, uninteresting house in Bermondsey. I think about it sometimes, that mailbag: of the hundreds of love letters, grocery bills, birthday cards, notes from children to their parents, that lay together, swelling and sighing as their thwarted messages whispered in the dark. Waiting, waiting, for someone to realize they were there. For it is said, you know, that a letter will always seek a reader; that sooner or later, like it or not, words have a way of finding the light, of making their secrets known.

Forgive me, I'm being romantic – a habit acquired from the years spent reading nineteenth-century novels with a torch when my parents thought I was asleep. What I mean to say is that it's odd to think that if Arthur Tyrell had been a little more responsible, if he hadn't had one too many rum toddies that Christmas Eve in 1941 and gone home and fallen into a drunken slumber instead of finishing his mail delivery, if the bag hadn't then been tucked in his attic and hidden until his death some fifty years later when one of his daughters unearthed it and called the Daily Mail, the whole thing might have turned out differently. For my mum, for me, and especially for Juniper Blythe.

Kate Morton, *The Distant Hours*, 2010

1. In this text, why is one letter being discussed?
 - A Because it had been sought after for a long time
 - B Because it ended up affecting several peoples' lives
 - C Because it had featured unexpected contents

2. What did Arthur Tyrell do?
 - A Delivered the mail late
 - B Failed to complete his duties
 - C Slept in the next morning

3. What started the whole chain of events?
 - A The secret message in a letter
 - B The Daily Mail dug something up
 - C Someone initially tried to cover up a mistake

Editor's Note: Deep Roots

Regard for the land runs deep in photographer Jim Richardson's life. When he flies over the fields of Cornwall, England, he feels the pull of places his ancestors farmed. He has even visited the homestead of his Celtic ancestors, who lived when Anglo-Saxons arrived in what would become England. His parents were descendants of yet other immigrants, drawn to the Kansas plains, where they could farm on a scale unimaginable in England.

Jim was not destined to continue his family's farming tradition. He was destined to document it, as this month's story on heirloom seeds shows. The article explains that the diversity of heirloom seeds, heritage varieties typical of particular areas, is critical to ensuring our food supply, and that a wide range of heirlooms is the best bet against disease and drought.

Such themes resonate with Jim. "My emotional landscape is forever haunted by the necessity of rain, shaped in childhood by parents who wondered when it would rain, and if not – what they would do," he said. "Some children fear divorce. I feared drought, the one thing that could destroy the security of a childhood life on the farm." For Jim, love of the land is elemental. Picture him in Ethiopia, seeing men harvesting oats by hand with sickles – a scene "straight out of some medieval tapestry," he said. Such hard work must be unrelenting drudgery, he thought. He got closer. "They were singing. This was the same land where hundreds of thousands died during the famine, yet there was joy in their voices and laughter." So Jim Richardson, who speaks the same language of the land as those harvesters, found himself laughing as well. – Chris Johns, Editor in Chief

National Geographic, July 2011

4. Why does Jim Richardson appreciate the countryside?
 - A He considers the scenery photographic
 - B He comes from a long line of farmers
 - C He wants to keep up family traditions

5. What does this month's story focus on?
 - A The importance of preserving original types of grain
 - B The necessity to develop new types of crops
 - C The need to replace old seeds with modified ones

6. Why is Jim Richardson's background mentioned at all in this text?
 - A He represents the opposite of what his family stands for
 - B He tends to live a healthy life
 - C He is deeply influenced by his heritage

7. What seems to be particularly essential for him?
 - A Exploring varying geographical locations
 - B Being in a professionally rewarding occupation
 - C Having an inner appreciation for soil

8. What in Ethiopia took him by surprise?
 - A The results of the harvest
 - B The kindred spirits encountered
 - C The local music traditions observed

Life Lessons: How to Spot a Liar

The following features four persons' tips for determining whether or not someone is lying.

Gregg McCrary, a retired profiler and analyst: As an investigator, I first try to assess how someone normally interacts. To do that, I begin an interview by asking questions that I know the answer to, like "What's your full name?" or "Where do you live?" Some folks are naturally animated and talk fast; others are more subdued. Once I know which type of talker a person is, I start asking him questions that I don't know the answer to. If his style shifts abruptly – going from calm to agitated or lively to mellow – chances are he's not telling the truth.

Jeffrey Hancock, an associate professor of communication: In my research on online lying, I've discovered that when people fib, they tend to use the first-person pronouns less often than people who are being truthful. Instead, they'll speak about themselves in the third person ("This is a girl who loves to ski") or even shorten their language ("Really into listening to jazz") – anything to give them psychological distance from the lie.

Julia Chung, an assistant principal: Ask most people what they were doing last week and they'll have to pause and think about it. That's even more true of teenagers, who generally don't have the capacity to tell an elaborate story on the fly. So when I call a person into my office and she seems totally rehearsed – there's zero hesitation before she answers a question – well, that's a dead giveaway.

Joseph Buckley, the president of a company training law-enforcement investigators: To sell us on the integrity of their answers, liars often use phrases emphasizing the validity of their statements, such as "to tell the truth" and "to be perfectly honest." These verbal tip-offs frequently invoke religion. Think of expressions like "I swear on a stack of Bibles" and "as God is my witness." Most truthful people don't need to go that far.

Real Simple, July 2011

9. According to Gregg McCrary, what is typical of liars?
 - A They strictly control their manner of speaking
 - B Their manner of speaking radically changes
 - C Their manner of speaking seems unaffected

10. According to Jeffrey Hancock, what is typical of online liars?
 - A They praise themselves
 - B They seem self-absorbed
 - C They avoid saying "I"

11. According to Julia Chung, how can one tell a young person is lying?
 - A She has a prompt answer for everything
 - B She thinks carefully about her response
 - C She often speaks uneasily

12. According to Joseph Buckley, what is typical of liars?
 - A They underline how religious they are
 - B They repeatedly point out their honesty
 - C They have no difficulty in impressing their listeners

Appendix 4. Vocabulary Knowledge Scale.

Note: In order to save some space, only the format of the test along with the list of words used in the test are given.

Name: _____

Group: _____

3. Valitse jokaisen sanan kohdalla omaa osaamistasi kuvaava vastaus. Täytä tarvittaessa myös tyhjät kohdat. Kohdissa III ja IV pyydetään kirjoittamaan ”synonym or translation”. Näissä kohdissa voit kirjoittaa samaa tarkoittavan englanninkielisen synonyymien tai suomenkielisen käännöksen kysytylle sanalle. Mikäli vastaat jonkun sanan kohdalla kohtaan V, vastaathan myös kohtaan IV.

Esimerkki:

Mark your answer here.	Categories	WORD: joke
	I	I don't remember having seen this word before.
	II	I have seen this word before, but I don't know what it means.
	III	I have seen this word before, and I think it means _____ (synonym or translation)
	IV	I know this word. It means <i>vitsi</i> _____ (synonym or translation)
X	V	I can use this word in a sentence: <i>She told me a funny joke.</i> _____ _____ (If you do this section, please also do Section IV.)

List of words used in this part of the test:

Strong, equipment, environment, interpretation, confident, precious, diary, exhilarating, doughnut, to amputate, infantile, rapacious, wallflower, parsonage.

Suomenkielinen tiivistelmä

Johdanto

Sanaston osaamista pidetään keskeisenä tekijänä vieraan kielen oppimisessa. Tämän tutkimuksen aiheena oli suomalaisten englannin kielen oppijoiden sanasto-osaaminen ja luetunymmärtäminen. Tutkimukseni tavoitteena oli selvittää, miten kaksi sanasto-osaamisen ulottuvuutta, sanaston koko ja syvyys, liittyvät luetunymmärtämiseen. Aiemmissa tutkimuksissa on tutkittu erityisesti sanaston koon ja luetunymmärtämisen välistä suhdetta, mutta sanasto-osaamisen syvyyttä ja sen yhteyttä luetunymmärtämiseen ei ole juurikaan tutkittu. Näin ollen tämän tutkimuksen toisena tavoitteena olikin tutkia, millaista lisätietoa sanasto-osaamisen syvyys voisi tarjota luetunymmärtämisen taitojen ennustamisessa ja selittämisessä. Tutkimuksessa oli mukana kaksi koehenkilöryhmää, joista toinen koostui lukio-opiskelijoista ja toinen englannin kielen yliopisto-opiskelijoista. Tutkimukseni kolmantena tavoitteena oli kuvailla eroja, joita näiden ryhmien suoritusten väliltä mahdollisesti löytyisi.

Tutkimuksen teoreettinen tausta

Termin *sana* määrittelyyn liittyy monia eri käsitteitä. Tässä tutkimuksessa *sanalla* viitataan *sanaperheeseen* (*word family*). *Sanaperhe* on laaja termi, joka sisältää sanan perusmuodon lisäksi myös sanan taivutusmuotoja ja johdannaisia. Näin ollen esimerkiksi sanat *lukea*, *luet*, *lukija* ja *lukematta* kuuluvat samaan sanaperheeseen.

Myös termin *sanasto* määrittelyyn liittyy useita käsitteitä. Yksinkertaisesti ilmaistuna *sanastolla* tarkoitetaan kielen komponenttia, joka koostuu sanoista ja niiden merkityksistä (Saville-Troike 2006: 191). Nykyään sanaston ajatellaan kuitenkin olevan hyvinkin monipuolinen käsite, joka ei koostu vain yksipuolisista sana-merkitys -pareista.

Sanasto-osaamisessa erotellaan tavallisesti kaksi ulottuvuutta: *sanasto-osaamisen koko* (*vocabulary breadth* tai *vocabulary size*) ja *sanasto-*

osaamisen syvyys (vocabulary depth). Sanasto-osaamisen koolla tarkoitetaan osattujen sanojen, tässä tutkimuksessa siis sanaperheiden, lukumäärää (ks. esim. Qian 1999: 283, Qian 2002: 515). Sanasto-osaamisen syvyyden määrittely on puolestaan hankalampaa. Termillä viitataan siihen, kuinka hyvin oppija tuntee osaamansa sanat (*ibid.*). Käytännössä sanasto-osaamisen syvyyttä lähestytään yleensä joko kehittyvänä taitona tai eri osataidoista koostuvana taitona (ks. esim. Read 2000, Schmitt 2010). Aiemmissä tutkimuksissa, joissa on tutkittu sanasto-osaamisen kokoa, syvyyttä ja luetunymmärtämistä, sanasto-osaamisen koko on nähty eri osataidoista koostuvana taitona. Tietämykseni mukaan samanlaista tutkimusta ei ole toteutettu niin, että sanaston syvyys olisikin käsitetty kehittyvänä taitona, ja siksi tutkimuksessani sanaston syvyys määriteltiin kehittyväksi taidoksi. Kehittyvällä taidolla viitataan siihen, että oppijan tietämys tietyistä sanasta kehittyy ja kasvaa: aluksi oppija ei tunne sanaa ollenkaan, seuraavaksi hän oppii tunnistamaan sanan ja antamaan sille jonkinlaisen merkityksen, ja lopulta oppijalla on elaboroitua tietoa sanan merkityksestä (Read 2004: 213–216). Nämä kaksi sanasto-osaamisen ulottuvuutta, koko ja syvyys, liittyvät hyvin läheisesti toisiinsa, mutta ovat silti erillisiä dimensioita. Siksi on tärkeää, että sanasto-osaamisen tutkimuksessa käsitellään molempia ulottuvuuksia.

Sanaston roolia vieraan kielen oppimisessa on tutkittu paljon. Sanasto-osaaminen on erittäin tärkeä osatekijä vieraan kielen oppimisessa ja sanasto onkin keskeinen osa kaikissa vieraan kielen taidoissa (lukeminen, kirjoittaminen, kuunteleminen ja puhuminen). Useat tutkimustulokset ovat osoittaneet, että oppijat, joiden sanasto-osaamisen koko on suuri, osaavat vierasta kieltä paremmin kuin oppijat, joiden sanasto-osaamisen koko on pieni (ks. esim. Meara 1996). Oppijalle ei kuitenkaan riitä monen sanan osaaminen, vaan opitun sanaston täytyy olla myös hyödynnettävissä parhaalla mahdollisella tavalla ja oppijan tulee osata käyttää opittua sanastoa (Nation & Meara 2002: 43). Vieraan kielen sanaston oppiminen on siis sekä määrällinen että laadullinen haaste: määrällinen siksi, että kunkin kielen sanasto koostuu tuhansista sanoista, ja laadullinen siksi, että sanaston oppiminen vaatii jokaisen sanan kohdalla usean eri aspektin oppimista (Laufer & Nation 2012: 163). Toisaalta on myös muistettava, ettei sanaston ja vieraan kielen

osaamisen suhde ole yksisuuntainen. Oppija, jonka sanasto-osaaminen on hyvällä tasolla, pystyy käyttämään kieltä rohkeasti, ja kielenkäyttö puolestaan mahdollistaa sanasto-osaamisen kasvamisen entisestään, mikä taas vuorostaan mahdollistaa oppijan osallistumisen useampiin ja monipuolisempiin kielenkäyttötilanteisiin, mikä edelleen johtaa sanasto-osaamisen kehittymiseen.

Tämän tutkimuksen kannalta on tärkeää käsitellä myös sanaston roolia erityisesti luetunymmärtämisessä. Oppijan lukiessa tekstiä matalan ja korkean tason tiedonkäsittelyprosessit toimivat vuorovaikutuksessa keskenään. Matalan tason prosesseilla tarkoitetaan muun muassa tekstin pienempien osien kokoamista sanoiksi, lauseiksi ja kappaleiksi. Korkean tason prosesseilla taas viitataan esimerkiksi tekstin syntaktisten ja semanttisten ominaisuuksien analysointiin ja tulkintaan. Tarvittaessa lukija voi turvautua lukustrategioihin. Lisäksi lukija integroi uutta tietoa vanhaan ja tarvittaessa käyttää jo olemassaolevaa tietoa uuden tekstin ymmärtämiseen. Lukeminen onkin monimutkainen prosessi, jossa lukija käyttää kielellisen tiedon lisäksi myös kognitiivisia taitoja.

Sanasto-osaamisen rooli vieraan kielen luetunymmärtämisessä on merkittävä, ja sen onkin arveltu olevan tärkein yksittäinen selittävä tekijä luetunymmärtämisessä (Nation & Coady 1988: 98). Muilla tekijöillä voi toki jonkin verran kompensoida riittämätöntä sanasto-osaamista, mutta tässä tapauksessa ymmärrys jää kuitenkin vajaaksi (Laufer 1997). Useat tutkijat ovat esittäneet arvioita siitä, kuinka paljon sanastoa vieraan kielen oppija tarvitsee lukemista varten. Lähtökohta lukemiselle on jossakin 3000 ja 5000 sanan välillä (Laufer 1992: 100, Nation & Waring 1997: 10), mutta monessa tutkimuksessa mainitut luvut ovat isompia. Yleensä esitetään, että lukijan tulisi ymmärtää noin 98% tekstistä, jotta lukeminen olisi sujuvaa ja ymmärrys riittävää (ks. esim. Nation 2006, Schmitt, Jiang & Grabe 2011: 26, Schmitt *et al.* 2015: 1). Käytännössä tämä tarkoittaa, että lukijan tulisi osata noin 8000–9000 sanaa (Nation 2006: 59). Toki esitetyt luvut riippuvat paljon myös tekstityypistä ja siitä, minkälainen ymmärtämisen taso nähdään riittävänä.

Luetunymmärtämisessä ei yksin riitä, että oppija osaa tarpeeksi monta sanaa, sillä myös sanaston syvyydellä on merkitystä. Mikäli oppijan sanasto-osaamisen syvyys on hyvällä tasolla, sanojen merkityksen hakeminen sanavarastosta helpottuu ja oppija pystyy käyttämään esimerkiksi kontekstia paremmin hyväkseen yrittäessään arvata tuntemattomien sanojen merkitystä (Qian 2002: 517–518, Qian 2005). Sanasto-osaaminen vaikuttaa ilmeisen voimakkaasti luetunymmärtämiseen, ja tässä tutkimuksessa haluttiin selvittää sanasto-osaamisen koon ja syvyyden roolia luetunymmärtämisessä. Kuvailen seuraavaksi tutkimuksen toteutusta.

Tutkimuksen toteutus

Tutkimukseen osallistui kaksi koehenkilöryhmää, joista toinen koostui toisen vuoden lukio-opiskelijoista (TOKA-ryhmä), ja toinen jo opinnoissaan pitkälle edenneistä Turun yliopiston englannin kielen pääaineopiskelijoista (YLI-ryhmä). Lukio-opiskelijoita oli yhteensä 39 ja yliopisto-opiskelijoita 19. Kokonaisuudessaan koehenkilöitä oli siis 58.

Kumpikin ryhmä täytti taustatietolomakkeen ja osallistui kolmeen testiin. Jokaisen testin täyttämiseen oli varattu kaksikymmentä minuuttia. Testin sai palauttaa aiemmin, mutta seuraavaa osiota ei saanut aloittaa ennen kuin kaikki olivat palauttaneet edellisen testin. Luetunymmärtämisen testi koostui kolmesta tekstistä, jotka oli valittu aiempien vuosien englannin kielen pitkän oppimäärän ylioppilaskokeista. Monivalintakysymyksiä oli 12, ja oikeasta vastauksesta sai yhden pisteen, eli maksimipistemäärä oli 12. Sanasto-osaamisen kokoa mitattiin monivalintatestillä (Nation & Beglar 2007: *Vocabulary Size Test*), joka koostui yhteensä sadasta kysymyksestä, joilla mitattiin englannin kielen 10 000 yleisimmän sanan osaamista. Maksimipistemäärä oli siis 100, ja tämä pistemäärä kerrottiin vielä sadalla, jotta saatiin tulokseksi koehenkilön osaamien sanojen lukumäärä. Sanasto-osaamisen syvyyttä arvioitiin puolestaan raportointimittarilla (Wesche & Paribakht 1996: *Vocabulary Knowledge Scale*), jossa koehenkilön tuli arvioida neljäntoista eri sanan osaamistaan asteikolla I-V. Jokainen vastaus pisteytettiin asteikolla 0-5, eli tämän osion maksimipistemäärä oli 70.

Tuloksia analysoitiin tilastollisin menetelmin ja apuna käytettiin *IBM SPSS Statistics 23.0.0.2* -ohjelmaa. Testitulosten välisiä suhteita arvioitiin *korrelaatiokertoimen* (r) avulla. Sanasto-osaamisen koon ja syvyyden roolia luetunymmärtämisessä tutkittiin *usean muuttujan lineaarisen regressioanalyysin* (*multiple linear regression*) avulla. Tämän analyysin avulla selvitettiin, kuinka suuri osa riippuvan muuttujan (eli luetunymmärtämistestin tulosten) varianssista voitiin selittää riippumattomilla muuttujilla (eli sanasto-osaamisen koolla, sanasto-osaamisen syvyyttä mittaavan testin tuloksilla ja näillä kahdella yhteensä). Regressioanalyysin tulokset esitettiin korjattuina R^2 -arvoina. Ryhmien välisiä eroja havainnoitiin sekä yksittäisten testien tuloksissa, korrelaatiokertoimissa että regressioanalyysituloksissa.

Analyysi ja keskeisimmät tutkimustulokset

Käsittelen tässä osiossa ensin jokaisen testin tuloksia erikseen ja siirryn sitten käsittelemään testitulosten välisiä korrelaatioita ja usean muuttujan lineaarisen regressioanalyysin tuloksia.

TOKA-ryhmän koehenkilöt tiesivät keskimäärin 5715 sanaa. Alimman tuloksen saanut koehenkilö tiesi 1500 sanaa, kun taas korkeimman pistemäärän saanut koehenkilö tiesi 8500 sanaa. YLI-ryhmän koehenkilöt tiesivät keskimäärin 8816 sanaa. Alin tulos oli 7300 ja korkein 9800 tiedettyä sanaa. Sanasto-osaamisen syvyyttä mittaavassa testissä TOKA-ryhmän koehenkilöt saivat tulokseksi keskimäärin 36 pistettä. Alin pistemäärä oli 16 ja korkein 58 pistettä. YLI-ryhmässä keskimääräinen tulos oli 60 pistettä. Alimman pistemäärän saanut koehenkilö sai testistä 52 pistettä, kun taas korkeimman pistemäärän saaneen koehenkilön tulos oli 66 pistettä. Luetunymmärtämistestissä TOKA-ryhmän koehenkilöt saivat keskimäärin 6 pistettä. Alin pistemäärä oli 0 ja korkein 11. YLI-ryhmän keskimääräinen tulos oli 11 pistettä. Alin pistemäärä oli 9 ja korkein täydet 12 pistettä.

Testitulosten välisiä korrelaatioita tarkasteltaessa erot kahden koehenkilöryhmän välillä olivat jälleen selviä. Sanasto-osaamisen koon ja sanasto-osaamisen syvyyttä mittaavan testin välinen korrelaatio oli hyvin vahva TOKA-ryhmässä ($r = 0.89$). Sanasto-osaamisen koko korreloi vahvasti

myös luetunymmärtämistestin tuloksen kanssa ($r = 0.65$). Korrelaatio sanasto-osaamisen syvyyttä mittaavan testin ja luetunymmärtämistestin välillä oli niin ikään vahva ($r = 0.59$). Korrelaatiot YLI-ryhmässä eivät olleet yhtä vahvoja kuin TOKA-ryhmässä. Sanasto-osaamisen koon ja sanasto-osaamisen syvyyttä mittaavan testin välillä oli vahva korrelaatio ($r = 0.72$), mutta sanasto-osaamisen koon ja luetunymmärtämistestin tulosten välinen korrelaatio oli heikompi ($r = 0.31$). Myös sanasto-osaamisen syvyyttä mittaavan testin ja luetunymmärtämistestin välinen korrelaatio oli heikompi ($r = 0.29$) kuin TOKA-ryhmässä. Kun tarkastellaan kumpaakin ryhmää yhdessä, korrelaatiot olivat vahvoja kaikkien testien välillä. Sanasto-osaamisen koon ja syvyyden välillä oli hyvin vahva korrelaatio ($r = 0.95$). Korrelaatiot sanasto-osaamisen koon ja luetunymmärtämistestin tulosten välillä sekä sanasto-osaamisen syvyyttä mittaavan testin tulosten ja luetunymmärtämistestin tulosten välillä olivat niinkään erittäin vahvoja ($r = 0.83$ ja $r = 0.95$).

Usean muuttujan lineaarisen regressioanalyysin tulokset erosivat myös selvästi kahden koehenkilöryhmän välillä. TOKA-ryhmän koehenkilöiden sanasto-osaamisen koko selitti noin 38% varianssista luetunymmärtämistestin tuloksissa. Kun analyysiin lisättiin sanasto-osaamisen syvyyttä mittaavan testin tulokset, sanasto-osaamisen koko ja syvyys yhdessä selittivät noin 39% luetunymmärtämistestin tulosten varianssista. Sanasto-osaamisen syvyyttä mittaavan testin tulokset selittivät yksinään noin 38% luetunymmärtämistestin tulosten varianssista. YLI-ryhmän tulokset erosivat selvästi TOKA-ryhmän tuloksista. Sanasto-osaamisen koko selitti noin 19% luetunymmärtämistestin tulosten varianssista. Yhdessä sanasto-osaamisen syvyyttä mittaavan testin tulosten kanssa nämä kaksi riippumatonta muuttujaa selittivät vain noin 15% luetunymmärtämistestin tulosten varianssista. Näin ollen sanasto-osaamisen syvyyttä mittaavan testin tulokset eivät lisänneet selitetyn varianssin arvoa. Sanasto-osaamisen syvyyttä mittaavan testin tulokset selittivät yksinään vain alle prosentin luetunymmärtämistestin tulosten varianssista. Kun tarkastellaan kahta tutkimusryhmää yhdessä, sanasto-osaamisen koko selitti noin 64% luetunymmärtämistestin tulosten varianssista. Sanasto-osaamisen syvyyttä mittaavan testin tulokset lisäsivät tähän noin 2 prosenttiyksikköä, jolloin nämä kaksi riippumatonta muuttujaa selittivät yhteensä noin 66%

luetunymmärtämistestin tulosten varianssista. Sanasto-osaamisen syvyyttä mittaavan testin tulokset selittivät yksinään noin 64% luetunymmärtämistestin tulosten varianssista.

Johtopäätökset

Testituloksia tarkasteltaessa erot kahden koehenkilöryhmän välillä olivat selviä. TOKA-ryhmän sisällä variaatio oli suurta: alimmat pistemäärät kaikissa kolmessa testissä olivat melko heikkoja, kun taas korkeimmat pistemäärät olivat lähellä YLI-ryhmän suoritustasoa. YLI-ryhmässä puolestaan variaatio oli selvästi pienempää ja ryhmän tulokset olivat melko yhteneväisiä. YLI-ryhmän kohdalla tämä ei sinänsä ole yllättävää, sillä kyseessä oli ryhmä pitkälle edenneitä englannin kielen pääaineopiskelijoita.

Testitulosten väliset korrelaatiot olivat vahvoja TOKA-ryhmässä ja kahdessa koehenkilöryhmässä yhteensä. YLI-ryhmässä korrelaatiot olivat jonkin verran heikompia, mutta silti positiivisia. Aiemmissa tutkimuksissa on osoitettu, että sanasto-osaamisen koko ja syvyys liittyvät voimakkaasti toisiinsa (ks. esim. Qian 1999, 2002, Rashidi & Khosravi 2010, Li 2015), ja tutkimukseni tulokset tukevat tätä havaintoa. Kuten tutkimukseni teoreettisesta taustasta kävi ilmi, myös sanasto-osaamisen koon ja luetunymmärtämisen on osoitettu liittyvän vahvasti toisiinsa. Tutkimukseni tulokset ovat linjassa aiempien tutkimustulosten kanssa. Sama koskee myös sanasto-osaamisen syvyyden ja luetunymmärtämisen välistä yhteyttä: vaikka aiempia tutkimuksia ei ole kovinkaan paljon, olemassaolevat tutkimukset kuitenkin viittaavat näiden kahden muuttujan välisen yhteyden olevan vahva, ja tutkimukseni tulokset tukevat tätä olettamusta. Aiempia tutkimustuloksia tarkasteltaessa on kuitenkin otettava huomioon se, että aiemmissa tutkimuksissa sanasto-osaamisen syvyys on nähty ennen kaikkea erilaisista osataidoista koostuvana taitona eikä kehittyvänä taitona, kuten tässä tutkimuksessa. Tämä on vaikuttanut etenkin käytettyihin testausmenetelmiin, joten aiemmat tutkimustulokset eivät ole suoraan verrannollisia tämän tutkimuksen tulosten kanssa. Tästä huolimatta vertailu kuitenkin osoittaa tutkimustulosten olevan samankaltaisia ja muistuttavan toisiaan.

Kun tarkastellaan sanasto-osaamisen syvyyden roolia luetunymmärtämisessä, aiempien tutkimusten määrä rajoittuu muutamaan tutkimukseen (ks. Qian 1999, 2002, Qian & Schedl 2004, Rashidi & Khosravi 2010). Vaikka tutkimuksissa käytetyt testit erosivatkin jonkin verran toisistaan, kaikissa tutkimuksissa käytettiin kuitenkin usean muuttujan lineaarista regressioanalyysia yhtenä tutkimusmenetelmänä. Toisin kuin tässä tutkimuksessa, aiemmissa tutkimuksissa on raportoitu pelkät R^2 -arvot korjattujen R^2 -arvojen sijaan. Korjatuissa R^2 -arvoissa on otettu huomioon se, että kyseessä on vain otos koko populaatiosta, ja siksi korjatut R^2 -arvot ovat usein pelkkiä R^2 -arvoja matalampia. Tämä siis selittänee osan aiempien tutkimusten ja tämän tutkimuksen tulosten välisistä eroista. Tutkimukseni tulokset ovat kuitenkin linjassa aiempien tutkimustulosten kanssa. Tutkimustuloksia vertailtaessa vaikuttaa siltä, että sanasto-osaamisen koko ja sanasto-osaamisen syvyys selittävät erikseen tarkasteltuina huomattavan osan varianssista luetunymmärtämistestin tuloksissa, etenkin TOKA-ryhmässä ja kahdessa koehenkilöryhmässä yhteensä. Aiemmissa tutkimustuloksissa on kuitenkin raportoitu sanasto-osaamisen syvyyden lisäämisen sanasto-osaamisen koon ja luetunymmärtämistestin tulosten väliseen analyysiin tuovan huomattavasti enemmän lisätietoa kuin tässä tutkimuksessa. Tämä voi johtua esimerkiksi käytettyjen testien välisistä eroista tai koehenkilöryhmien välisistä eroista. Aiempien tutkimusten ja tämän tutkimuksen tulokset ovat kuitenkin samankaltaisia ja tukevat olettamusta, jonka mukaan sanasto-osaamisen syvyys on yhtä tärkeä tekijä luetunymmärtämisessä kuin sanasto-osaamisen kokokin.

TOKA- ja YLI-ryhmien välillä oli selviä eroja itse testitulosten lisäksi myös korrelaatioiden ja usean muuttujan lineaarisen regressioanalyysin tuloksissa. Testitulosten väliset korrelaatiot olivat selvästi korkeampia TOKA-ryhmässä kuin YLI-ryhmässä. Samoin regressioanalyysin tulokset olivat korkeampia TOKA-ryhmässä, mikä viittaa siihen, että sanasto-osaamisen koko ja syvyys (sekä yhdessä että erikseen) selittivät suuremman osan luetunymmärtämistestin tulosten varianssista kuin YLI-ryhmässä. Kun otetaan huomioon YLI-ryhmän korkeat keskimääräiset pisteet luetunymmärtämistestissä, tutkimukseni tulokset viittaavat siihen, että

edistynyt YLI-ryhmä suoriutuu luetunymmärtämisessä hyvin huolimatta sanasto-osaamisen koosta tai syvyydestä. On todennäköistä, että YLI-ryhmän koehenkilöt pystyvät käyttämään paremmin hyödyksi myös muita kykyjä luetunymmärtämisessä, esimerkiksi lukemisstrategioita tai muuta kielitietoa. TOKA-ryhmän koehenkilöt eivät olleet yhtä vahvoja englannin kielen taitajia kuin YLI-ryhmän koehenkilöt, ja TOKA-ryhmän tuloksia tarkasteltaessa näyttää siltä, että heidän kohdallaan sanasto-osaamisen koko ja syvyys vaikuttavat suuresti luetunymmärtämiseen.

Tutkimukseni tulokset viittaavat siihen, että sanasto-osaamisen syvyyden mittaaminen on yhtä toimiva keino ennustaa ja selittää oppijoiden luetunymmärtämistaitoa kuin sanasto-osaamisen koon mittaaminen. Varovaisuutta on kuitenkin noudatettava erityisesti edistyneitä oppijoita tutkittaessa. Tutkimuksestani käy kuitenkin ilmi sanasto-osaamisen syvyyden monitahoinen rakenne, ja tämän vuoksi lisätutkimusta tällä alueella tarvitaan. Sanasto-osaamisen koon ja syvyyden vaikutusta muissa vieraan kielen taidoissa, kuten kirjoittamisessa, kuuntelemisessa ja puhumisessa, ei myöskään tunneta tarpeeksi.