UNIVERSITY OF JYVÄSKYLÄ

TEACHING AND TESTING VOCABULARY IN FINNISH UPPER SECONDARY SCHOOL: TEACHER PRACTICES

JARKKO KEKKI

MASTER'S THESIS

ENGLISH

DEPARTMENT OF LANGUAGE AND COMMUNICATION STUDIES

MAY 2018

JYVÄSKYLÄN YLIOPISTO

Tiedekunta – Faculty Humanistis-yhteiskuntatieteellinen tiedekunta	Laitos – Department Kieli- ja viestintätieteiden laitos
Tekijä – Author Jarkko Kekki	
Työn nimi – Title Teaching and testing vocabulary in Finnish upper secondary school: teacher practices	
Oppiaine – Subject Englanti	Työn laji – Level Maisterintutkielma
Aika – Month and year Toukokuu 2018	Sivumäärä – Number of pages 85 + 3 liitettä

Tiivistelmä – Abstract

Sanaston hallinta on perustavanlaatuinen osa kielitaitoa, ja se edeltää ja ennakoi osaamista kaikilla kielitaidon osa-alueilla (esim. kirjoittaminen, puhuminen, lukeminen ja kuunteleminen). Tämän tutkimuksen tavoitteena oli tutkia sitä, mitä sanan osaamisen osa-alueita (esim. ääntäminen, kirjoitusasu, kollokaatiot) lukion englanninopettajat arvostavat, mitä he testaavat sanakokeissa, ja mistä syistä opettajat pitävät sanakokeita. Koska näitä asioita ei juuri ole tutkittu Suomessa, tutkimuksen tarkoituksena on tuoda uutta tietoa asiasta ja ennen kaikkea toimia suunnanantajana jatkotutkimukselle.

Tutkielmaa varten haastateltiin yhdeksää lukion 1. vuosikurssilla opettavaa englanninopettajaa. Opettajilta kysyttiin haastatteluissa heidän käsityksistään ja toimintatavoistaan sanaston ja sanakokeiden suhteen. Aineisto analysoitiin käyttämällä ainestolähtöistä sisällönanalyysia.

Tutkimuksen tulokset osoittavat sen, että ylioppilaskirjoitukset hallitsevat lukio-opetusta ensimmäisistä kursseista lähtien, ja opettajien valitsemat metodit ja sisällöt ovat sellaisia, että ne harjoittaisivat mahdollisimman hyvin kirjoituksia varten. Toisaalta opettajilla on hyvin erilaisia tapoja opettaa sanastoa, ja sanakokeiden muoto vaihtelee suuresti opettajasta toiseen, sikäli mikäli opettaja niitä pitää.

Tuloksista ilmenee myös, että opettajat pitävät sanaston laajuutta tärkeämpänä kuin sanaston syvyyttä tai sanastojen välisiä verkostoja. Lisäksi kiteytynyt kieli (*formulaic language*) on opettajille vieras käsitteenä, sekä osittain myös ilmiönä. Sen painoarvo opetuksessa ja testauksessa on erittäin vähäinen lukion ensimmäisellä vuosikurssilla.

Asiasanat – Keywords EFL, vocabulary teaching, vocabulary testing, assessment, formulaic sequences, vocabulary knowledge, lexical competence

Säilytyspaikka – Depository JYX

Muita tietoja – Additional information

TABLE OF CONTENTS

1 INTRODUCTION 1

2 VOCABULARY 3

2.1 Defining words 3
2.2 Multi-word units 5

3 LEARNING AND TEACHING VOCABULARY 8

3.1 Types of vocabulary knowledge 9
3.2 Incidental and intentional learning 14
3.3 Repetition and learning 18
3.4 Language of instruction 20
3.5 Teaching formulaic language 22

4 VOCABULARY TESTS AND VOCABULARY IN FINNISH SCHOOLS 24

4.1 Types of tests and test items 25
  4.1.1 Types of tests 25
  4.1.2 Types of test items 30
4.2 Test reliability and validity 34
4.3 Vocabulary testing and teaching in Finnish schools 35

5 RESEARCH QUESTIONS 38

6 DATA AND METHODOLOGY 39

6.1 Data collection 39
6.2 Participants 41
6.3 Methods of analysis 42

7 EFL VOCABULARY TEACHING AND TESTING PRACTICES IN FINNISH UPPER SECONDARY SCHOOLS 44
7.1 Vocabulary knowledge in the teachers' perceptions 46
7.1.1 What aspects of lexical competence are important? 49
7.1.2 The importance of breadth and depth 52
7.1.3 Formulaic sequences in the interviewed teachers' teaching 54
7.1.4 Using context in teaching 58

7.2 Vocabulary tests 61
7.2.1 Spacing of and reasons for giving vocabulary tests 63
7.2.2 Areas of lexical competence tested 66
7.2.3 Types of test items 69

8 DISCUSSION 73

BIBLIOGRAPHY 77

APPENDICES 86
Appendix 1: Framework for the Interview 86
Appendix 2: Test item types 89
Appendix 3: Original Finnish interview quotes 91

1 INTRODUCTION

Currently up to a quarter of the world's population speaks English as a first language (L1), second language (L2) or studies it as a foreign language (FL) (Crystal 2012: 6). In Finland, by the 3rd school year 90% of all pupils study English as their first foreign language, and 99% of Upper Secondary School (USS; *lukio* in Finnish) students follow the advanced English syllabus (The Federation of Foreign Language Teachers in Finland SUKOL 2017).

Learning vocabulary is an essential part of language learning, and an extensive vocabulary is a prerequisite to fluent language performance (Hinkel 2006). In other words, unless a learner knows enough words in English, they will have difficulties in both expressing themselves and understanding what is being said or written. Indeed, one of the teachers interviewed for this study said that: "you can understand written text and spoken language even without a deep understanding of the grammar if you have a large vocabulary*"* (Saana).

Over 98% of English teachers in USS are formally qualified to teach English (Kumpulainen 2016: 69), meaning that they have a Master's degree which includes both pedagogical and English studies. Consequently, teachers can be expected to have a high expertise in their subject matter. The teachers, however, also have a high degree of autonomy in their teaching and are quite free to use any teaching and assessment methods they deem most efficient (Tirri 2014, Sahlberg 2007). Since teachers have a free rule over their teaching methods, great differences between teachers' teaching practices are to be expected.

While the system allows for great variation in methods, the National Core Curriculum (NCC) sets some boundaries to the content and objectives of the courses (Finnish National Board of Education (2016), FNBE henceforward). For example, for the second English course in the USS, NCC says: "The students enhance their ability to act as active interlocutors and language users who listen to what others have to say. They develop skills in forming opinions and negotiating meaning" *(*FNBE 2016: 117). As can be seen, these objectives are rather vague and only hint towards the kind of language the learners are to learn and practise during the course. In practise, the kinds of content that are taught are often dictated by two factors: firstly, by the course books the teacher chooses to use, and secondly, by the kinds of language skills assessed in the matriculation examination at the end of the three-year USS.

In this cross-fire of the responsibility brought about by high autonomy and the objective that students do well in the matriculation examination, teachers have to make daily decisions over which skills to practise and dedicate time to. Personal preferences and those aspects that teacher training programmes focus on are likely to be emphasized. However, having a large, functional vocabulary is an essential skill for any EFL learner to have. At face value, learning new words would seem simple enough: study them from a glossary and then practise them in use. Nevertheless, when learning vocabulary, there is almost endless information about each word to be learnt. While some aspects are fairly straightforward and relatively easy to learn, such as the meaning, orthography and pronunciation, others may take years of practise with and exposure to the language, such as the collocational and syntagmatic relationship-networks. Even more difficult may result how to ascertain that learners have acquired all the required knowledge. A common solution to this is giving short vocabulary tests that measure those aspects of lexical competence each teacher deems essential for the students. Much like with teaching methods, which testing method, if indeed any, a teacher uses is completely up to them.

As can be seen, teachers have a high degree of autonomy over their teaching and testing methods, and consequently the variation in the used methods is likely to be considerable. While there is a considerable body of extant research on language teaching in Finland, little attention has been paid to vocabulary knowledge or to vocabulary tests, despite their ubiquity. This study aims to shed some light on the kinds of lexical competence teachers focus on in their instruction in USS, and at the types of vocabulary tests they give. One of the aims is to see whether there are gaps in teachers' knowledge on vocabulary testing, and, consequently, to find possible areas of development in English teacher education in Finland. To do this, 9 teachers were interviewed on their teaching and testing practices, and transcriptions of the interviews were analysed through data-based content analysis.

Before the presentation and analysis of the data, the some of the approaches to vocabulary learning, teaching and testing will be looked at. Chapter 2 presents some of the different ways *word* has been tried to define, and especially how units that include several words are comprehended. Chapter 3 takes a closer look to what there is to learning and teaching words and vocabulary.

Chapter 4 presents different types of vocabulary tests and test items, and what makes a test valid and reliable. Furthermore, the state of vocabulary testing and teaching in Finland is also examined in the chapter.

Chapter 5 presents the research questions and aims of the study, while chapter 6 introduces the participants of the study, and the methodologies concerning data collection and analysis. Chapter 7 is divided into two parts; the first part presents and analyses how the interviewed teachers perceive vocabulary knowledge and how this is reflected in their teaching. The second half presents and analyses the types of vocabulary testing they do. Finally, chapter 8 concludes the study with a discussion of the results and suggestions for further studies.

2 VOCABULARY

2.1 Defining words

When defining *vocabulary*, common sense would say that vocabulary is the total sum of words a person has stored in their mind and that are retrievable from the mind when needed. While seeming like a good definition, it only brings about the next question, namely what a word is. *Word* is a surprisingly elusive term to define, as can be seen by looking at any textbook on the issue. Carter (2012), for example, dedicates a whole chapter for the definition of *word*, and several approaches to defining it can be listed: orthographical, phonetic, phonological and morphological. This chapter looks into how *word* has been defined in linguistic literature.

The most commonly used practical definition of a word is the orthographic definition, where each word is marked out by a space on either side, such as *hat, a, from, yesterday* and *football*. This is how we count words when writing an essay or playing games such as charades or Scrabble. This definition, however, brings about immediately a problem: is *lightbulb* one word when written together and two words when written separately *light bulb*? Compound words are problematic, because some of them are written without a space, others with a space, and some use a hyphen between the different elements of the word (such as *mother-in-law*), and there are no clear rules to why any of the options for spacing might be the conventionally accepted one. Indeed, as Saenger

(1997: 9–10) points out, historically numerous languages have used what is called *scriptura continua*, or writing without spaces, and also some modern languages, such as Chinese, are written in a continuous string. So, while the orthographic definition of a word may serve for everyday purposes, for scientific use it is too ambiguous.

The phonetic and the phonological definitions try to define words by their acoustic environment. The phonetic approach considers parts of speech separated by a pause words. Since speakers of a language rarely make the pauses in expected places in normal speech, as a look at a careful phonetic transcript reveals, this approach has only a limited use. The phonological approach tries to define words based on the sound-system of the language. For example, in English a word has typically one stressed syllable (such as *admire*, *admirable*, *admirability*). Some words, however, do not tend to receive any stress in normal speech (such as *and*, *at*, *in*, *or*, *his*) while in fixed expressions only one main stress occurs: *construction site*, *ballet dancer*, so this approach is not without its pitfalls either. (Singleton 2000: 7–8).

Word inflection and polysemy bring forth another set of problems: are *come* and *comes* different words, or are the different meanings of *foot* in *foot* of a stairs, *foot* of a mountain and *foot* of a table just one word? Is to *give up* just one word or two separate words? An attempt to solve these problems is the concept of *lemma*, which is the term used for the form of a word that can be found in a dictionary (Karlsson 2004: 187–188). *Come*, *comes, coming* and *came* are all represented by the lemma *come*, while lexeme is the meaning or the idea of the whole set of those words. If the meaning of words is related through a semantic field, as in foot of a page, mountain or a table, where foot always means ‘suggestive of a foot in position or function’, the words are considered polysemes. *Foot* can also mean ‘a unit of poetic meter’, where it is homonymous as the previous *foot*, or has the same spelling and pronunciation. But since this meaning of *foot* is semantically more distant while sharing the same etymology, the question whether these words are polysemes or merely homonymes is contestable. Regardless, defining word poses problems even for lexicographers, whose work it is to define words.

These are only some of the attempts to define *word* in scientific literature. For the purposes of this study, however, it is not enough. While the orthographic definition that word is *any unit delineated by blank spaces on both sides* has the problems described above, it is a useful tool in discussing the topic, and is adapted for the time being. A great part of the vocabulary of any language is likely to be made up of strings of several words that are bound to each other more or less tightly. The next chapter explores these units.

2.2 Multi-word units

Since this study is concerned with how vocabulary is being taught, we need to take a wider approach to what vocabulary is from the focus on single words presented in chapter 2.1. An approach to tackle the problems with the definition of a word as discussed in the previous chapter, Cruse (1986: 24-32) introduced the term "lexical unit" as opposed to a "word". A lexical unit needs to fill two criteria:

(i) a lexical unit must be at least one semantic constituent
(ii) a lexical unit must be at least one word

And to illustrate these, Cruse gives two examples:

1. the prefix *dis-* in *disobey* is not a lexical unit because, although it is a semantic constituent, it is smaller than a word.
2. the *pulled* in *Arthur pulled a fast one* is not a lexical unit because, although it is a word, it is not a semantic constituent, but rather an operator element of the semantic constituent "*pull a fast one"*.

There "word" is the smallest element of a sentence which has positional mobility, or which can be moved to another position without making the sentence ungrammatical. Semantic constituent, in turn, is part of a word, phrase or sentence that has an independent meaning in the context. While *light bulb* is both a word and a lexical unit, *give* and *up* are two words forming a lexical unit, and *come* and *comes* seem to be different forms of the same lexical unit. In *this will cook Arthur's goose,* the word *goose* is not a semantic constituent because it lacks an independent meaning in the context, much like the word *pulled* in the previous example. (Cruse 1986: 23-45).

Cruse's definition of a lexical unit does, however, still have its defects. The various meanings of the word *foot* are not addressed which is problematic, since, as Bogaards (2001: 326) points out, most words are polysemic, or they have several meanings, and consequently cover more than one lexical unit. Cruse’s lexical unit, however, has the strength that it is more readily applicable to languages that use different writing systems, such as Chinese, where word-boundaries are not marked.

As seen above, *vocabulary* includes units made up of two or more words that are not readily separable, such as *pull a fast one* in the example above. Singleton (2000: 47) calls these *lexical partnerships*, which he then divides further into different categories, such as collocations and fixed expressions. Wray (2002: 9) uses the term *formulaic sequence*, which is an all-encompassing term to include idioms, frozen phrases, routine formulae and myriad other terms that have been used in the scientific literature to refer to these multi-word units.

What formulaic sequences have in common is that they are usually made up of more than one word, and carry meaning beyond the face-value of the word, such as *give up* or *it’s raining cats and dogs*. On the other hand, the term also includes collocations and set phrases, where the meaning is the sum of its parts, but the words are more or less fixed to a certain order, or that there are certain other words that the words are frequently paired with: *salt and pepper* and not **pepper and salt* or *torrential rain* but not *torrential shower.* Wray (2009: 9) defines formulaic sequences as ”a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.” The advantage of formulaic sequences over other terms, such as Cruse’s lexical unit, is that it is easy to understand, and is not overly lengthy. A principle cited by Dixon and Aĭkhenval′d (2003: 9) states that “a definition should not be more difficult to understand than the word it purports to define”.

Schmitt (2004: 9) argues that formulaic sequences may also contain *semantic prosody* where certain constructions tend to appear with certain linguistic elements. To exemplify, Schmitt (2004: 9) shows how *bordering on* appears usually with a word that denotes an undesirable state of mind, as in *bordering on arrogance*, and consequently opposite meanings, such as *bordering on modesty* would be considered, if not incorrect, at least awkward or unnatural. As Wray (2002: 52) argues, what is nuclear about a word is not its meaning, but rather, its use. Therefore, the dictionary definition or translation of a word is only a part of vocabulary knowledge, though admittedly not an insignificant one.

Schmitt (2004: 26–27) adds, however, that the kind of semantic prosody on an abstract level described above cannot be included in the meaning of formulaic expressions but should rather be regarded as another aspect of the nature of a lexicon, but admits that due to the fluidity of the term each research study should define what they mean by formulaic sequences. In this study, therefore, I use the term formulaic sequence in the sense that Wray (2009: 9) defined it as "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar", adding to it Schmitt's (2004) *semantic prosody* to ensure fluidity of the term. This is important because formulaicity is a scale (Schmitt 2004: 26), where we would find fixed phrases such as "How do you do?" in the one end and semantic prosody in the other, and in order to avoid any artificial boundaries that might unnecessarily and unfoundedly limit the scope of the study all of them are included.

The term formulaic expression is a useful tool as it includes many widely different aspects of the collocational and contextual use of a word that would be otherwise difficult to sum up. Due to this wide informative load that they carry, formulaic expressions are commonly associated with *fluency* of speech (Nattinger and DeCarrico 1992: 32, Wray 2004: 249, Kuiper 2004) and presumably with other areas of language use such as writing, and as such might be expected to constitute an important element in EFL teaching.

Defining a *word* is a complicated issue. On the one hand we have to consider that the concept of a "word" may include everything from single word units to longer formulaic expressions, while on the other hand it is to be acknowledged that such definition fails meeting the standards that are expected from an academic text. For the purposes of this study, we assume that word can be made up of several discrete orthographical units, that come together to create a single meaning. Therefore, when the methods of teaching and assessing vocabulary knowledge are discussed in chapters 3 and 4, the term includes, besides orthographical words, longer formulaic sequences.

3 LEARNING AND TEACHING VOCABULARY

As seen in the previous chapter, defining what a word is proves challenging. Similarly, attempts to define what it means to know a word are fraught with complications. Words not only tend to be polysemous, but each of them also follows certain grammatical and morphological rules, has its own collocations and belongs to a certain register, among other characteristics. Section 3.1 explores how these aspects have been treated in linguistic literature, and also explains what knowing a word means in the context of this study.

Vocabulary learning[1] is essential, as it indicates later success in other language abilities, such as writing, reading, speaking and listening (Zhong 2012: 23). Indeed, without a considerable vocabulary a competent performance in all L2 skills is not possible (Hinkel 2006). A solid lexical base is central for learners' ability to infer and process L2 vocabulary (Albrechtsen, Haastrup and Henriksen 2008: 200). However, as Haastrup and Henriksen (2000: 222) point out, no model covers all aspects of L2 vocabulary acquisition, and as a consequence some inferring and adopting has to be done from related fields, especially the organisation of L1 mental lexicon and the more general second language acquisition theory.

Because of the reasons mentioned above, teachers must have the tools for teaching vocabulary and for assessing learners' development. The teacher's role is particularly important because, as Meara (1982: 30) states, even if the learners themselves claim to know the words, identifying them reliably is a challenge for many learners. Furthermore, Singleton (1999: 144–145) points out that it seems that L2 learners have difficulties learning synonyms for L2 words, and that learning new words for the same idea is seen as "a waste of time and effort". Carter (2012:190) regrets that when there are books giving practical advice to vocabulary teaching, they usually are not concerned with how words are actually learned, and the proposed teaching strategies are based on "at best, untested assumptions". Sections from 3.2 to 3.5 attempt to compile the research on some of the aspects vocabulary teaching, while vocabulary assessment is discussed in chapter 4.

---

[1] Both learning and acquisition can have various meanings, but for the purposes of this study, the terms are used interchangeably to mean *to become informed of, to acquire knowledge of or skill in by study, instruction, or experience*.

In the early 2000s, Bogaards (2001: 322) criticised that a lot of research on vocabulary learning either focus on a narrow meaning of vocabulary, namely single words or compounds, or scientists do not define what they mean by a word clearly. He found especially problematic that for researchers vocabulary acquisition means learning new words, while ignoring the new meanings for known words. Since then, however, there has been extensive research on formulaic language, and some of it has also been focused on pedagogical aspects. Chapter 3.5 focuses on this area of language research.

3.1 Types of vocabulary knowledge

Nation (2008) argues that before teaching a language, the teacher should in some way assess the learners' vocabulary knowledge. This section presents the most widely used theories of what vocabulary knowledge actually is, and in chapter 4, different types of language tests are discussed.

In a study that focuses on vocabulary tests that teachers give to students, it is important to address what it means to know a word. A whole body of research and theories address this issue, with each researcher having invented a new term or given an existing term a new definition, some of which are discussed here in an attempt to create if not a synthesis, at least outline some of the ideas that are concerned with vocabulary knowledge. Three different groupings arise which are not necessarily mutually exclusive and are even often included within each other: 1. vocabulary knowledge as having different aspects. 2. vocabulary knowledge as an interconnected network and 3. receptive vs. productive language knowledge.

For Wallace (1982: 27), knowing a word involves as many as nine different aspects: recognizing and recalling it, relating it to a concept, using it in a grammatically appropriate way, spelling and pronouncing it correctly, using it in the right collocation and in the appropriate register, and knowing its connotations. Singleton (2000: 161) lists five aspects to knowing a word, which synthetize those of Wallace: pronunciation, orthography, meaning, morphological behaviour and syntactic behaviour. Influential is Nation's (2001) proposal that vocabulary knowledge has three aspects to it: form, meaning and use, which can be broken down to their constituents: pronunciation, spelling and word parts; meaning, associations and grammar; collocations, register and frequency. Whichever way vocabulary knowledge is partitioned, it is clear that it entails many different aspects and layers.

Anderson and Freebody (1981: 92) first made the commonly used distinction of the *breadth* of vocabulary knowledge and the *depth* of vocabulary knowledge. They defined breadth as how many words a speaker knows and depth as how well those words are known. While the breadth (also sometimes called the *width*) of vocabulary knowledge is fairly straightforwardly the number of words known by a person, however difficult to measure, the depth or how well words are known is a much more complicated matter, and consequently there has been a great number of propositions as to how to define it. Some of those proposals are explored below.

Perfetti (2007: 359) introduced the term *lexical quality* to study the depth of vocabulary knowledge. First, *lexical knowledge* consists of three aspects that are very similar to those of Nation: knowledge about forms, meanings and use, where use is the meaning of a word together with its pragmatic features. Then, lexical quality is the learner's grasp of those three components and how well the aspects of word knowledge are connected through bindings. A learner's mental vocabulary includes words that have very high lexical quality and others that have low lexical quality: those with high quality have a *tightly bound* orthography, phonology, grammar and meaning, and those with low quality have incomplete information. Perfetti's theory's weakness is its focus on receptive written language, and consequently it is too narrow for the purposes of this study. As Li and Kirby (2015: 613) say, "depth of vocabulary can be seen to involve at least precision and multiplicity of meaning, collocational use of words, and morphological knowledge about word structure." Therefore, a broader approach is needed to understand the depth of vocabulary knowledge.

A common view is to see the depth of vocabulary knowledge as a network of interconnected words, or as Aitchison (1994: 72) rather colourfully puts it, "a gigantic multi-dimensional cobweb". Aitchison (1994: 63) also says, "knowing the meaning of a word involves being able to relate it to other words. A full understanding of the meaning of many words requires knowledge of words which are found with it and related to it." In other words, it is not enough to know just separate words, but a learner has to also have knowledge of the words that appear in close contact with it or are related to it through meaning or have derivational kinship. This is especially important in the light of Henriksen's (Albrechtsen et al. 2008: 25) argument that words are not stored randomly in a learner's mind, but rather they form a network, and when language is being used the lexical items are accessed through this network. This would mean that words that are poorly networked will be correspondingly more difficult to be accessed when using a language, implying that network building is imperative in learning vocabulary.

Meara and Wolter (2004: 89) abandon the terms breadth and depth, calling them "unfortunate", and employing terms *size* and *organisation* instead, although how they differ from breadth and depth is not made entirely clear. Meara and Wolter hold, however, that L2 lexicons are less well-structured than those of L1 speakers, whose networks are denser and more highly organised. The less organised nature of the networks might be a cause to the phenomenon described by Meara and Wolter (p. 90) that L2 speakers' associations to any given word are much wider than those produced by L1 speakers; since the network is less well-organized, the connections are made with nodes further apart.

The lexicon network of L1 may help a great deal in the acquisition of L2, as it the existing network may work as a building block, and may be applicable to L2 network with only minimal adjustment (Singleton 1999: 80). Problems arise when the network of L1 and L2 vocabularies differ due to cultural dissimilarities between the language communities. The learner has not only to learn a new network, but they also must unlearn the pre-existing network from their L1. For that to happen, they have to become aware of the different networks, and undoubtedly the teacher plays an important role in this process.

Henriksen (1999: 304–307) distinguishes three interrelated dimensions of vocabulary knowledge. The first is the *from partial to precise knowledge dimension*, where the learner gradually learns the meaning of a word, or its translation to L1. The second is called the *depth of vocabulary knowledge dimension*, where, besides the meaning of the word, the depth involves the "rich meaning representation of the word" (p. 305). This rich meaning representation refers to the relations words have to other words, these being the "paradigmatic (antonymy, synonymy, hyponymy, gradation) and syntagmatic relations (collocational restrictions)" (p. 305–306) as well as the syntactic and morphological restrictions that the lexical unit may have. The third is the *receptive-productive dimension*, as the productive abilities of a learner can be vastly different to their semantic knowledge of a word. To Henriksen acquiring a word is a combination of the first two dimensions, and this she calls the *semantization process*, where the learner goes through both of these dimensions simultaneously, constructing the lexical patchwork to all directions at once. Zhong (2012: 29–30) concludes that the first two dimensions in Henriksen's model are basis for comprehension or mastery of vocabulary knowledge. The comprehension enables the learner to use language appropriately, while the performance itself is the receptive-productive dimension (see Figure 1). Henriksen (1999: 309) suggests that L2 vocabulary acquisition research focuses mainly on students learning the form or the first dimension of vocabulary knowledge. In my view this is true also to language teaching, where teachers mostly focus on teaching and students focus mostly on learning the meanings of words but largely neglecting the rich meaning representation or the mastery of said words.

Hulstijn (2012: 3) points out that a lexical entry in a person's mental lexicon contains both semantic and formal features, and links to other words, which are "intrinsically or associatively related to each other" and to their counterparts in other languages the speaker knows. The relationships between the different entries in this mental lexicon resemble to great extent the networks described above, especially that of Singleton (1999). Hulstijn adds that the relationships between lexical entries can be of varying strength. Applying it to Henriksen's (1999) model, weak relationships would be those on the from partial to precise knowledge dimension, while stronger relationships have acquired a rich meaning representation.

Zhong (2012: 24), like Henriksen, divides vocabulary knowledge into receptive and productive vocabulary, but she points out that studies measuring the receptive/productive aspect have been mainly concerned with the breadth of vocabulary. The division into receptive and productive language is widely held, and for example the Common European Framework of Reference for languages assesses receptive and productive skills separately (Council of Europe 2001). Among the researchers there are, however, slightly differing views of what those skills exactly are (Zhong 2012: 28). In this study, Nation's (1990) definitions are adopted as they are succinct and feel instinctively right. Nation defines receptive vocabulary skills as the ability to perceive words from written or spoken input, and retrieve their meanings from memory, while productive skills is the capacity to retrieve and produce a word to express a meaning in written or oral output.

Basing on the literature reviewed, Zhong's (2012: 30) hypothesis for the relations between dimensions of vocabulary knowledge (Figure 1) is adopted with slight elaboration. The depth dimension and the partial-precise dimension interact through a complex network, where the relationships between L2 words and the different aspects of their meaning, form and use together with their L1 correspondents get reinforced or weakened in a continuous process. New occurrences of words activate the process, and the teacher through their teaching methods is an active player in affecting which nodes are consolidated and which get weakened.

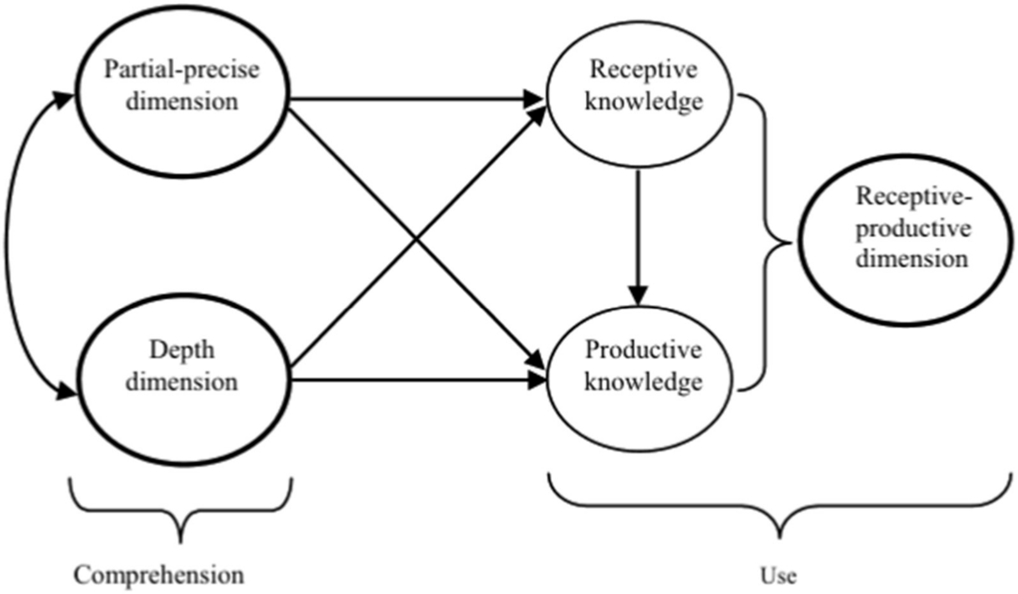

Figure 1. Dimensions of vocabulary knowledge (Zhong 2012)

In the face of all the previously mentioned terminology and ways of perceiving vocabulary knowledge, the term *lexical competence* will be used in this study. While this term itself carries its own historical baggage, such as Meara’s (1996: 3) statement that “the basic dimension of lexical competence is size”, in this study it will include all the different aspects and dimensions of vocabulary knowledge. In other words, it will include the meaning-form-use network and the receptive-productive dimensions shown in Figure 1. Here, like before, I have chosen a sufficiently wide definition so that the different approaches that may arise in the data can all be analysed within the framework of the study.

As can be seen, *knowing* a word involves so many aspects, and so many elaborate processes are involved in it, that it may seem overwhelming, if not impossible, for the teacher to teach all there is. Section 3.2 explores some of the methods which a teacher may utilize to advance the learners’ lexical competence.

3.2 Incidental and intentional learning

This chapter explores the different traditions on vocabulary learning. Vocabulary learning is often divided into two categories: incidental and intentional learning, which in the field of vocabulary learning are also called implicit and explicit learning, accordingly (Hulstijn 2012: 2). Hulstijn points out that the words are not wholly synonymous, but as they are commonly used interchangeably in literature, in this study they are treated thus.

Implicit and explicit vocabulary learning are analogous with implicit and explicit language knowledge more generally. As Ellis (2010: 35) puts it, implicit knowledge is “extracted from experience of usage”, so that a child learns L1 not by being dictated its grammatical rules, but rather from a constant exposure and from a feedback in communicative context. Similarly, incidental language learning is defined as a by-product of reading or listening, where the mental processes are not focused on learning new vocabulary but rather on understanding what is being said (Huckin and Coady 1999: 182). It is commonly accepted that native speakers learn most words incrementally, or through repeated exposure to a word, and incidentally (Hulstijn 2012: 3), and the learning process carries on through adulthood (Singleton 2000: 170). It is questionable if this is true also to foreign language (FL) learners, because there is usually significantly less exposure to FL than there is to L1.

Another aspect to consider is the learners’ age. As USS students are typically aged 16-19, the learning profile may be very different of that of children learning their L1. A study by Schmitt and Zimmerman (2002: 149) illustrates the different processes involved in L1 learning and adult FL learning. They showed that morphological derivations in L1 are learned incrementally, but in FL lexicon the words often fossilize before the stage is achieved where derivations are usually learned. In other words, even if FL learners might learn the derivatives incrementally, some language forms have become so established in their internal lexicon that improvement may prove if not impossible, at least difficult. While the results cannot be generalised to all areas vocabulary knowledge, the study demonstrates that FL learning process that takes place after childhood differs significantly from that of L1, and that incremental learning may bear less importance in FL learning in general than it does in L1 learning.

Krashen (1989) famously postulated that most of vocabulary learning is incidental and happens best through extensive reading. Krashen (1989: 455) suggests therefore that reading books should be favoured instead of vocabulary and spelling tests, because "teachers will teach to the exam, and students will study for it". In other words, the only motivation that both the teacher and the students have is that they perform well in the test, instead of studying so that they know how to use the language. Elley (1997) supports this, and reports several studies conducted on elementary school students across the Pacific that consistently show that extensive reading is an efficient method of L2 vocabulary learning, especially if the meaning of the words were explained by the teacher and there was repeated exposure to the lexical items over the span of days or weeks. As to more advanced students, Carter (2012: 194) posits, "beyond a certain level of proficiency […] vocabulary development is more likely to be mainly implicit or incidental". This means that highly advanced learners merit little from special vocabulary instruction, and instead learn most of their new lexical items through reading.

There are, however, limitations to the learning via extensive reading. Firstly, it can be slow, only a few words are remembered (Laufer 2003, Horst 2005). Secondly, the learners need to have a rich vocabulary to begin with: an optimal text is such that the reader already knows at least 98% of the vocabulary at the onset (Nation 2001: 233). Thirdly, as Sökmen (1997: 238) points out, inferring a meaning from context[2] is risky, as the meaning of the word is often guessed wrongly. Fourthly, as shown in a study by Elley (1989), if the text is uninteresting to the students or it is culturally too distant to them, the number of words learned plummet. These restrictions mean that the gains can be slow, the learning outcome is uncertain, and the method is most suitable to students that already have large vocabularies. Due to these limitations, incidental vocabulary learning might not be relevant to Finnish scholastic settings, as this level of proficiency is very unlikely to be obtained by more than a small minority of the students.

In my view, there is even a larger problem with learning vocabulary through reading: it completely ignores spoken situations. After all, written language and books in particular form only a small part of language use and input. That students may be able to define a word does not equate to being able to use it either in spoken or written communication, which for me *is* the primary use of a language. Nevertheless, reading is important, and as Krashen (1989: 455) says, encouraging students to read on their free time regardless of the level of the texts is advantageous for L2 learning, making the reading of more difficult texts also easier.

---

[2] Sometimes a difference between *context* as "the real-life situation where language is used" and *co-text* as "the other words that the target word appears with" is made. In this study, context includes both meanings.

Schmitt (2008: 330–331) discusses the vocabulary needed to understand spoken discourse. While the percentage that needs to be comprehended is lower than that of written text, it is still above 90%, although the studies cited by Schmitt are concerned with *listening comprehension* and not with real communicative situations, where negotiation would likely lower this figure. It is still clear from these figures, however, that a good command of vocabulary is imperative for adequate comprehension of spoken language. This was further demonstrated in a study by von Zeeland and Schmitt (2013), where similar results were produced.

So far, I have been discussing incidental learning. Opposed to it, intentional learning, as stated above, is learning where learners are knowingly trying to learn new words. Intentional vocabulary learning methods include as varied approaches as reading words from a list or having fairly free conversations on set topics with the aid of a related word-list. Many studies demonstrate the advantages of intentional vocabulary learning; Norris and Ortega (2000) found in their meta-analysis of 49 studies that explicit instruction settings were systematically more efficacious than implicit settings. Nation and Webb (2011: 307) went as far as to state that learning vocabulary intentionally causes both faster learning and more vocabulary being learned “without exception”.

When learning a new language, the learner may be highly dependent on the teacher for learning new words, indeed, Nation and Webb (2011: 1) state that the “most important job of the vocabulary teacher is to secure that students are focusing on the vocabulary most relevant to them and that there are enough learning opportunities”. This may be even more relevant for students learning L2 in L1 environment, or as Lin and Hirsh (2012: 119) put it, where “exposure is minimal”, as they are not exposed to as much L2 vocabulary outside the formal learning settings as learners living in L2 environment. Although it could be argued that a Finnish learner of English might be exposed to a great deal of English in Finland, it is unlikely to be as immersive as living in the L2 environment.

The role of the teacher is further highlighted by Zhong and Hirsh (2009), who suggest that the patterns of vocabulary development are influenced by the types of vocabulary tasks the students are given. If the learners are being taught mainly receptive vocabulary knowledge, they will mainly develop their receptive skills. Consequently, teachers should use as many methods of teaching new vocabulary as possible.

In contrast with the theories explored above, Hulstijn (2012: 4) asserts that neither incidental nor intentional vocabulary learning is necessarily better than the other, but whether vocabulary learning takes place depends on other factors, such as making the students look up the meaning of unknown words. Hulstijn (2012: 3–4) also argues that for people with large vocabularies it is irrelevant whether the words have been acquired incidentally or intentionally. It is more important, Hulstijn says, "*how* one processes information" than if the person is consciously trying to memorize the word. That is the reason why people do not remember everything they have ever been exposed to, but rather only things that go through elaborate memory processes are remembered (Eysenck 1982: 203). Additionally, new information has to be connected to pre-existing knowledge (Baddeley 1997: 123), and to retain it, the information, in this case vocabulary, has to be repeated frequently. The effect of repetition on vocabulary retention is discussed in chapter 3.3.

Others have tried to unite the two seemingly opposing aspects of vocabulary learning. Kirsner (1994) hypothesised that explicit instruction may be a required to create a link between the form and the meaning of a new word before implicit learning is possible. Intuitively, this would feel to be the case especially for in the lower language proficiency levels. Basing on important advances in psychology and neurosciences, Ellis (2010) argues that in L2 language learning explicit and implicit knowledge are in a dynamic process, where a form, in this case a word or a lexical unit, is first learned explicitly, but it then passes to the implicit cognition. This highlights that even though much of our language knowledge and skill is automatic and implicit, in L2 it is highly advantageous to first learn the contents intentionally.

3.3 Repetition and learning

Words, whether learned incidentally or intentionally, are seldom learned only after a single exposure to the word. Usually the nature of vocabulary learning is incremental, or the words are learned gradually and consciously after repeated exposure (Chacon-Bertrand 2010; Frishkoff, Perfetti, and Collins-Thompson 2011). Once a word is learned, it is not certain that it will be remembered permanently. Indeed, Baddeley (1997: 123) states that our memory works so that frequent "maintenance rehearsal" is needed for a word to stay in our minds. This section discusses how repeated encounter and use of the word reinforces the attainment of a word. In the context of this study, repetition plays an important role, since teachers' goal in vocabulary tests is to ensure that learners are faced with the vocabulary again, both while studying for the test and while doing the test. There are three aspects of word repetition that influence retention: number of repetitions, the spacing of repetitions and the type of repetitions.

There is no consensus on the number of encounters with a word needed in order for a word to be remembered, but both Schmitt (2008: 348) and Matsuoka and Hirsh (2010: 57) synthetize this number to be around 10. Naturally some words are learned after fewer repetitions while others require a considerably higher number, but it seems that 10 encounters is enough so that in most cases the word is remembered. 10 exposures (either receptive or productive) to a word might not seem much, but it is questionable whether even this low figure is reached in Finnish EFL classes with other than the most common words.

The number of exposures to a word is not, however, the only aspect of repetition that affects word attainment. It is also influenced by the spacing of the repetitions. When the time dedicated to studying is continuous, or is without interruptions, it is said to be *massed* (Cepeda et al. 2006: 354). When, however, there are intervals in the studying process, it is called *spaced* or *distributed* (ibid.). In the classroom context, this would mean that each lesson constitutes one massed event, and separate lessons over a week or a semester are spaced learning events. A meta-analysis by Cepeda et al. (2006) on studies that measured the effects of distributed practice on verbal learning shows that for adults distributed practice is invariably beneficial for retainment. This means that repeated encounters with the word over the period of several days or weeks helps learning new words without exception. Due to this effect, vocabulary testing is propitious for vocabulary learning as students ideally are exposed to the words at least three, spaced, times in the process: when studying for the test, while doing the test, and when receiving feedback on the test. Few studies have been conducted on the effects of spacing on children or adolescents, but they seem to show a similar trend (Seabrook, Brown and Solity 2005; Sobel, Cepeda and Kapler 2011).

In a study by Rossiter, Abbott and Kushnir (2016), teachers reported assessing student vocabulary on average in every seventh class, or almost every week. The majority of the teachers in their study, however, only tested vocabulary that had been learned recently, even though for optimal vocabulary retention words should be retested, and thus revisited, frequently.

Besides the number and spacing of the repetitions, the type of repetition is important. Eysenck (1982) states that words that go through elaborate mental processes are remembered better, so a mere repetition of the word may not be enough. Similarly, Baddeley (1997:121) points out that the more ways a word is encoded into the memory, the more likely is the retainment, especially if this coding is semantic (as opposed to visual or phonetic). In other words, the more nodes is created in the depth dimension network discussed in section 3.1, the more likely the learner is to remember the word. These repetitions should not, however, be too similar, as Elgort et al. (2015: 506–507) explain: repetition from extensive reading or varied contexts helps to abstract the meaning of the word. In other words, if the word is given always in the same context, the student will have difficulties in creating a rich meaning representation for the word. Reading novels or shorter, thematically related texts provide favourable repetitions of new words (Matsuoka and Hirsh: 2010), that is to say occurrences of a word in a similar context and use, but with enough variation as to overcome the previously mentioned problems.

While extensive reading seems to be a functioning method for learning vocabulary, exactly how much vocabulary can be learned through reading alone depends on various other factors such as how well the texts are prepared and on the learners' initial vocabulary (Webb and Chang 2015). Even if extensive reading worked always, it would be difficult if not impossible to do this in most Finnish EFL classrooms due to limitations of time and resources. In 2014 a study by the Finnish Ministry of Education (2014: 122) showed that for the majority of secondary school EFL teachers the course book is the most important tool for teaching. As a result, texts in course books constitute the major part of texts that are read in Finnish EFL classes. EFL course books invariably contain short texts on a varying range of topics with little repetition of vocabulary, so the favourable conditions are not met. However, Paribakht and Wesche (1996) demonstrated that if students have to practise the words after reading, the level of vocabulary retention is significantly higher as opposed to only reading the words in a text. Since classroom teaching typically provides multiple instances of practising the words after reading a text, this lack of repetition from extensive reading may well be compensated by the activities.

A lot of the research on word repetition emphasises the role of extensive reading and incremental vocabulary learning as words are often repeated in long texts (Cepeda et al. 2006; Schmitt 2008; Matsuoka and Hirsh 2010), but intentional vocabulary learning gains equally from spaced repetition. For example, a study by Zechmeister and Shaughnessy (1980) showed that spaced learning has positive effects on learning from word lists. Bilingual word lists are ubiquitous in Finnish EFL classrooms, as a glance at nearly any course book shows, and consequently the study is particularly relevant for the Finnish context. Even though the study was conducted with college-level native English-speakers, and thus the results may not be generalisable to L2 learners, Zachmeister and Shaughnessy made an interesting finding, namely that learners are fairly accurate at predicting which words they will be able to recall later on. Knowing this, teachers could exploit the information, and for instance instruct their students to pay more attention to the words they believe they are not yet capable of recalling.

3.4 Language of instruction

When a teacher chooses the language of instruction, there are three main issues to consider. Firstly, the use of L1 may not be possible due to there being several different L1 in the class. In Finland this has not been a major issue until recently due to little immigration, but this will continue being a growing challenge. Secondly, the teacher may not feel confident enough in the language and consequently may opt out from teaching in L2. Thirdly, the pedagogical advantages and disadvantages, if any exist, of using L1 have to be weighed. This section will explore those advantages and disadvantages.

Schmitt (2008: 337) asserts that using L1 in language teaching is unpopular among many teachers, since L1 will inevitably always influence L2 learning. Therefore, teachers may shun away from using L1 in order to minimize that influence. However, L1 should be used to our advantage when possible. Especially compelling seems Schmitt's (2008: 337) argument that psycholinguistic studies show L1 to be active in L2 lexical processing even in advanced learners. If this indeed is the case, L1 seems to be present whether the teacher wants it or not, and consequently it should be considered as a tool for teaching rather than a hindrance.

Using L1 in L2 instruction can even be highly advantageous, since "the more opportunities that can be found for formal transfer between foreign and mother-tongue, the better the chances of retention" (Carter 2012: 188). This seems to tie logically both to the multi-dimensional idea of vocabulary knowledge (section 3.2) and to the positive effect of repetition on learning (section 3.3). Carter (2012: 196) specifies that especially on lower levels bilingual word lists and translation tasks can be a valuable tool in learning new vocabulary. They should naturally be used together with other pedagogical methods to ensure ample possibilities for retention.

Carter (2012: 190) says that learning words does not mean merely connecting words to their L1 meaning, but on the other hand, when forming the initial meaning-link with a L2 vocabulary item, it seems that using L1 is the most effective method. Jiang (2004: 104) goes as far as to suggest that learners will create a connection to a L1 equivalent regardless of the teaching strategies used. The results from a number of studies that especially for less proficient learners using L1 translations increases the word retention significantly (Schmitt 2008: 337; Carter 2012: 196) seem to support this idea. After the meaning-link is formed and the learners are familiar with the word, using L2 for the more contextualised types of word knowledge seems to be more beneficial (Schmitt 2008: 337).

Different languages may also play different roles in language learning. Read (2000: 167) states that translations from L2 to L1 measures learners' receptive skills while L1 to L2 translations assess their productive skills. Even if the distinction were not that clear-cut, the fact that different languages play a different role in language learning is an important notion. Read compares giving monolingual tests to using monolingual dictionaries, and monolingual teaching in general has the same problems. Firstly, the most common words are nearly impossible to explain using words of similar frequency range. This could be circumvented with the use of images, real-life objects or gestures, but in USS the words are typically so abstract or specialised that this method seems unlikely to be useful. Secondly, learners may be unfamiliar with or they might have forgotten the L2 words that are used to explain the new lexical items, creating a further obstacle to learning. Also, if students create a link to the L1 counterpart irrespectively, as Jiang (2004: 104) suggests, the purposeful avoidance of L1 may be unnecessary when the teacher and the learners share the same language.

3.5 Teaching formulaic language

No word is an island. Whenever words are used in contact with other words, they follow certain grammatical rules. This is crystallised by Stubbs (1993: 17) “Every sense or meaning of a word has its own grammar: each meaning is associated with a distinct formal patterning”. When students are taught vocabulary, the words carry with them grammar that has to be taught with them. And as stated in section 2.2, most words are polysemic, and therefore the load that must be taught can get overwhelming to both to teacher and learners. Formulaic sequences may facilitate this process, as the expressions are learned and stored holistically (Wray, 2002).

During the last 15 years, it has become more and more obvious that formulaic language teaching should be an essential part of language teaching for three reasons: firstly, formulaic sequences are ubiquitous in language, and have even been called “the very centre of language acquisition” (Nattinger and DeCarrico 1992: xv). Some estimates say that in English “up to 50% of the language is formulaic in some way” (Coxhead 2018: 113). Secondly, how well a learner uses formulaic language corresponds with their level of language fluency, and thirdly, learners find formulaic language challenging because they lack the intuitive resources for formulaicity that native speakers possess (Meunier 2012: 112, Martinez and Schmitt: 2012). For those reasons teachers are responsible for providing the students with exposure to and practice with formulaic sequences.

While FSs are a crucial part of lexical knowledge, a lot of EFL textbooks, have been shown to not reflect the formulaic language that can be found in language corpora (Meunier 2012: 114). This seems to hold true also to Finnish EFL textbooks, which focus heavily on idioms and as a result ignore lexical phrases, as shown by Ylisirniö (2012). Ylisirniö’s study did not, however, include collocations or phrasal verbs other than incidentally, and again, looking at modern USS English course books, whole lists of common phrasal verbs can be found in almost any book. Nevertheless, when it comes to formulaic sequences in general, the disparity between course book language and corpus data appears to be real. Consequently, if a teacher relies heavily on the course book, it is likely that FSs are underrepresented in their teaching.

To balance the absence of FSs in EFL course books, teachers need to look for material elsewhere. Opportunely, recent years have seen the publication of some formulaic sequence lists for the use of researchers and teachers alike, the most important of which might be the Phrasal Expressions List by Martinez and Schmitt (2012) and the Academic Formulas List by Simpson-Vlach and Ellis (2010). How well these lists have pervaded EFL classes is open to question, but they exist there for the interested: claiming that there are no practical resources for teachers on formulaic language does not hold water anymore.

Despite the attention formulaic sequences have received, research on how formulaic language should be taught is still scarce (Coxhead 2018: 114) and may not be directly applicable to the Finnish classroom context. Few comments can be made on the subject, however. As seen in sections 3.1 and 3.3, learning new vocabulary is a slow process, and it gains from explicit teaching. Alali and Schmitt (2012: 155) assert that this must be true also to formulaic sequences. Research also suggests that formulaic language teaching may benefit from using specialised methods. Extensive exposition combined with making the learners aware of the FSs present in texts improves perceived fluency of expression (Schmitt 2008: 340). Also, when formulaic language is taught cognitively, learning outcomes are improved considerably (Alali and Schmitt 2012: 172). This means making the students analyse for example the function of the word *off* in such idioms as *take off* and *see off*. On the other hand, Wray (2002: 7) argues that native speakers' mental processes treat formulaic language as single unit rather than piecing it to its constituent parts. How much L2 learners follow the patterns of native speakers in uncertain, and anecdotal evidence from my personal experience as a teacher would suggest that learners are keen to analyse even very idiomatic expressions, such as *hold your horses*. Considering the argument brought forward by Alali and Schmitt, this learners' analytical response to FSs may in fact enhance the learning process.

Overall, formulaic sequences are crucial for the (perceived) fluidity of language and they form a large part of the language. There is little research on how they are best taught, but the most efficient methods may differ from those used in teaching single vocabulary items. Some resources exist for teachers to use, but the application and use of those resources is up to the teachers' knowledge, motivation and time.

4 VOCABULARY TESTS AND VOCABULARY IN FINNISH SCHOOLS

Testing is ubiquitous in school life because it is foremostly a tool for assessing students' performance, or as Huhta (2017) concisely put it, "testing is assessment with a test-like measuring instrument". When vocabulary is tested systematically, it measures cumulative vocabulary learning, usually on regular intervals (Rossiter et al. 2016: 7). Regular testing has the advantage that it incites students to review the vocabulary, whose benefits were discussed in section 3.3. Regular, smaller tests also have the advantage that students are less likely to cram when the contents of the course are tested on cumulatively. Besides being an useful tool for assessment, frequent testing also provides the teacher with information on the learners' vocabulary learning, and as such is a tool of monitoring both the teaching efficiency and learners' development.

When revising for the test, learners will encounter the words again which is shown to reinforce word retention (see section 3.3). Testing may also create an effect on both learners and teachers where they will focus on certain parts of syllabus that otherwise might be neglected (Wall 2012: 79). Wall (2012: 79) argues that this is this case mostly in high-stake testing or testing that affects the final grade significantly, which is not necessarily the case for a lot of vocabulary testing, but Nation (2001: 375) points out that regular vocabulary testing may have real, long-term positive washback effects.

Admittedly, the effect is not straightforward and it does not necessarily take place every time or with all learners. Furthermore, other circumstantial factors play an important role on how much learners study for a test (Wall 2012: 84) and the effects differ from a learner to learner, and also the same test may have a different effect on the learner on different dates (Wall 2012: 86-87). There are those who challenge vocabulary testing altogether, such as Hughes (2003: 179, 184), who asserts that giving too much importance to vocabulary (and grammar) testing may create a negative backwash effect, in particular if language teaching focuses on communication since the tests conflict with this and students see good performance in tests as a goal in itself.

Hughes (2003: 27) argues that those areas of lexical competence that are not tested will be ignored in teaching. For this reason it is important that teachers test the skills they deem essential for the learners to have, and that they are aware of which aspects their tests focus on. There are a number of different ends to which vocabulary tests are given: section 4.1.1 introduces these, while section 4.1.2 shows the different types of test items that may appear in vocabulary testing. Wall (2012: 84) postulates that bad test design may cause negative washback and good test design positive. Section 4.2 discusses this from the perspective of test reliability and validity. Finally, section 4.3 presents how vocabulary is being tested and taught in Finnish USS.

4.1 Types of tests and test items

4.1.1 Types of tests

Vocabulary tests are a very old method of language assessing and teaching, and consequently a myriad of different types of tests exist, and they have been classified in a number of ways. This chapter does not claim to include all of them, but rather is an overview of some of the most commonly used test types both in practise and in research on vocabulary learning. There are two qualities according to which vocabulary tests may be categorised: the content and format of the test, and the purpose of the test. After these have been discussed, different vocabulary testing methods will be discussed.

Read (2000: 8–9) divides vocabulary tests into having three dimensions based on what is being asked: *discrete* and *embedded; selective* and *comprehensive;* and *context-independent* and *context-dependent*. In discrete tests vocabulary knowledge is measured as an independent ability, while in embedded tests it is only a part of a larger skill set, such as “reading-comprehension ability”. Read (2012: 308) argues that most vocabulary testing is discrete in their nature, because tests usually focus on pre-selected words, and they focus on specific aspects. Selective tests measure the use of certain, pre-determined words, and contrastingly comprehensive tests measure the overall fluidity and range of expressions used in a spoken or written test. In context-independent tests vocabulary knowledge is measured independently from the structures of the language, whereas context-dependent tests place the words in whole texts or discourses.

Vocabulary tests can be divided into four main types depending on their purpose, as listed by Nation (2001: 373). Firstly, a vocabulary test can measure the learners' *proficiency*, or how much vocabulary they know. Secondly, a test may be used for *diagnosis* to see where students face difficulties. Thirdly, it can be a *long-term achievement test*, whose aim is to see whether the goals of a large segment, such as a whole course, have been achieved. Fourthly, the test may measure *short-term achievement*, where the focus is on a recent group of words, such as the vocabulary of a single unit in a course book.

A number of tests have been devised to test the learners' language proficiency. Well-known examples of this kind of tests are TOEFL and IELTS. These do not, however, measure only the learners' vocabulary breadth and depth, but they are also concerned with, for example, the learners' command of grammar and communicative skills. In the Finnish context YKI or the Finnish National Certificate of Language Proficiency is a proficiency test that measures functional and communicative language use, but also pays attention to vocabulary and grammar (FNBE 2011: 8, 15). Laufer and Nation's (1995) Lexical Frequency Profile (LFP), is another typical example of a vocabulary proficiency test. It aims at measuring the proportions of words from different vocabulary frequency levels used in learners' writing. This, they claim, is a good indicator of lexical richness or the breadth of vocabulary and is used to assess learners' vocabulary proficiency.

An example of a vocabulary test used for diagnosis is the Vocabulary Levels Test devised by Nation (1990) and modified by Schmitt, Schmitt and Clapham (2001). It provides information on the general vocabulary size (or the breadth of vocabulary) of the learner, and in particular on how well the learner knows high-frequency words, and is used to discover a learner's vocabulary level. Even though Nation (2001: 373) states that it "*is a diagnostic test*", I would argue that it can be either a proficiency test or a diagnostic test, depending on the intended use. When used as a diagnostic test, Nation suggests, the teacher may create a vocabulary learning programme suited for the learners' needs based on the results. In any case, tests that measure the breadth of vocabulary are mainly useful in diagnosing or measuring proficiency as they aim at creating a general profile rather than measuring achievement in a certain time period.

Long-term achievement tests are an integral part of nearly every course in the Finnish upper secondary school, and also matriculation examinations that students sit at the end of the USS assess how well learners have assimilated the broad contents dictated by the syllabus. Similarly, *unit tests* are a common tool of assessment in Finnish language classes (Huhta 2017). A unit test is concerned with either the vocabulary or the structures or both of a certain unit of a course book, and as such is a good example of a *short-term achievement test*. This study is concerned with short-term tests, so the subsequent test types are examined from the point of view of their suitability to this kind of testing. Since tests that measure breadth of vocabulary are typically proficiency tests, they will not be discussed here further.

Nation (2001: 374–375) lists three qualities that short-term tests should have:

1. Easy to make, because they may not be used again.
2. Easy to mark, because learners need to know quickly how well they performed.
3. Fair, or that they should not test too much considering a short learning time and they should relate to the contents in a predictable way.

Read (2000: 5) notes that language testing may let us know about two aspects of learners' language ability: language knowledge and strategic competence. In the context of this study, language knowledge corresponds to lexical competence, while strategic competence is the ability to use that lexical competence for instance in communicative situations and to make up for the shortcomings through alternative expressions. Of Nation's four types of tests listed above, language proficiency tests most often measure strategic competence, as Read (2000: 4) points out, and the other types may be more focused on language knowledge.

A criticism towards vocabulary testing has been brought forward by Ishii and Schmitt (2009: 6). They point out that tests often measure only partial vocabulary knowledge, so that if a learner uses the word *nutrient* correctly, it may hide the fact that they do not know the adjectival form *nutritious*. If deficiencies in lexical competence are to be detected, test that measure depth of vocabulary knowledge are to be used. Tests that are concerned with measuring lexical competence may evaluate different aspects of the depth of vocabulary knowledge. Below some of the approaches will be discussed.

Paribakht and Wesche (1996) devised a Vocabulary Knowledge Scale (VKS), improving on earlier similar propositions. VKS is a tool for assessing the from partial to precise knowledge continuum. In it learner self-reports on their knowledge on any given word on a scale from I to V, where I means that they have never seen the word before and V is that they can use the word in a sentence (see Item H in Appendix 2). The strength of VKS is that it shows whether the learners actually know the word or just guess the right answer. Additionally, it allows productive language use, which some of the other test types fail to do. Admittedly, VKS has its pitfalls, namely that is very linear, even though it may not even be possible to represent a learner's knowledge of a word on a single, linear scale (Read 2000: 136). Waring (2002) also notes that the scale ignores polysemy of vocabulary, which is problematic since, as already stated several times in this study, most words are polysemous.

Zareva (2005) tested how well Henriksen's (1999) proposition of vocabulary as a three-dimensional construct can be applied to vocabulary testing. The study was conducted using a modified version of VKS, and such has the same weaknesses and limitations as Waring's (2002) VKS. Also, it is questionable whether most EFL teachers know how to analyse the data using the tools used by Zareva, or indeed if teachers or students, for that matter, know how to read the analysed data. If these barriers are overcome for example through a different presentation of the data, Zareva's proposition for assessing students' vocabulary knowledge could be implemented in classrooms, especially amongst advanced learners.

An influential approach to testing the depth of vocabulary knowledge comes from Read (1998, cited in Schmitt, Ng and Garras 2011), whose Word Associates Test (WAT) has inspired a number of similar tests. These tests measure network of vocabulary knowledge, and Read's WAT does this by asking the testee to relate certain predetermined words to their collocations and near-synonyms. For each word that is tested, 8 words are provided, half of which are collocates and synonyms and the other four are distractors, from amongst which the correct words are to be recognised. While this kind of setting offers ample space for guessing, an issue Read was aware of too, there is another problem with this kind of tests: they are time-consuming to make, infringing Nation's first requisite for a short-term test. Additionally, if WAT is used for assessment on a language course, the teacher has to secure that all the paradigmatic and syntagmatic associates that are being tested on have been taught to the learners.

Nation (2001: 375) proposes that as a part of a short-term achievement test, students could be asked to list words in the same word family as the given word. Bogaards (2001: 323) criticises this idea, noting that which words belong to any particular word family are not as clear to language earners as to native speakers, e.g. words like *ad*, *add*, *adder, additive, addictive* and *addition* could confuse even an advanced learner. Schmitt and Zimmerman (2002) in turn devised a sentence completion test where morphological skills were tested, which in practice is a cloze test where the learner has to provide the right derivative. They argue that knowing different members (such as *evade, evasive, evadable, evader, evasiveness*) of a word family is imperative for productive language use, and those learners who lack derivative skills are unable to use a significant part of their vocabulary. This test differs from Nation's device in that instead of listing discrete words, learners have to produce the correct derivation into a context, and as such is likely to yield more reliable results. While as such it might work, they are likely to be time-consuming to make, and because English derivatives are not always regular, expecting learners to know which derivational suffix is used in each case may be too much from USS1 learners.

As a tool for measuring both the breadth and depth of learners' vocabulary knowledge, Ishmii and Schmitt (2008) devised a series of four tests in an attempt to test areas in vocabulary knowledge that teachers have found to be problematic in their teaching. Firstly, they measured students' approximate vocabulary size. Secondly, they tested how well students know the polysemous meanings of words. Thirdly, they tested their grasp of derivatives, and finally, the students were tested on near-synonyms. This multiple-phased test can provide both the students and the teacher with data on how well the students compare with the baseline norm and, in particular, what are the weaknesses and strengths of each student. Ishmii and Schmitt admit that it takes some effort from the teacher to devise the test, but once done, it is relatively easy to modify and to target to each group. One of its strengths is how the data is represented in a way that is readily understandable by most students. The test could be criticised of paying little attention to productive language use, as the only phase that measures productivity is the third part where students are asked to write the various derivatives of given words.

A problem with all the tests mentioned above is that they are only concerned with written language, ignoring oral skills completely. Furthermore, Read (2000: 5) criticizes discrete tests for assessing only the language knowledge and not strategic competence, which means that instead of measuring how well learners can use their vocabulary for communicative purposes the tests focus on how much they know about the spelling, grammar and other such aspects of vocabulary. However, if the test is used to measure only language knowledge and is not taken as an implication of their general language level, the criticism is void. More on the validity of tests, see section 4.2.

Above some of the suggestions to testing lexical competence have been discussed. As can be seen, there are numerous of types of tests aimed at assessing different areas of learners' lexical competence. A majority of the tests mentioned above are easy to conduct and to mark, and consequently, besides fulfilling the criteria set by Nation (2001), they are appealing to EFL teachers (Fitzpatrick 2007: 116). However, I would argue that they are best used for long-term achievement, proficiency or diagnosis tests. Word networks are complex structures which are learned gradually, and measuring acquisition of vocabulary network in weekly short-term achievement tests seems overly demanding. However, since testing directs both teaching and learning processes, testing the depth aspect is admittedly propitious for vocabulary learning. Giving the students broad tests that cover many aspects and require a lot from the learners may have a negative washback effect, and therefore a balance has to be found. This could be achieved by using a range of different test item types in one, relatively short test. The next chapter presents these different test item types.

4.1.2 Types of test items

Since this study focuses on short-term achievement tests or unit tests that measure lexical competence, here we are interested only in the types of test items those tests may have so that this section is not unnecessarily swelled up. Nation (2008: 145) lists the kinds of test items that are useful depending on the reason for testing, but as discussed before, there may be several reasons to give unit tests (motivation, awarding a grade, diagnosing etc.). Therefore, items from various categories are included here. Some of the types of test discussed in section 4.1.1 may work as test items on their own right (such as WAT), and they will not be repeated in this section.

There are two ways to test vocabulary: orally or through writing. While spoken tests are a functioning method of testing vocabulary knowledge and may be more suitable to some learners than written tests, it is likely out of all tests that are given, oral tests form a minority. In practice, however, Read (2000: 178–179) asserts that oral interviews are out of the question, since they are extremely time consuming. Testing the learner's depth knowledge on only a small number of words may take as long as 30 minutes, which categorically rules them out in regular USS teaching. Consequently, the only practical method is having written vocabulary tests.

Nation (2001:344–362) compiles a list of commonly used test item types. First of them is a true/false item, where the learner is provided with a sentence, such as "*Horses, whales and dogs are mammals"*. After the sentence, they write T if it is true, N if it is not true, and X if the learner does not understand the sentence. In the example, the new word being tested could be *mammal*. However, the student has to know also all the other words in the sentence before they can decide whether the sentence is true or not. Consequently, when designing these items the teacher has to be careful to use only high-frequency words or words the learners are supposed to know already.

The next item on Nation's list is a definition completion test, where a number of definitions are given, and the learner has to match the words to their definitions (see Appendix 2, item B). To increase the difficulty of the item, there can be more words than definitions. This type tests learners' receptive recognition abilities, and it also requires the learner to know all the words in the definition. This, Nation (2001: 359) points out, tests partial vocabulary knowledge. On the other hand, much like in the previous item, productive abilities are ignored since the only thing the students are require to do is to copy the right word from the list and write it to complete the definition.

A popular method of testing vocabulary is the multiple-choice item (Nation 2001: 349), where learners either choose the correct definition or synonym for the word. The problem with multiple-choice items is that they may inflate the score, because students can guess the correct answer. This seems to help especially the students with lower scores (Kremmel and Schmitt 2016: 379). Again, only receptive skills are tested.

There are different ways of using L1 translations in tests, and most test items may be modified to include L1. For example, in the previous item L1 words would have to be matched with their L2 definitions, or in the multiple-choice item either the choices or the target word is in L1. Other item types require the learner to do the translating: there could be a sentence with an underlined word that the learner has to translate, or there could be just a single word to translate, either from L1 to L2 or vice versa. Nation (2000: 351) claims that there is “a general feeling” against using L1 in vocabulary testing, but I would claim such sentiment is mostly absent amongst Finnish EFL teachers. As discussed before, using L1 and L2 may have very different roles in the learning process; in the case of using translation test items, translations from L2 to L1 measure receptive skills while translations from L1 to L2 are more likely to measure productive skills (Read 2000: 167). Using L1 may also facilitate test item devising. For example, creating multiple-choice item in L2 where the learners understand all the different choices can prove very challenging and on lower levels may even be impossible.

Gap-fill or cloze tests are a common assessment method, and several different modifications exist around the same idea, or that a word or a phrase has been omitted from the text (McCray and Brunfaut 2018: 52). The task of the learner is to restore the sentences. L1 translations may be provided, or all the missing words of the exercise can appear mixed in a word-bank next to the task. Also, the first letter(s) of the missing word can be provided (Nation 2008: 151). The problem with providing first letters of the word is that it may prompt the learner to give the correct word even if they do not remember its meaning or understand the sentence which it completes. This is especially true to short-term achievement tests, where students are often tested on a small glossary that may contain only few words that begin with a particular letter. Also, while the test-maker controls which words are omitted, and as such can focus the learners’ attention to certain words, there is also criticism that gap-fills may not measure what the test-maker intended to measure (McCray and Brunfaut 2018: 53). For example, gap-fills might measure more global reading skills instead of how well certain words have been learned. Regardless of the ongoing discussion, gap-fills are used widely by L2 teachers.

While the research discussed above focuses on single words, most of the test item types above can be used to test formulaic sequences. This means that instead of a single word a multi-word structure has to be translated, matched with its definition or associated with its collocates. It is likely that in practise teachers do this, especially with the more common phrasal verbs and other routine FSs, but research is scarce on the subject.

Even if use of context generally makes recognizing words easier (Nation 2001: 353), in certain test item types it may have the opposite effect on test grades. For example when translating whole sentences from L2 to L1 there is a risk that the learner does not recognise or recall the other words, thus being unable to translate the whole sentence correctly even if they know the target word. Similarly, if the learner is asked to produce a sentence where they use the word, inability to recall certain closely associated words may create a situation where the learner performs below their own level. On the other hand, if the ability to use the word is deemed essential, then the learner should also remember collocating vocabulary. Zhong (2012: 29, 48) discusses controlled productivity test items and free productive tasks. A controlled productivity item could be *"She was reading a ne______"*, where the target word is *newspaper*, and a free productive task could be for example "*Write three sentences about newspapers."* While controlled take less time from learners to do, free composition shows learners' command of the word better. On the other hand, they are also more difficult to grade reliably, and unless devised carefully, they may not trigger the target word.

Nation (2001: 375) also suggests asking the students to list collocations to a word or other members of the word family. However, unless these have been specifically taught on the course, they cannot be tested lest the test reliability be compromised. This kind of test item is possibly more suitable to highly advanced learners as it requires complex depth knowledge unlikely to be found in USS.

This section has discussed the strengths and weaknesses of some of the most commonly used test item types. However, when a test is designed, the teacher has to consider also whether the test measures what it is supposed to test and that the results are reliable. These test properties will be discussed next.

4.2 Test reliability and validity

When teachers design vocabulary tests, besides the type of test items they are using, they have two aspects to consider, namely reliability and validity. These aspects secure that tests work properly.

Reliability is *“the accuracy with which a test measures what it is supposed to measure”* (Daller et al. 2007: 15). This means that if a person performs the same test several times, they get the same result each time. If the results are different each time, it could be caused by a faulty test design, for example too short a test. Nation (2008: 153) posits that for a test to be reliable, it should have at least 30 items, although he does not state what this figure is based on. Also, the test format should be familiar to the learners so that attention and potential is not lost because test-takers do not understand how the test works (Nation 2008:153).

Validity in turn is “*the extent to which a test measures what it is supposed to measure*” (Daller et al. 2007: 16). This means that, for example, when testing the retainment of the glossary of a course book unit, the test measures the glossary and not the grammatical structures or glossary of another unit. Huhta and Tarnanen (2011: 216) point that even though validity sounds like a simple task, ensuring it is one of the major challenges in assessment. Even when tests are designed with the best intentions and utilizing the most recent research on the field, the way students do an activity may not match with the intended method (Hosenfield 1976: 120, cited in Nation and Webb 2011); for example, when students are supposed to practise grammar in reading contexts, they may often skip the reading and only fill in the grammar parts. Similarly, when a test is supposed to assess a general skill, like learners’ lexical competence, it proves difficult to create a test that indisputably measures this, unless proved by ample research, as Hughes (2003: 31) points out. To avoid this, the test-maker has to decide what aspect of lexical competence is being tested, and relatively valid results will be obtained (Hughes 2003: 30-33).

Kremmel and Schmitt (2016: 391) state that teachers may interpret vocabulary test results as suggestive of their general language level, even if the result only tells how the students performed in a particular vocabulary test on a particular day. A solution to this is to assess learner competence amply and for example to use different types of test items. Nation (2008: 153) notes that a valid test suits the level of the learners, which for experienced teachers should be easy to notice. In addition, to ensure validity, learners should take the test seriously. This last aspect could be ensured by making the test count in the final grade, or to make it a high-stakes test.

While teachers often state having separate vocabulary tests and grammar tests, Arnaud and Béjoint (1993) say that vocabulary tests and grammar tests actually measure the same aspects of lexical competence, and this relates to the fact that words always carry their grammar with them. In a study about vocabulary tests by Kekki (2017), students reported learning also grammar from vocabulary tests, while the teacher stated that the grammar was tested only in the final exam. Indeed, it seems that when productive language is assessed in a typical vocabulary test, reliability, as understood from strictly statistical point of view, may not be possible to reach with current testing methods (Daller et al. 2007: 16). Therefore, in the context of this study, more pertinent is scorer reliability, or how likely it is that the scorer will assess the same answer the same way, regardless of circumstantial factors, such as who took the test or what time of the day the grading takes place. This may be secured by laying out clear assessment criteria against which the answers are compared. However, when the tests are graded, the criteria that is being used has to allow the amplest variation possible in productive tasks, so that different answers that still are correct receive the same grade.

All in all, when deciding on the test items and assessment criteria, teachers have a lot to consider. The next section discusses how vocabulary is tested and taught in Finland.

4.3 Vocabulary testing and teaching in Finnish schools

This chapter explores how vocabulary is being tested and taught in Finland. There has been very little research on vocabulary testing in Finnish schools, which is all the more surprising seeing how ubiquitous it is. While in USS there is no standardised testing before the matriculation examination and the curriculum does not require vocabulary testing, the NCC (FNBE 2016: 240–241) dictates that teachers assess learner progress continuously and dynamically with the learners. In practise, this often translates to different kinds of tests and to vocabulary tests in particular. Personal experience tells me that if not all, most language teachers in Finland give vocabulary tests. Indeed, in a wide national study nearly 3,500 9th graders (6% of the cohort) reported vocabulary tests being one of the most used teaching methods by their EFL teachers (Härmälä 2014: 142). Additionally, Huhta (2017) states that unit tests are a common tool of assessment used by Finnish language teachers.

While Read (2012: 308) states that a multiple-choice type is a typical vocabulary test item type, in Finland vocabulary tests are stereotypically lists of words in both L1 and L2, to which the test-taker has to provide translation. Even if this is the stereotype, in practice vocabulary tests used by Finnish EFL teachers are likely to contain different types of test items from complementing test types so that various aspects of the vocabulary knowledge are tested. Most USS EFL course books are accompanied with vocabulary tests, and these typically contain various test types.

NCC does not dictate which words or expressions have to be learned on any given course, but rather paints general, abstract goals. For example, the goals for the 3rd English course (ENA3) in their entirety are:

> The students' multiliteracy is expanded and advanced in the course. The students produce texts in different genres, emphasizing linguistic accuracy typical for each genre. Themes include different cultural phenomena, English language media, and creative activity. (FNBE 2016: 118)

The goals, as can be seen, leave a lot of room for interpretation. But, even if the course descriptions do not include any goals to that are to be learned, NCC sets that the goal for USS students is to reach B2.1 level in the evolving language proficiency scale (ELPS), which is an application of the CEFR. In the ELPS, level B1.2 states that in text production "the student uses a reasonably extensive vocabulary and common idioms" (FNBE 2016: 258) and level B2.1 requires that the learner "masters a relatively large vocabulary" (ibid.: 259). *Mastering* vocabulary sounds like having at least reasonable competence in the depth aspect of vocabulary knowledge.

Martinez and Schmitt (2010: 26) comment on the difficulty of incorporating formulaic sequences in teaching unless using them gives some sort of advantage, such as elevated scores on essays. Finnish teachers are lucky in the sense that standardised testing only happens at the end of the USS, so at least in theory the teachers can give the additional value to the use of formulaic sequences when assessing student performance. As it happens, FSs are alluded to a few times in NCC. Firstly, there are common idioms mentioned above in the text production skills. Secondly, in communication strategies level B1.2 states that "the student is capable of […] using suitable expressions in interaction situations that deal with familiar topics" (ibid.: 258), where *suitable expressions* could be interpreted as being the FSs one typically encounters in interactive situations. The same category on level B2.1 states that learners should know how to "sometimes use standard phrases, such as 'That is a difficult question', to gain time" (ibid.: 259). This, in fact, is the one of the two instances in the whole ELPS where an actual example is given to illustrate a point, and it is a clear instance of formulaic language. I think based on these examples it is safe to interpret that ELPS places some weight on FSs, and expects teachers to convey their importance in fluent language to learners.

Besides FSs being alluded to a couple of times and one phrase being given as an example, nothing concrete is said in NCC as to what kind of words and expressions the learners are expected to master, and local curricula follow the same pattern. Therefore, the decision which words are taught is left for teachers to choose. In practice this will mean that the vocabulary in the course book, often chosen by the teacher themself, will set the objectives for vocabulary learning of the course. Since it may be difficult to know on what basis the makers have chosen the vocabulary used in the books, knowledge on word frequency lists might be useful for teachers as an accompanying material. Fortunately, these lists are readily available online.

A study by Albrechtsen et al. (2008: 202) suggests that the size of vocabulary is of foremost importance for learners and should be ensured first, supplemented by expanding the networks between the words. They also show that as far as at the university level it is important to "keep consolidating [the learners'] basic vocabulary", as the learners do not show stable mastery of high frequency words. The implication of these results is that teachers should be aware of high frequency word lists and use them actively as a part of their instruction methods. However, a study by Lahtikallio (2016) suggests that Finnish EFL students both in secondary school and USS have a much wider vocabulary than is expected from them. On the other hand, it seems that USS students might need more instruction on the depth aspect of vocabulary knowledge.

Albrechtsen et al. (2008: 198) argue that the L1 teacher and L2 teacher should work together to improve the complex processes that learning receptive and productive language skills require. They state, "Instruction in declarative lexical knowledge should also be addressed in first language instruction" (ibid.: 198), or that it shouldn't be only implicit as the learners would also gain in their L2. A common problem with some students is a vagueness of expression, where they settle for a general language instead of using precise words to express the idea. Accepting roundabout expressions or synonyms in vocabulary testing may reinforce this habit. Since the new curriculum stresses teacher cooperation across subject boundaries, the kind of action Albrecthsen et al. call for could be executed in Finnish USS.

All in all, while the NCC sets some limitations to what teachers may do, also a lot of freedom is left to the teachers, and in practise each teacher and the course book decide the contents of each course. The aspects of vocabulary knowledge that teachers deem important, however, has been studied little if at all in Finland. Furthermore, even though vocabulary tests are so ubiquitous in EFL teaching in Finland, little attention has been paid to them in research. This study attempts to shed light on the kinds of vocabulary tests and vocabulary test items that teachers use.

5 RESEARCH QUESTIONS

This study aims to find more about vocabulary teaching in Finland and what kind of vocabulary tests are given in EFL courses in Finnish upper secondary schools through the following questions:

1. What types of lexical competence do teachers value?
2. How are these values reflected in vocabulary testing?
3. Why do teachers give vocabulary tests, if they do?

With the first question, I want to find out what aspects of lexical competence teachers find important. As shown in chapters 2 and 3, learning vocabulary is not a straightforward issue, but instead there are many different ways of knowing a word. By answering this question, I hope to find out whether there are some commonalities between the teachers' values, and are there some areas of lexical competence that are ignored or, contrastingly, valued by all or most of the teachers.

The second question is focused on vocabulary testing. As was seen in chapter 4, vocabulary testing plays a big role in Finnish EFL teaching, and there are different ways of testing lexical competence. By answering the second question I aim to find what kinds of test item types USS EFL teachers use, and to see if they are consistent with the types of lexical competence they reported valuing in response to the first research question. Since nearly all teachers use vocabulary tests, it is important to study if the teachers are aware of the types of lexical competence different test item types are used to assess.

The third question follows naturally from the topic of this study. Again, as little research on the topic of vocabulary studies in Finland exists, this is included in the study for the purposes of future research, and it also helps in creating a profile about each teacher so that questions 1 and 2 are more readily answered.

6 DATA AND METHODOLOGY

This study is mainly concerned with *unit tests* as they are the prevailing method of short-term assessment, but should any other methods of testing arise from the data, they will be examined equally.

The present study is a qualitative study, which has implications on both the data collection and analysis. Hirsjärvi, Remes and Sajavaara (2009: 164) have compiled a list of common research methods for qualitative studies, which I will be summarising here. Firstly, qualitative studies tend to collect the data from people through for instance surveys or interviews. Secondly, the study does not aim to test a hypothesis or theory, but rather I will analyse the data comprehensively and in detail. While the study is not based on any hypothesis as such, all research surges from need and curiosity to find more about the subject, and the researcher will always have some assumptions about the results. Thirdly, the participants are chosen deliberately rather than using a random sample, and finally, the research plan is flexible and it should be adapted to the circumstances. Finally, during the analysis the data is treated to be case-specific and unique. In this chapter I will describe in more detail the methods that have been used.

6.1 Data collection

The present study is an exploratory study, and it is characteristics of exploratory studies that they investigate little known areas (Hirsjärvi, Remes and Sajavaara 2009: 138). As was established in chapter 4.3, vocabulary testing in language teaching in Finland is largely uncharted territory. As a consequence, the assumptions the study and the interview questions are based on have been derived on one hand from the theory and on the other from personal, anecdotal, experience in lieu of using earlier studies as a spring board for the research.

The data was collected through an interview because it is an efficient, natural and socially acceptable method of collecting descriptive data about the world in a way that is familiar to both sides of the interview (Dörnyei 2007: 134). Also, interviews are, as Mann (2016: 2) states, "widely held to be a fundamentally useful way to understand informants' beliefs, experiences and worlds". As this study aimed at discovering values, teaching practices and reasons for testing, interview was deemed a suitable method to collect the data.

The interview was piloted with one English language teacher, and the required adjustments were made to the structure of the interview. The actual interviews were conducted in February 2018. They were carried out face-to-face when possible; when not, they were conducted via Skype or other similar platforms. The interviews were conducted using a semi-structured interview (Dörnyei 2007: 136). This method was chosen because it is flexible enough so that were any new openings to emerge during the interviews, I could response to them to find more about the topic (Dörnyei 2007: 37). These responses or probes include asking for clarification and for further explanation, such as asking for justification or reasoning for opinions that the interviewees state (Hirsjärvi, Remes and Sajavaara 1997: 205) or asking for clarification for what the interviewees mean by the words they use (Dörnyei 2007: 138). The interviews lasted from 30 to 50 minutes, as some teachers were chattier than others. For the framework of the interview, see Appendix 1. As the interviews advanced, I noticed certain probes yielded more thoughtful responses than others, and those probes became a part of the interview. The slight variation between interviews does not undermine the reproducibility of the study, since as Richards (2005: 74) points out, the differences in the results are "food for analysis, not a problem for consistency", and on the other hand variation is part of the nature of semi-structured interviews, as described above.

All the interviewees were native Finnish speakers, and consequently the interviews were conducted in Finnish. The choice of using L1 was deliberate: to ensure firstly, that the interviewees would fully understand the questions and secondly, to minimize any pressure they might feel, seeing that they were interviewed and questioned on their teaching methods, a situation that in itself is bound to make some people feel uncomfortable. Any linguistic terminology, such as the breadth and depth of vocabulary and formulaic sequences were explained to the teachers to ensure that the answers are comparable.

The interviews were made up of three parts. The purpose of the first third of the interview was to see how much the teachers have thought about vocabulary learning and what kind of approach they have to vocabulary in general. The second part aimed at finding how much time and effort the teachers dedicate to vocabulary instruction and practise and what they think are on the one hand the responsibilities of the teacher and on the other those of the learners in the learning process. The third part was focused on vocabulary testing and on the reasons teachers have and the ways they test vocabulary learning.

At the end of the interview the teachers were presented with a list of common test-item types (see appendix 2), and in addition to reporting whether they used any given test item type, the interviewees were prompted to give their opinions on each type, if they had any. After we had gone through the test-item types, the teachers were asked if any that they used were missing from the list. The purpose of presenting them with the list was to decrease the probability of forgetting any test-item types the teachers might use; indeed, this happened in the pilot interview.

The ethicality of data collection was ensured by giving the teachers as much information as possible on the purposes of the study and on how their privacy will be secured (Dörnyei 2007: 65). Only the audio of the interviews was recorded and later transcribed. On one hand, the visual gestures of the interviewee are not relevant for the study, and on the other hand, a better privacy is secured for the interviewees since their faces were not stored.

6.2 Participants

The subjects of this study were 9 qualified Finnish upper secondary school (USS) teachers who are all giving compulsory English courses. This was made to ensure a relative homogeneity between the teachers; being USS teachers they all have to follow the national core curriculum. Some of the teachers I approached directly through e-mail, having found their addresses on school websites, and others were found through an English teacher network on a social media platform. All the interviewees were female, as male teachers could not be reached or they did not respond to my emails. Also, male teachers are harder to reach in first place as 70% of all language teachers in Finland are female (Kumpulainen 2016: 65). Teachers from different geographic areas and with differing teaching experiences were chosen for the interest of maximum variation sampling. This means that the participants were markedly different (albeit not by their sex), and if there are commonalities across the spectrum, we may assume that they are reasonably stable within the wider group (Dörnyei 2007: 128). On the other hand, the purpose of this study is to see what kinds of ideas Finnish EFL teachers have about teaching and testing vocabulary; since the area has been studied little, this study does not aim to create an all-encompassing theory but aims to serve as a basis for further studies on the topic. The participants' teaching backgrounds are discussed more in chapter 7.

The sample size of 10 was chosen firstly to ensure that there is enough material for the analysis. Secondly, since there is practically no research on vocabulary tests used in Finnish schools, 10 participants seems enough so that should there be follow-up studies, the results may give a meaningful point of departure for those studies. The audio file of first interview, however, was corrupted and could not be used, so 9 interviews were used in the analysis.

6.3 Methods of analysis

The analysis was performed using data-based content analysis, because it is a practical way of reducing large amounts of unstructured data into a manageable size. Also, it allows understanding the data in a way that normal reading does not permit. (Riazi 2016, Dörnyei 2007).

After the interviews were conducted, they were transcribed removing all the data impertinent to this study, such as pauses, hisses and intonations (Tuomi and Sarajärvi 2009: 109). Other content editing such as restructuring the orations was not made at this step, since as Dörnyei (2007: 247) points out, we do not know at this point what might be relevant in the analysis. The participants were also given pseudonyms at this point.

The next step was coding the data, as suggested by Dörnyei (2007: 250–252). During the transcribing and after it pre-coding took place. This meant rereading the transcriptions and making journal entries about my impressions over the texts. This shaped and structured the actual coding, which followed. The coding had two phases: initial and second-level coding. In the initial coding, each interview was read carefully through, highlighting content relevant to the topic, and giving it a label. The second level consisted of comparing the different responses and finding more abstract commonalities between them, clustering them into wider categories.

The final stage comprises interpreting the data and drawing conclusions (Dörnyei 2007: 257). The interpretations grew naturally from the data, and the selection of the themes was based on their salience and on how they related to other important categories. However, when analysing the data, some of the framework was derived from the existing theory. As Tuomi and Sarajärvi (2009: 95, 117) suggest, I combined old theory and empirical data to formulate a new theory in chapters 7 and 8.

The answer to the first research question was drawn from an analysis of interviews, where the teachers were asked both about their values and about their vocabulary teaching practices. Since lexical competence includes various different aspects, and restraints of time and resources limit teachers from focusing on all of them, they are forced to make choices which aspects to pay less attention to. The second question was answered through two interview phases followed by the analysis of the data. Firstly, teachers were asked what aspects of lexical competence they assess in their tests, and secondly they are asked over the kinds of test items they use in their vocabulary tests.

An example of the method of analysis is how the teachers were grouped on their perceptions on the importance of FSs in vocabulary learning. First, all the references to FSs were highlighted and giving them labels such as "*encountered all the time*" and "*full of them*", and the labels were usually taken from the data itself. Then, these were combined under a more abstract label, "*omnipresent"*. All the interviews underwent a similar process, and then these groupings were contrasted to each other and to the theory discussed in chapters 2, 3 and 4. When formulating the abstractions, special care was taken to ensure that the analysis is faithful to the interviewees' line of thought, and all results were checked against the whole interview to minimise erroneous conclusions. On the other hand, the analysis is always a product of the researcher's interpretation and as such it is inevitable that the results are somewhat subjective. Regardless of its weaknesses, qualitative research is a functioning method of doing research on new, little-known areas (Dörnyei 2007: 39-42), and as such is an appropriate research method for this study.

7 EFL VOCABULARY TEACHING AND TESTING PRACTICES IN FINNISH UPPER SECONDARY SCHOOLS

Most if not all the teachers reported at some point of the interview two things that affect a lot of teaching and testing choices they make, namely the quick pace that is required in USS courses and the focus there seems to be on grammar in the first two USS English courses.

Teachers also reported on having little theoretical expertise on vocabulary learning and teaching, since during their studies the topic was not covered to any depth. Some also stated their knowledge on the subject matter was based mostly on what they had learned during their own studies, which in some cases goes back to the early 1990's or 1980's, Kirsi stating that "*All this theory didn't exist back then"*. However, many of the teachers had had a long career in USS, and as such they have learned in practise the kinds of problems learners have and also what kind of lexical competence is required form the students in the matriculation examinations.

As discussed in section 6.2, all the teachers were teaching at the time of the interview in USS. One of the teachers (Tuula) had 20 years of teaching experience in USS, and she had never taught on other levels. Contrastingly, Siiri had been teaching less than two years, and this was her first year in USS. Table 1 compiles the information on the teachers' teaching experience.

Table 1. Teaching background of the interviewed teachers.

Teacher	Length of teaching career in years (rounded up to next full year)	Teaching level
Saana	20	Most of career in primary school, has been teaching several years in USS.
Nina	4	Always in USS. Uses a system where students advance independently on an online platform.
Ulla	15	Always in USS.
Siiri	2	First year teaching in USS, previously secondary school.
Tuula	20	Always in USS.
Mira	10	Last 5 years in USS, before that substitutions at varying levels including USS.
Kirsi	Over 25	Last 8 years in USS, before that various institutions, but all post-compulsory school age group.
Sinikka	20	20 years at an USS that runs an International Baccalaureate (IB) Diploma Programme, while teaching English also to regular USS students.
Maria	2	First year in USS. Teaches at a school where a significant part of the learners have dyslexia or learning difficulties.

The interview quotations will be given in the analysis only in English since the audience of this study speaks English. Also, for this study the wording is not important but rather the content of the quotes is what matters and therefore the original quotes need not be present here. For the sake of reliability, however, each quotation is numbered and they are listed Appendix 3 both in Finnish and English.

I will first discuss what the teachers think about the process of learning vocabulary in general. This is an important point to explore, as it sets the basis for how in the whole the teacher approaches vocabulary teaching and testing. Then, in the section 7.1.1, I will examine which aspects of lexical competence do the teachers think are important for their USS1 students. The subsequent sections analyse how the breadth/depth division and formulaic sequences are present in the teachers' thinking and teaching. Finally, the use of context in teaching vocabulary will be discussed in section 7.1.4.

Chapter 7.2 discusses the teachers' attitudes toward vocabulary testing, and explores the timing of tests and motivations for giving them. Finally, the areas of lexical knowledge teachers reported testing and the test items they used are examined in chapters 7.2.2 and 7.2.3.

7.1 Vocabulary knowledge in the teachers' perceptions

Teachers approached the broad, open-ended question of "What is vocabulary learning?" from various directions. Following the line of the Common European Framework of Reference for languages (Council of Europe 2001), dividing lexical competence to productive and receptive was shared by several of the teachers. For example, Saana stated that:

> (1) "It's a long process, you learn passively little by little, and then it changes at some point into active [knowledge], or I mean some of it becomes active." - Saana

In Example 1, Saana recognises that not all words ever form part of a learner's productive vocabulary knowledge, but instead some of them stay in the receptive knowledge dimension shown in Figure 1 (Zhong 2012) in section 3.2. Saana also acknowledges that vocabulary learning does not happen quickly, but rather is a gradual process, as discussed for example by Chacon-Bertrand (2010). However, the statement also seems to turn the learner into a passive subject in the vocabulary learning process, but I believe it is only a matter of unfortunate wording. This is because receptive and productive knowledge are sometimes called active and passive knowledge. Therefore, more in line with Saana's other responses would be that they meant that the process from receptive to productive vocabulary knowledge is gradual, as proposed by Henriksen (1999). In any case, Saana seems to suggest that once a word "becomes active", or enters the learner's productive vocabulary, it is stable. Saana, much like all the other teachers interviewed appear to ignore the fact that words are forgotten without frequent maintenance (Baddeley 1997: 123), a point that is discussed further in section 7.2.1.

Similarly to Saana, Maria also pointed out that a lot of the words never get to the "second stage" or become part of the productive vocabulary. Mira elaborated on this by dividing the receptive vocabulary knowledge to those that you understand from context and those that you understand without a context. This requires the word to have become cemented enough so that contextual clues are not required. This bears a striking resemblance to Henriksen's (1999: 304–307) concept of semantization process, where the depth dimension of a word is learned while the word is going through the receptive-productive dimension. Tuula expands on what productive knowledge to them is: knowing how to use words in speech and writing with the correct meaning and in the right context. In Tuula's approach the learner has to be fairly advanced in the depth dimension before their language can be productive. Both the depth dimension and use of context in teaching will be discussed in subsequent sections.

Another approach to vocabulary learning was the communicative point of view: these teachers considered that lexical competence is above all communication, an opinion shared most strongly by Nina, Ulla and Sinikka. However, when their descriptions of what communication actually means are looked at more carefully, they seem to go back to what was discussed previously. For example, Sinikka states that:

> (2) "I guess it [=learning vocabulary] means that the student acquires means of using the language so that they can get by using a foreign language. In different situations, [knowing] what kind of language and vocabulary you need and stuff like that." - Sinikka

while Ulla says that:

> (3) "Learning vocabulary means that you are able to communicate in as many registers as possible and in various ways." - Ulla

In Examples 2 and 3 we see that both Sinikka's and Ulla's focus is on interaction, consisting of the depth dimension and productive knowledge of Zhong's (2012) proposal on vocabulary knowledge. To Ulla and Sinikka, learning vocabulary is not an end-goal itself, nor is it related to passing examinations, but rather the aim is to become independent users of language in authentic situations. For this to become reality, certain steps need to be taken. First, the learner has to attain an advanced depth dimension of enough words so that the kind of range of language use required in Examples 2 and 3 is possible. Secondly, the mastery of words has to become so consolidated that productive language in the various registers and situations is possible. However, communicative language also includes various suprasegmental areas of lexical competence such as intonation and juncture, and consequently it largely falls out of the scope of this study. In any case, at least the kinds of aspect Li and Kirby (2015: 613) referred to have to be mastered until the criteria set in Examples 2 and 3 are fulfilled.

However demanding to the learner the communication-oriented approach may seem like, the third group that shared commonalities had an even more all-encompassing view of vocabulary learning. They stated that *"everything is vocabulary*" (Ulla), and that *"it's the starting point of all language learning*" (Kirsi), while Siiri listed nearly every aspect of language skills as being vocabulary learning:

> (4) "It means that a student learns new words, different contexts of use, pronunciation, learning to recognise the word when they hear it, learning to write it and knowing how to use it in different contexts and maybe a good student also understands, of course not immediately, but when you can use words such as 'bitch', 'hoe' and 'woman'" - Siiri

Example 4 includes at least six of the nine aspects listed by Wallace (1982: 27), and the answer is something the teacher came up with on the spot, with no preparation. Indeed, from the answers it is clear that teachers recognise that teaching vocabulary is more than trying to get the learners memorize lists of words. All of them acknowledge that there are numerous aspects to each word, even if the approaches differ from teacher to teacher.

Another matter altogether is what teachers consider to be a word in the first place. While most of them recognised that sometimes several orthographical words may come together to create a single meaning, the orthographical division seems to be the predominant approach to understanding what a word is, undoubtedly influenced by word-counts used in essays. Even if this is the case, teachers are clearly aware of formulaic sequences, and they dedicate time for teaching about them. The uncertain answers that the teachers gave to the questions concerning the more abstract nature of vocabulary were most likely caused by the fact that they had not given much thought to it or that they had not put those ideas into words before. However, when they answered questions about their teaching methodology and especially the vocabulary tests, it became clear that the ideas were there. These ideas will be discussed in subsequent chapters.

Now that we have a conception of how teacher perceive the process of vocabulary learning, we may see what aspects they deem important for their learners, and how they approach the those aspects in their teaching methodology.

7.1.1 What aspects of lexical competence are important?

As seen in section 3.1, lexical competence includes a whole range of different vocabulary-related skills, and teachers are highly aware of this. Consequently, when teachers were asked which aspects they find important for USS1 learners to know about words, the answers included everything from spelling to fluency of productive use. Commonalities did arise, too, and all the teachers mentioned knowing the meaning of the word as the key element to be learned about a word, or as Kirsi put it succinctly: “Meaning is the be all and end all of vocabulary knowledge”.

Another aspect that was well presented in the replies was the orthography, although some teachers give leeway to the learners; it is sufficient to spell the words *“more or less correctly*” (Maria) or that knowing the spelling is “*helpful*” (Mira). A good pronunciation of the words is deemed an important skill, but several mentioned that it is not feasible to focus on it given the fast pace of USS courses that the teachers reported.

Other than these three aspects, little generalisations or categorisations can be made. For that reason, a brief description of each teacher’s valuation is given here, so that some comparisons and connections can be made later in this chapter.

Starting with Saana, they say that the students need to know mostly those areas of lexical competence that Nation (2001) listed under the meaning and form aspects: meaning, morphology, pronunciation and spelling. In other words, Saana seems little concerned with the use aspect at the USS1 level, focusing mainly on what perhaps could be thought to form the traditional approach to vocabulary knowledge.

According to Kirsi, in USS important are mainly those aspects that are assessed in the matriculation examination, especially orthography. The matriculation examination assesses learners’ written receptive and productive skills and receptive oral skills, and as a result, productive oral skills bear lesser importance. The influence of the matriculation examination on the aspects of vocabulary knowledge that are taught in USS English classes may be presences in the responses of other teachers, too, but no one makes it so clear than Kirsi. Indeed, they say that

> (5) “Of course it is important that they learn how [the word] is pronounced, but since for the time being all assessment [in the matriculation examination] is carried out in writing, it is the writing [that is most important]”. - Kirsi

In Example 5, Kirsi points out that even though learning to pronounce a word is important, the way the matriculation examination is arranged leads to the written language and orthography in particular receiving heightened attention. This contrasts heavily with Sinikka, who stated that meaning and pronunciation are the two most important aspects of a word, followed by orthography and knowledge about the usage the word. In general terms, fluency of spoken expression is very important to them, and throughout the interview Sinikka stressed the emphasis that they give to oral language. In other words, their aim is to shift the learners’ semantical knowledge from the receptive to the productive dimension (Zhong 2012). Here we can see the influence of the IB experience, as they are more used to teaching content in English instead of explicit language instruction.

While Sinikka stresses the importance of oral production, Nina considers that in USS1 the use and register are of secondary importance. Instead, meaning and orthography are what receive most attention from them. Those skills that Nation (2001) counts under the use aspect are, according to Nina, fine-tuning which can be focused on after USS1. This could be influenced heavily by their use of an online teaching platform, where the students advance independently and the majority of exercises are written exercises, marked automatically by a programme. When marking is performed by a machine, there is typically no room for orthographical errors, causing an emphasis on correct spelling. Similarly, electronic platforms are typically programmed to accept only certain answers, while synonyms or roundabout expressions are marked faulty. As a consequence, the learners need to focus on the vocabulary and the exact phrasings that appear in each unit. Whether this is pedagogical depends on the point of view, but for the discussion on synonyms in vocabulary tests, see section 7.2.2.

Like Nina, Siiri also seems to emphasize written language, but they stress that the learners should know above all *how* to use a word in a sentence. This is a skill that in fact requires the mastery of several areas of lexical competence, e.g. at least an understanding of the morphological and syntactical behaviour of a word, but also some knowledge of the collocations a word may have and knowledge of the various meanings of a word, all of which have to become part of the learner's productive language. Mira shares this view on the importance of productive language use:

> (6) "Most useful would be that the learners themselves can produce [language], but at least they should be able to recognise the word and preferably without a context". – Mira

In Example 6 Mira states that productive language skills are the most practical skills for a language learner. However, should this prove too demanding, receptive skills are also helpful. Mira also added that the ability to recognise isolated words is relevant since in the matriculation examination words may appear without a context. Here Mira shared with Kirsi the view that those aspects of lexical competence that are useful in the matriculation examination are important for USS1 learners. Also, Example 6 shows an understanding of what Henriksen (1999: 304) calls from partial to precise dimension, or that the meaning of the word is learned gradually. When a learner has only vague idea of what a word means, contextual clues are an important aid for correctly guessing or in retrieving the word from the memory.

Ulla and Tuula had more trouble in defining the aspects of word knowledge that they perceive are important for USS1 learners. Ulla thought that what there is to know about a word depends on the word itself. Concrete words, such as *a cat* are more straightforward to learn and to use, but for abstract words and polysemous words the learning is much a slower and complicated matter. Ulla stressed the importance of knowing the different meanings of words, the word class and the cultural context in which the words may be used, among other things. Ulla also stated that even though pronunciation is important, a lot of students already have a strong idiolect when they start USS, and within the time constraints of USS courses it is not possible to focus on this aspect of lexical competence. Like Ulla, Tuula also thought that the required lexical competence depends on the word. For some words, they said, it is enough to have receptive abilities, while learners need to be able to use other words actively. However, in contrast with Ulla, Tuula stated that it is essential that they practise pronunciation of words and listening, so that they recognise familiar words upon hearing them. This suggests that Tuula deems receptive and productive oral skills essential to the students, in which they share the value with Sinikka.

Finally, Maria saw that ideally the students would know the orthography of a word, could pronounce it understandably, use it in a context and understand the word in a context. In other words, Maria values both receptive and productive oral and written skills. Later in the interview, however, they stressed the productive oral skills, reporting that they dedicate as much time as they can to different kind of oral exercises. In this aspect Maria is similar to Sinikka, to whom not only was it essential that learners could produce fluent oral language, but that they also had communicative abilities.

Above we can see significant differences between what the teachers think are essential lexical competences for USS1 learners. While some focus more on the written aspects of vocabulary knowledge, often influenced by the looming presence of the matriculation examination, others concentrate in the learners' pronunciation and productive skills. Despite the differences, Henriksen's (1999) first dimension, or the meaning is present in the answers by all the teachers, and could be even considered to be the central focus of most teachers. In this light, it is intriguing to see what importance teachers may give to the rich meaning representation or the depth dimension of a word (Henriksen 1999: 305–306). This topic will be discussed in the next section.

7.1.2 The importance of breadth and depth

This chapter reports what importance the interviewed teachers give to teaching breadth and depth of vocabulary, as discussed in section 3.1 (see, for example Aitchison 1994: 63, Henriksen 1999: 305–306). A common line of thought for many teachers was that USS1 learners need to expand their vocabulary quickly, because in the secondary school the course book texts were short with a limited vocabulary. This idea was accompanied with the perception that even the advanced learners at this level had a very narrow lexis, and three teachers mentioned specifically that they had next to no knowledge of synonyms to the words in their vocabulary, a case in point being the word "nice", which kept cropping up across the interviews. There are also students, they reported, that may be very fluent in their expression but have a very narrow vocabulary or that the vocabulary it is limited to certain, highly specialised fields, such as vocabulary needed in online gaming of shooting games.

Most if not all teachers found the breadth of vocabulary a more pressing matter than depth at USS1. Nina pointed out that at this level it is important that the learners learn new words rapidly, because unless there is sufficient breadth to the lexis, it is impossible to get any depth to it either. Siiri thought in similar lines, saying that first the focus is on breadth and then it can be shifted to depth. A stronger statement for the importance of the breadth of vocabulary came from Saana:

> (7) ”[The vocabulary] is really important, it affects everything; from understanding the language to language production: you can understand written text and spoken language even without a deep understanding of the grammar if you have a large vocabulary, so it is really important.” - Saana

In Example 7 we see that Saana argues for having a broad vocabulary rather than extensive and detailed knowledge about the grammatical rules. It may be argued that the depth aspect of vocabulary includes several aspects that traditionally are considered to be a part of grammar, such as the syntactical and morphological behaviour of a word (Henriksen 1999: 305–306), and it appears that Saana ranks these below the breadth of lexicon in importance. Still, Saana does not disregard depth completely, stating that the depth should be focused on once the basic vocabulary is broad enough, whatever “broad enough” may mean. Other teachers are mostly in line with the opinions shown above, although they may not be as extreme as shown in example 7, but instead they are more in line with Nina and Siiri. However, when this is collated with the aspects they deemed important in the section 7.1.1, there is an interesting collision. Nina and Siiri, among others, thought that it is essential for the learners to learn how the words are used. I would argue that the depth dimension is in fact knowledge about how the words are used. This apparent contradiction will be discussed at the end of this section.

Some teachers were divided on the issue of the relative importance of vocabulary breadth and depth. Mira mentioned that since there are great differences between the students, they need to be instructed accordingly. Expanding the breadth, she said, is more important for the less advanced learners while advanced learners benefit most from instruction that focuses on depth. However, Mira admits that especially in USS1 grammar is preferred over vocabulary building, although this is not necessarily by their own choice but the structure of the course books lead to this. Regardless of the reasons, Mira states that usually there is not enough time for teaching vocabulary, even though all the students would benefit significantly from it.

Tuula thinks similarly to Mira, stating that:

> (8) ”Maybe it’s the breadth at that stage [that is essential]; depth is something that only during the USS we get into the point where we can make the students aware of the fact that the word is not just a translation.” - Tuula

Tuula expresses in Example 8 that while the breadth may initially have a more central place in language learning, one of the goals of USS is to get to a point where the learners realise that English words are not direct translations of their L1 counterparts, but rather that they operate based on fundamentally different parameters. This means that Tuula aims that the students become aware of the differing lexical networks that words have in different languages (Singleton 1999:80) and that this spans the whole USS, starting from the first year. Nina also mentioned that making students aware of the differences between languages is a major challenge. They gave an example of a recent situation, where student was faced with the expressions “give in” and “give up”, both of which have the same translation to Finnish. In this case, the student needs to create two separate nodes in the network of words (Albrechtsen, Haastrup, and Henriksen 2008: 25) where in their L1 there is only one.

All in all, there seems to be a preference to teach the breadth of vocabulary. It does not, however, necessarily implicate that teaching depth is neglected altogether during USS1 or that they do not consider it an essential area of knowledge, as seen above. The situation is quite the contrary, indeed; many of the teachers thought that learning how to use vocabulary is one of the main objectives of USS1. A comprehension of how to use a word is a combination of learning its meaning and depth dimensions, while productive language is a performance of that comprehension (Zhong 2012: 29–30). Additionally, despite teachers expressly stating that breadth is more important to them, a lot of the methods they use are more than suitable to teaching depth and may even work better to this end than expanding the breadth. This use of context to teach vocabulary will be discussed in the section 7.1.4.

7.1.3 Formulaic sequences in the interviewed teachers’ teaching

While formulaic language includes many different kinds of categories (Wray 2002: 9), in the interviews phrasal verbs stood out as a category of FSs that teachers are faced with regularly. Most likely this is because they are such a frequent part of the English vocabulary, and of the spoken language in particular. In addition, they are heavily represented in English course books, drawing further attention to them from the teachers. Also, since the meaning of phrasal verbs is not the sum of their parts, they may have to be focused on more so that their meaning becomes clear to the learners. Phrasal verbs were by no means, however, the only category of FSs that arose in the interviews, as can be seen in this chapter.

Formulaic sequences appeared to be an important part of language teaching in the English classes of the interviewed teachers, which is only natural given the pervasive nature of FSs (Nattinger and DeCarrico 1992: xv). All the teachers reported dedicating time and effort to teaching and practising formulaic language to some extent, and they saw it as an important aspect of vocabulary building. Teaching FSs is not a simple task, however, and Tuula said that:

> (9) “The challenge of teaching a language is precisely how to explain why certain words go together.” - Tuula

In Example 9 Tuula captured in few concise words all the problems that FSs poses to a language learner. Like Meunier (2012: 112) pointed out, learners do not have the intuitive resources for formulaic languages that native speakers have. Since there usually are no clear reasons why certain words appear next to each other, teachers would often find themselves answering the learners’ questions with a response like “it just goes like that” when there is a clash between the collocational networks that words in different languages have.

Problematic as they may be, FSs are seen as a key element to a language. Sinikka said that collocations are exactly what is characteristic of each language, adding that unless a learner understands the collocations, they will not understand how the language works. While this may be a slight exaggeration, collocations and other FSs arguably form a major part of texts[3] in English (Coxhead 2018: 113) and learners need to have a good grasp of collocations and formulaic language if they want to speak the language fluently.

Despite the challenges that teaching FSs poses, the interviewees agreed that instruction on them is generally speaking helpful and necessary, but quantifying how much teachers actually teach FS proved to be difficult. Teachers stated that FSs are often mentioned in passing or that they use drills or productive exercises aimed at practicing FS. They also reported that the course book texts have a lot of them, although there may be an overrepresentation of certain types of FSs, such as idioms (Ylisirniö 2012). However much time each teacher dedicate to teaching FS is left unanswered, but when answering follow-up questions the teachers invariably stated that they form a considerable part of the content of their courses.

---

[3] “Text” understood in its broad sense, including speech, videos, songs etc.

All that being said, there are still fundamentally different approaches between the teachers as to how important FS are deemed. These differences may partially arise from the different meanings they give to the term "formulaic language", even if the meaning of the term in the context of this study was explained to the teachers during the interview. In their responses, some teachers focused principally on idioms, others were more fixed on phrasal verbs, while yet others talked mostly about teaching collocations. It may be that each teacher talked chiefly on the topic they are most familiar with and that may be most heavily present in their teaching. In spite of these differences of focus, two groups could be observed among the teachers.

The first group that emerged from the data are those teachers who thought that formulaic language is present in teaching at all times. Maria stated:

> (10) "[FSs] appear in the texts all the time, so yeah they do exist there, and also in concrete communication, but there's no separate exercises." - Maria

Example 10 shows that Maria is clearly aware of the ubiquity of FSs both in the course book texts and in the interaction that takes place in the classroom. Regardless of how omnipresent FSs are, Maria reported not giving the students separate exercises to practise them. It is likely that FSs are present in most exercises, but that Maria does not focus on them specifically. Indeed, Sinikka and Nina also stated that that all texts are full of formulaic language, but they questioned whether a single method to teach them even exists. It seems that being highly conscious of the pervasive nature of FSs and the range of different phenomena it includes, the teachers consider that all of them cannot be taught or paid attention to on a regular USS English course.

The situations that they described where they do teach FSs are almost organic, where they both react to any encountered FS and provide formulaic language related to single words that appear in the materials. In other words, they are expanding the learners' depth dimension by presenting them knowledge on the formulaic behaviour of certain lexical units. Especially Nina reported having ample possibilities of doing this, as their instruction is mostly self-directing. It means that they have time tackle parts of the text individually with learners a phrase at a time, and Nina reacts to any FSs in the manner that the situation demands.

The second group could be classified into a group that, in Ulla's words, deal with FS "*whenever they are encountered*". If indeed up to 50% of all language is formulaic (Coxhead 2018: 113), this seems unlikely. It is more likely that Ulla referred to the most common FSs and to those that are more clearly formulaic, such as idioms, greetings and phrasal verbs, rather than the less easily perceivable sequences such as *bordering on* and its semantic prosody (Schmitt 2004: 9). Ulla also stated that when working with the vocabulary and the text, they might make the students pay attention to certain phrases that appear in them. Siiri methods in treating FS are similar, namely that students are provided L1 translations to central phrases in the texts, to which they have to find translations. Since formulaicity seems to be largely ignored in textbook exercises (Meunier 2012 114), any additional focus that is drawn to them is bound to be helpful. Even if the course book did include varied formulaic language and had diverse exercises focusing on FSs, a further emphasis placed on them would not be unnecessary due to the sheer multiplicity of FSs. Additionally, Siiri's method utilizes the L1 of the learners, which helps the poorer learners in particular (Schmitt 2008: 337; Carter 2012: 196).

Most teachers fit into these groups, but there are two outliers. Firstly, Kirsi did not report spending much time on FSs. Kirsi said that they had a printout which has several pages of idioms and other formulaic sequences, and that in the last year of USS they drill these in one lesson. Besides that, Kirsi mentioned that the new course book they are using has lists of idiomatic expressions on different topics, but the responsibility of practising them falls on the learners. Kirsi also stated that it is good to know some idiomatic expressions so that they can use them when writing essays. This reflects that Kirsi considers FSs mostly a novelty with which to zest their written language. All in all, there seems to be limited focus on FSs, and it is mostly up to the learners to dedicate time on them if the course book offers opportunities for practise.

The other outlier is Mira, who, unlike Kirsi, considers FSs an essential aspect of fluent language use, but finds teaching them problematic due to differences in learners' language proficiency.

Groups typically have students with widely differing levels of English, and Mira said that for some students most FSs are too complicated an issue altogether. Collocations in particular pose problems to the less proficient students, and Mira reported that collocational expressions may become fixed in their vocabulary so that the learner starts using the collocating words together always. This is undoubtedly a fault in the networking process, creating the situation where if one of the lexical items is accessed, the other follows it. The process also strengthens itself, as with every repetition the expression gets more cemented in the network (Chacon-Bertrand 2010). Mira also believed that the highly proficient learners will learn the FSs through encounters with and usage of the language, and for those learners separate instruction is unnecessary.

While Mira and Kirsi did not dedicate much time to teaching formulaic language, for the rest FSs constitute an important aspect of their teaching methods. Research suggests that if formulaic language is taught cognitively, for example analysing the function of a preposition in a phrasal verb, the learning outcomes are better (Alali and Schmitt 2012: 172). In such analysis the words go through the kinds of elaborate processes Eysenck (1982) argues to facilitate retention. Some teachers reported doing this kind of analysis, for example, Maria had their students compare the literal translations of idioms and phrasal verbs with their actual meanings and make connections between the two, if possible. Nina in turn stated having the learners analyse the functions of prepositions in phrasal verbs, exactly as Alali and Schmitt (2012: 172) suggest.

As can be seen, teachers view the role of FSs in a language very differently. Several teachers pointed out, however, that often the meaning of opaque FSs is taught the most clearly when they are set in a context, such as a whole sentence or in a dialogue. Next chapter examines how the interviewed teachers use context in their teaching.

7.1.4 Using context in teaching

Aitchison (1994: 63) stated that knowing a word requires knowledge about the words that appear with it. Extensive knowledge about a word also includes information on the register to which the word belongs to. It may be possible to teach this explicitly, but the easiest way to do it is to present the word in a context, and then analyse the setting. This information can also be learned incidentally, but as discussed in section 3.2, there are limitations to incidental learning. It seems that the best way to learn these aspects is intentionally, and presenting the word in a context; lexicon network is best constructed when words appear together with other words (Elgort et al. 2015: 506–507, Matsuoka and Hirsh: 2010). For the most part, the teachers that were interviewed for this study seem to agree with these arguments. This chapter explores the ways the teachers use context in their teaching.

Many of the interviewed teachers found using context a crucial part of vocabulary teaching. As was seen in section 7.1, teachers required that learners know how to use a word in the right context before it can be thought to form a part of their productive vocabulary. For this to happen, the word has to be presented to the learners enough times in different contexts, as argued by Eysenck (1982). Indeed, Tuula stated that it is the responsibility of the teacher to provide the learners with varied materials and different contexts for the words to appear in.

While Ulla agreed that in order to acquire a rich meaning representation of a word, learners need to encounter it in varied contexts (Elgort et al., 2015: 506–507), they also stated that course books do not offer sufficiently varied language this to take place. On the other hand, Ulla continued, there is not enough time in USS English courses to expose student to the range of texts that is required for this to take place. Consequently, USS students are left with the responsibility to deepen their vocabulary through consumption of different types of media on their free time.

Sinikka emphasized the significance of the use of context in as authentic situations as possible, often through oral exercises. These methods, she said, develop learners' productive skills, so that they are able to use the correct words in correct contexts. This echoes Zhong's and Hirsh's (2009) observation that how vocabulary is learned is influenced by the types of teaching methods and vocabulary exercises that teachers use.

As discussed in the previous section, Nina dedicates a large proportion of classes to going through the texts and exercises individually with those students who need assistance at that moment. During these situations questions over FSs often arise, and Nina said that the best way to teach multi-word units is through context, or as she so catchingly put it, "*context is the king*". A recent example of such situation they described was the difference between *give in* and *give up*:

> (11) "[The student and I] discussed because they had put 'give up' to Quizlet because the Finnish translation said 'luovuttaa' and Quizlet said that the right answer is 'give in', so they asked that doesn't 'give up' also mean that, so I told them that not in this unit and in this context, and then I gave them two sentences and gave them two different meanings to 'luovuttaa' and then they could connect which one is which, and then they understood that one means 'antaa periksi' and the other means 'lopettaa tekemästä' or 'ei enää jaksaa'; if I had just told them that this one means this and that one means that, they wouldn't get it, so these two situations show it better." - Nina

Quizlet is a platform for practising vocabulary, and it typically has bilingual glossaries. In the situation described Example 11, the phrasal verb *give in* appeared in isolation, causing problems and confusion with a similar expression *give up*. The teacher used context to clarify that there are two translations to *luovuttaa* depending on the context. This is clearly a case of the kind of clash between L1 and L2 vocabulary networks described by Singleton (1999: 80), further complicated by two opaque FSs. Nina also said that collocations and other formulaic sequences are often present in the exercises, causing similar problems.

While using similar methods to that shown in Example 11, Maria also advocates of incidental learning through context. The course books leave little room for incidental learning, since they are designed with the precise idea that learning is intentional. Consequently, Maria describes using a profusion of external material, including but not limited to videos, news articles and songs. The goal of this, they say, is to expand the learners' vocabulary and to give the new uses in context, hoping that the students assimilate some of the information. Most teachers, however, seem to avoid using external material. The main reason for this is, I would argue, are the effort it takes to find suitable material. Not only has the content of the material fit in the course, but also the level of language in the text has to be suitable for the learners, since ideally there would be only a few new words (Nation 2001: 233).

External materials can also be used for intentional learning, in which case the teacher needs to prepare some exercises concerning the text if full use is to be taken from it. Since the course book texts are well prepared and have accompanying exercises, there is little incentive to do it. Furthermore, most teachers said that the course schedules typically do not allow covering the course book fully, let alone using external materials.

According to Schmitt (2008: 337), after the meaning-link has been consolidated between L1 word and its L2 counterpart, it might be beneficial to use L2 context to teach how the word is used in context. Mira, however, claimed that even the more advanced learners in their group gained from using L1 translations of whole texts. Mira said that translating the whole text gave the learners insights into the meaning and behaviour of words that otherwise would not be achieved. Mira also said that they would give examples of different contextualised uses of certain key words in the texts. Siiri reported using very similar methods with similar reasoning. This seems to conflict with Schmitt's argument, but it might be possible that the USS1 learners Siiri and Mira taught have not achieved the proficiency where the benefits of using L1 start decrease.

The ways teachers use context seems to correspond quite well with the valuation they gave to the different areas of lexical competence in section 7.1.1. For example, Sinikka, who emphasized the importance of oral communication, reported using oral exercises. Contrastingly, Siiri, to whom written language plays a more important role, used written translation exercises. As can be observed from this section, the interviewed teachers have a range of teaching methods were the context is utilized to further the learners' lexical competence. Even if those methods differ sometimes drastically, all the teachers seem to acknowledge the value of context in vocabulary learning.

7.2 Vocabulary tests

This chapter concerning vocabulary exams the interviewed teachers give is divided into three sections. The first section discusses how often teachers gave test and the motives teachers reported for giving vocabulary tests. The second section is dedicated to what kind of lexical competence they declared to asses, while the third part is focused on what test items they used. Both the second and third sections also include a discussion on how these compare with the aspects of vocabulary knowledge they deemed important earlier.

All the interviewed teachers except Sinikka gave vocabulary tests, although also Tuula reported not having them necessary on all the courses. In Tuula's case, however, not giving vocabulary tests on certain courses it is due to the nature of those courses, rather than a categorical decision of not having them. That only 1 out of the 9 teachers interviewed does not give vocabulary tests keeps well in line with Härmälä's (2014:142) findings that most Finnish EFL teachers use vocabulary tests. Sinikka acknowledged that she is part of the minority, and went as far as to call herself "*a renegade*", though saying this facetiously.

Sinikka, nonetheless, did state that she has given vocabulary tests at the beginning of her teaching career, but she stopped doing it after some time. She said:

> (12) "I realised they were utterly useless, firstly we lose time from the class, then I have to correct them, and the students didn't study for them, so it was a complete waste of time and energy, they could spend those 15 minutes doing something else. So someone swishes through the test in two minutes, while other tinkers 12 minutes with it, so the first is whistling for 10 minutes and just waits, because I can't just go and take the [slower student's] paper from them. So I prefer them doing something else those 15 minutes." - Sinikka

As can be seen in Example 12, there were several aspects that Sinikka found problematic. Firstly, valuable time was lost doing the tests. Second point, also relating to time, is that some students may spend most of the test time doing nothing if they finish the test quicker than other students. For the reasons of fairness, however, enough time has to be allotted for the test so that also the slower students may perform to the full extent of their abilities. The drawback of this is that it can create situations where 19 learners are waiting for the last one to finish their test. Were this to happen repeatedly, the loss of time would indeed grow significant and alternative options to vocabulary testing might grow more appealing, as happened in Sinikka's case. The third point Sinikka made is that the learners did not study the vocabulary for the test, which, in the end, is one of the main arguments for having vocabulary tests. It could be that Sinikka's students did not perceive them as being high-stakes tests, where a positive washback effect is typical (Wall 2012: 79).

Another point that Sinikka made was that memorizing word lists is not useful and could be compared to reading phone directories, or elaborate mental processes that promote learning are absent (Eyesenck 1982). Sinikka also criticized that if vocabulary tests are given, a week later those learners that do study will have forgotten everything, implying that learners are studying only for the exam, as argued by Krashen (1989: 455). It could be that with longer tests the learning outcomes would be more substantial and the test would also be more reliable (Nation 2008: 153), but extending the tests would mean increasing the time allotment, exacerbating the problems raised above. All in all, Sinikka opposed separate vocabulary testing strongly and saw no benefits to giving them, asserting that they are "*a complete waste of time and energy"*. Nevertheless, she reported that the final exam includes a lot of vocabulary from the course, but since it is beyond the scope of this study it is not explored here further.

As the rest of the teachers reported that they give vocabulary tests, the following chapters will discuss their motivations, goals and methods of vocabulary testing.

7.2.1 Spacing of and reasons for giving vocabulary tests

This section examines some general aspects of the interviewed teachers' vocabulary tests. I start by how often teachers tend to give tests, followed by the motivations for giving vocabulary tests in the first place.

The frequency with which the teachers give vocabulary tests varied greatly from one teacher to another. One extreme is Mira, who said that she has a test for each unit of the course book, which in practise translates to having a vocabulary test each week. The other extreme is represented by Tuula, who said that they test half of the vocabulary midterm, and the other half is tested in the course exam at the end of the course. Tuula added that more would be a waste of time, in which she echoed Sinikka's position on vocabulary testing. The rest of the teachers were somewhere between Mira and Tuula, but most have a vocabulary test for each unit, which in practise means having a test at least every other week, which is surprisingly similar to what Canadian ESL teachers reported in Rossiter et al.'s (2016) study.

None of the teachers repeat vocabulary from earlier units in their vocabulary tests, even though as seen in chapter 3.3, repeated, spaced exposure to the words when studying for the test would be propitious for learning vocabulary (Cepeda et al. 2016), and indeed frequent repetition is necessary for vocabulary retainment (Baddeley 1997: 123). Several of the teachers were quick to point out that the words are asked again in the final course exam, whereas Tuula stated that once the learners have *"cleared out the vocabulary"* in the unit tests they are not explicitly asked about it in the final exam. These methods seem to be at odds with Rossiter et al.'s (2016: 7) argument that cumulative testing produces better learning outcomes and lowers the likelihood that students cram for the test, forgetting everything soon afterwards.

Concerning repeating words from earlier units, Nina said:

> (13) “I think it’s a little bit unfair if there are some weaker [learners] in the group, and then you’re like ’Surprise, here’s this word from the previous unit’, but of course sometimes there may appear words that we have studied earlier and they are supposed to know them but because we’ve had them before, but yeah, there’s little overlap between the tests.” – Nina

In Example 13 Nina stated that she does not include in vocabulary tests words that have not been in the latest unit, because she regarded that it would not be fair to the weaker learners. This implies that Nina feared that the lower-achieving students would have forgotten the vocabulary that had been studied earlier. One could argue that it is even more of a reason to test those words so that students would gain from the backwash effect, but maybe the demotivating effect of such testing practices would be greater than the possible positive learning outcomes. Ulla said that learners protest if words from earlier units are tested, and she had grown tired of having the same argument every time. However, as Nation (2008: 153) indicates, the learners should be familiar with the test format that they are given. If they knew that also older words are included and were expecting it, there might be significantly fewer protests.

As to why give tests in the first place, motivating learners to study vocabulary seems to be a key reason. All the teachers, with the possible exception of Sinikka, recognised the encouraging effect vocabulary tests may have on learners’ studying habits (Read 2000: 170). Maria said that she gives the tests principally to motivate students to study, adding that since the learners in her groups could not be expected to do their homework, revising for the test forces them to encounter the words repeatedly. Saana thought in a similar vein, going as far as to saying that:

> (14) ”For many students it’s the only motivator to study, that’s my main reason for testing”.
> - Saana

Example 14 shows that Saana held the opinion that unless they are forced by an external motivator, a significant portion of the learners will not study vocabulary. In fact, she said that it is the principal reason to test vocabulary in the first place. By the same token, Ulla said that she might not give vocabulary tests were it not for the learner’s insisting that they have them. In other words, even the learners themselves are aware of the coercing effect tests have on their studying practices. While Mira recognised the motivational effect of testing, her reasoning was somewhat different from the rest. She said that the main objective is to encourage the weaker learners to encounter English also outside classes, because, she reasoned, the stronger learners are likely to seek situations where they use English, be it productively or receptively.

Vocabulary tests serve also as a tool for assessment, as became obvious from the answers the teachers gave. How much it affects the course grade varies greatly from one teacher to another. Some teachers, like Saana said that the effect is small that it is to all intents and purposes negligible. Maria in turn stated that she already has more than sufficient material for assessment, but since the vocabulary tests are graded, they form a part of the final grade. In contrast with the rest, Nina admitted that the tests are used mainly for assessment, and the motivational effect is a by-product. Consequently, Nina's tests have a significant role on the final grade. Vocabulary tests bearing a heavy weight in assessment is not limited to Nina. Also Siiri said that the vocabulary tests make up for 40% of the final grade, which definitely makes them high-stakes test, and as such should motivate the learners to revise vocabulary (Wall 2012: 79).

Maria and Tuula also said that vocabulary tests are good tool of measurement for how much the learners have learned, and shows whether there is need for revision. This means that Maria and Tuula use vocabulary tests also for diagnosing (Nation 2001: 373). Tuula added that in USS1 this diagnosis can also be self-directing, so that the learners themselves see the results of their work and where they might need to centre their attention.

Even though Mira tests vocabulary nearly every week, she explained that it is not necessarily because she enjoys giving them, but rather there is a larger problem in the USS system. Firstly, she stated that the Finnish school system is test-orientated, implying that students expect to be tested on everything they are supposed to learn. Secondly, due to a heavy workload, USS students have to make choices what to study, and unless they know they will be tested on the vocabulary they will opt for studying other subjects. For these reasons Mira found it imperative to have vocabulary tests as students would not study otherwise.

7.2.2 Areas of lexical competence tested

This chapter describes what areas of lexical competence the teachers stated that they assess in their vocabulary tests. In the next section will be with the types of test items they use. These are examined separately, as many test items can be modified to assess nearly any area of lexical competence and linguistic competence in general.

Saana reported asking in vocabulary tests mostly the meaning and orthography of the words and how the words are used in context. She stated expressly that she does not test pronunciation or register, however, Saana will remark if slang words are used. This is mostly in line with what she stated earlier, namely that besides pronunciation they find meaning and orthography the essential areas of lexical competence.

Nina said that she tests orthography and morphology, but more complex aspects of lexical competence such as collocations do not form a part of her tests. However, Nina penalizes if the learner uses colloquial language in their answers where a standard form is required, e.g. writing *wanna* instead of *want to*. In this sense, she does require some perception of the register words may belong into, at least when it comes to the most common words. Siiri is mostly on the same track as Nina, but she said that syntax, FSs and collocational structures are often present in the tests, even if they may not have the central focus.

Ulla stated that she requires their learners to know morphology and syntax in vocabulary tests, especially for the basic vocabulary, while for the less frequent words meaning and orthography is already demanding enough. Tuula echoed Ulla in all aspects, and it is possible that the similarities in their answers is a result of the commonalities in their teaching careers, as they both have taught in USS for over 15 years.

Maria said that vocabulary tests focus heavily on orthography. There are two reasons for this; firstly, as they are written tests, they will inevitably test learners' orthography, even if the test does not expressly focus on this aspect because the responses have to understandable even if mistakes are allowed. Secondly, Maria explains that whole USS trains for the matriculation examination, and as long as orthography is important there, it has to be emphasized throughout USS. Maria asserted that her vocabulary tests test always both vocabulary and structures, which for her meant the collocations and depth aspect.

Mira agreed with Maria in that matriculation examination affects also vocabulary tests, and she tests mostly orthography and meaning, whereas the depth aspect of words is seen less relevant. Mira stated that they might comment on the register if a learner used a word that is totally out of line, but besides that, little attention goes to testing depth.

None of the teachers test pronunciation, despite the fact that several teachers said it is an important part of knowing a word. There could be several reasons to this, the least of which not being the fast pace of USS English courses, a fact which teachers expressed with annoyance. Another reason was stated by Nina, who said that "*it would require an oral test*", something that the teachers are uninclined to do, possibly for the complexity of organizing it. Mira also pointed out that it would be problematic due to the various accents of English there are. A solution could be, in Maria's words, that the learners pronounce the words so that "*they are understandable"*, yet this poses other problems, such as *whom to* has the pronunciation be understandable. Due to these limitations it is not unreasonable for the teachers to opt out from testing pronunciation.

In section 7.1.1 it was noted that all the teachers found knowing the meaning of a word an essential aspect of lexical competence, and orthography was also a highly valued skill. As can be seen above, teachers carry these ideas to their vocabulary tests, even to the point where: *"Vocabulary tests are pretty much about orthography"*, as Maria says.

When asked about partial-precise dimension (Henriksen 1999: 304–307), teachers found the idea amusing, questioning whether this could even be tested. However, Tuula said that quite often tests have items where the learner is instructed to choose from among options the word that best fits the context or the phrase, and this kind of tasks may indeed measure partial vocabulary knowledge.

Despite the prevalence of FSs in English (Coxhead 2018), none of the teachers reported testing on them, or at least they do not pay any extra attention to them in their tests. A couple of teachers mentioned including phrases in their tests, but as the term covers everything from verbs conjugated in tenses to everyday expressions and noun phrases, little can be deduced from that. Regardless, it is likely that some common FSs, e.g. sentence builders such as *"it is possible that*" and phrasal verbs, are tested especially by those teachers who include phrases or entire sentences in their vocabulary tests. There are two reasons to assume this: firstly, purely by their ubiquity some are bound to end up in the tested phrases. Secondly, many FSs are intriguing to pedagogues and are the kind of competence that teachers appreciate.

No one assesses pronunciation separately except for Maria, although she did not do it in the vocabulary tests but rather has separate tests for pronunciation and speaking. Similarly, Sinikka said that she can easily hear the level of each learner's pronunciation during the classroom activities. Earlier Sinikka said that in her classroom there are plenty of spoken activities, so this seems plausible enough, although how evenly and equitably she may listen to every learner's speech can only be guessed.

As discussed in chapter 4.3, accepting synonyms or roundabout expressions in vocabulary testing may have adverse effect on the language learning in general (Albrechtsen et al. 2008: 198), while it could also be taken as a sign of the fluency of expression. The teachers that were interviewed are also divided on the issue, but most teachers think that synonyms are acceptable if they fit in the context. For example, Maria said:

> (15) "I accept it if it fits in the context but if it means something else or is the right word but not the right register or context I deduce a bit. They use a lot of roundabout language and I try to award them, especially with the weaker learners I've tried that they write everything they know and can, so even the weaker can go around these really well, sometimes the word or phrase is expressed in a very complicated way, but hey, I always strive for that communication would be the main thing." - Maria

In Example 15 Maria pointed out that the less-advanced students may not remember the exact word that is in the unit, but she may use a different word or a set of words as long as the answer is communicative and expresses what is required. Maria in fact seemed to value also strategic competence (Read 2000: 5), accepting even convoluted answers as long as they were communicative.

Kirsi shares the notion of Example 15 that the learner does not have to remember the exact word of the unit, accepting a synonym as long as it means the same and it could be used in an authentic situation. Even though this was a common line across the interviews, not all the teachers were so lenient. Nina gives only half a point for a synonym from outside the glossary since for her, she declares, the whole purpose of vocabulary testing is to broaden the learners' vocabulary. This is in line with what was noted earlier, namely that Nina considers broad vocabulary more important than other language skills, such as grammar.

7.2.3 Types of test items

Since all the teachers teach several groups, there is ample room for experimenting with different types of tests and test items. Consequently, the more experienced teachers in particular are likely to have had enough possibilities to discover the kind of testing practices that best suit their teaching methodology.

When it comes to devising the test itself, many teachers reported using the vocabulary tests that publishers provide with the course book. Teachers opt for using these because they believed they have already been thought out by whoever created them, and teachers perceived that they tend to include the important phrases and vocabulary of each unit. Saana said that there is a selection of different test item types for each unit out of which they choose those items she considers most suitable for the group. The kinds of test items she uses will be discussed later in this chapter.

There were also those teachers who create their own vocabulary tests, such as Kirsi and Ulla. Kirsi used translations of words from L1 to L2 and vice versa, and consequently could easily choose the words to be tested each time. Ulla also reported creating their own vocabulary tests, since the publisher's tests are too long and the test item types too complicated for the purpose. She said:

> (16) "I give short vocabulary tests, […] there are 10-15 items and then there might be some sentence translations, I have noticed that for those who are good it does not matter if there are 1 or 100 items, they will study them in any case, and others will not study no matter what, so you can tell it with shorter tests". - Ulla

In Example 16 Ulla stated that the length of the test is irrelevant, as the more diligent learners will study regardless while the less hard-working learners are unlikely to revise in any case. Devising her own tests, Ulla can create such tests as they deem most suitable, even though Nation (2008: 153) advises to give tests that have at least 30 items to ensure their reliability.

Some teachers, such as Siiri, draw both from the tests provided by the publisher and from their own creativity. Siiri stated that some of the publisher's test items are "*ridiculously difficult*" so she devices some or even most of the test items herself. Since in Siiri's courses vocabulary tests form 40% of the grade, it is understandable that the level of the tests is not too high so as not to drag down the final grades unnecessarily. Also, being a tool for motivation for the learners to study the vocabulary, too difficult a test could have a demotivating effect on them (Wall 2012: 84).

The test item types that teachers used vary a lot, and once again there seem to be as many different opinions on which items are the most functional as there are teachers. While some teachers thought translation tasks should never be used, others reported using them exclusively. Here these different approaches are examined more closely. The test item types referred to here with letters can be found in Appendix 2.

Out of the 10 common test item types, there were two that none of the teachers reported using. Firstly, item A faced criticism for several reasons. Mira said that it tests more than just the target word, and Maria thought that it could be used to test grammatical structures. Nina thought that it could be a good method to test the poorly performing learners, but for the majority it would be too easy. Ulla and Siiri thought the exact opposite, thinking that it could be more suitable later in USS due to its difficulty. As can be seen, there is huge variation between the responses, and in this sense item A represents quite well the kind of range of responses each item prompted.

The other item that none of the teachers had used in USS vocabulary tests was the item I, a slightly modified version of VKS (Paribakht and Wesche 1996: 180). In fact, none of the teachers reported seeing it before, but they thought it could be used for self-assessment while its usefulness for summative assessment was questioned. Its practicality as a way to measure partial word knowledge was not apparent to the teachers, and Siiri in fact stated that the first two spots could be removed to improve the item. It seems that teachers do not regard partial vocabulary knowledge something that could or should be assessed.

Siiri stated that she does not use translations at all in tests, saying:

> (17) "I hope to rid the students from the mindset that language is first translating words into Finnish and then saying them." - Siiri

For that reason, she prefers using types B, C and D, where the lexical competence is tested through other L2 words. What she said in Example 17, however, conflicts with her teaching methods, because she stated that translations indeed are a great way of learning vocabulary. It could be that Siiri aims at creating the initial meaning link with the use of L1, as suggested by Carter (2012: 90).

It seems that after this link has been made, however, Siiri aims at weaning the learners from using L1 translations, and avoiding such tasks that require the connection to be made is part of the process.

Tuula, on the other hand, uses all the different test items shown except for item A and I, aiming to have varied tests. Doing this is an effective way of testing different areas of lexical competence and probably learners too find the variability in test items refreshing. Similarly, Nina used all items B-H except D, but this was due to it being a new type for her, and she said that she might adopt it to her tests. Also, Nina said that her vocabulary tests used to include sentence translations (Item G), and while as such they worked well, marking them was complicated and time-consuming, especially since she uses peer-grading where learners mark each other's tests.

Ulla, who devises the tests herself, used sentence translations (Item G), cloze testing (Item H) and different multiple-choice items (B, C, D). The use of those different types of test items permits for assessment of different aspects of lexical knowledge. Earlier Ulla said that syntax and morphology are key skills to that she wants to test, and Items G and H respectively are good ways of assessing them, while B, C and D can be used for various purposes.

Kirsi said that she only ever uses type F, both from L1 to L2 and vice versa. There may be some phrases or FSs, but everything more complicated than that takes too long a time to devise and to grade. However, since Kirsi uses only single-word translations, their validity as a measurement of lexical competence may be questioned (Hughes 2003: 31). But, since she uses them principally to encourage vocabulary revision and the tests' weight in the course assessment is tiny, this kind of testing, while currently unpopular amongst teachers, has fewer problems than might be expected.

In a stark contrast with Kirsi, Saana stated that she uses all types B-H other than F. Saana said that words may have so many different translations depending on the context that it is difficult to grade them and to focus them on what is being tested. The other tasks that either present the word in a context or are multiple-choice tasks (especially C) can be targeted more precisely.

Mira stated that when devising the test, only the kinds of structures and vocabulary has been taught can be included in the tasks. For that reason types A, B, C and D are problematic, because when creating them one has to be careful not to include in any of the choices vocabulary that they have not faced during the classes. Mira pointed out that she uses whole sentence translations (type G), but this is only possible because they translate the entire unit texts together in class. This, she continued, ensures that everyone has access to the information required to do the task. In other words, Mira used mostly single word or phrase translation tasks (Items E, F, H), and used whole sentence translations only if it has been translated together in class before. Mira reported that productive language tasks have their pitfalls. An example she gave was an actual task where the learners had to write a Christmas wishlist using the vocabulary of the unit and it had to include possessive structures. The lists then had items such as "*grandmother's benevolence*", which despite of fulfilling all the task criteria does fail in some other areas. While it seems that in 7.1.1 Mira deemed productive skills important, it may not directly transfer to their testing methods. On the other hand, in short-term achievement tests, productive tasks have to be targeted in such a way that they trigger the desired vocabulary. But, as seen in the anecdote, this may cause further problems.

If Mira avoided having sentences and productive language in her tests, Maria stated that she never tests just vocabulary but mastery of the structures is equally important. Consequently, she asserted that items E and F did not seem functional to her, except some phrases and FSs could be tested with item E. Maria reported having used a modified version of C, where the learners write the definition in L2. Maria also said that gap-fill type H is a good task to measure the breadth of vocabulary, as the learners may use the near-synonyms and roundabout expressions they know. However, Maria pointed out that she is a new teacher in USS, and that she is still figuring out what is the most functioning method for her.

Based on the interviews, teachers use very different methods of testing vocabulary, and no two teachers reported using the same combination of test item types. Teachers agreed only on that type A was not practical to USS, even if they gave opposing reasons to it. Similarly, type I was deemed more suitable for self-evaluation rather than for actual short-term vocabulary testing. Next I will discuss the general results and implications of the study.

8 DISCUSSION

Based on the results of the analysis, it is clear that teachers deem good vocabulary skills essential or even prerequisite for both receptive and productive language. Vocabulary learning was in the interviewees opinion, besides the basis for communication, "*the starting point of all language learning"* (Kirsi).

The present study aimed at answering the following questions:

1. What types of lexical competence do teachers value?
2. How are these values reflected in their vocabulary testing?
3. Why do teachers give vocabulary tests, if they do?

As may be expected, the differences from one teacher to another were great in most if not all of the topics. However, there were some interesting commonalities and equally interesting differences.

The teachers reported valuing principally those skills that are the primary focus in matriculation examination, namely meaning and orthography, and some even said that good performance in the matriculation examination is their main objective as a teacher. This is noteworthy, since the learners at in question were still two to three years away from the examination. Teachers indeed do teach for the test, and, students will presumably study for it, just as Krashen (1989: 455) suggested.

Some of the teachers also deemed important knowing how to "use" a word – a vague word which can include nearly all vocabulary knowledge areas depending on the teacher. Nation (2001) defines the *use* aspect of knowledge consisting of collocations, register and frequency, but in the light of the interviews it is very unlikely the teachers meant this. Instead, it seems that the teachers generally wanted their students to have flexible productive language skills, which, however, is not always reflected in the vocabulary tests.

While teachers knew about the existence of at least some types of formulaic language, it seems that its ubiquity and centrality in language learning is something of which they were not fully aware. This manifests itself in that relatively lot of time and effort phrasal verbs and idioms, while other types of FSs may be slighted. Partially it could be caused by the fact that teachers were not aware of the existence of less apparent FSs, such as semantic prosody.

Teachers focused mostly on expanding breadth and stated that this is the principal goal in USS1. However, a study by Lahtikallio (2016) suggested that it might be more important to focus on the depth aspect. The concept of vocabulary as a network proposed by Aitchison (1994) did not seem to appeal to the teachers, and teachers rarely pay attention to vocabulary network creation, at least intentionally. The teacher who seemed to pay the most attention to vocabulary networking and FSs does not give vocabulary tests but rather practises those skills through other means. Because vocabulary networks and FSs are such an important skill, I would suggest that the universities that train EFL teachers focused more on assessment and the role of FSs in language fluency.

The answer to the second research question can be boiled down to a single sentence: the teachers' claims of what they hold important may not be reflected in their teaching or assessing practices based on the interview results. Generally it seems that there is a decreasing slope to fewer aspects of lexical competence if we compare which aspects teachers deem important to which they report testing. Only one of the teachers reported testing pronunciation even though most found it an essential language skill. Furthermore, teachers who stated that having functional, productive language abilities was important to them reported using test types that allow little productive language.

This mismatch of ideals and reality could be down to several reasons. Firstly, the teachers may not have thought about the topic previously, and they may have said that certain areas of lexical competence are important to them only because it was the first thing that entered their head; for instance, when discussing about formulaic sequences with Tuula, she noted, "*I have never thought about this before"*. Similarly, Nina said, "*I have never actually put these things in words before*", which implies that a follow-up interview could have been useful so that they would have had more time to reflect on the topic. There is also the possibility that what teachers stated doing may not be what they are actually doing, but this could be proven only by a larger study that followed their teaching and testing practices over a period of time.

Also, they might have been were replying what they felt like was the "correct" answer to the questions, although at the beginning of the interview I expressly said that there were no right or wrong answers and I was not assigning any values to the answers but the study. If they did this, they might have ended up giving incongruous answers. Indeed, several teachers mentioned that they graduated such a long time ago that “*all the theory must have changed since then”* (Kirsi), and they may have felt somewhat uncomfortable for a moment. Finally, it could be down to the structure of the interview or indeed to the meanings that I have given to the statements made by the teachers. However, when drawing conclusions, I first drew a general picture of each teacher’s values and methods, and kept these in mind when analysing the data further.

The interesting result of this study is, however, that despite of the commonalities in the teaching methods and in the ideas of what vocabulary learning is, there are huge differences in the testing. The test items that teachers use vary from open-ended productive written tasks to single word translations. Vocabulary tests are given mostly to motivate learners to study, but they are also used for assessing the learners’ lexical competence, but also other reasons came up. One teacher does not do any short-term vocabulary testing, having abandoned them years ago as “waste of time and energy”.

Many of the differences in the teaching and testing methods undoubtedly arise from the great autonomy Finnish teachers have in their classes but equally from their personal values and experiences. The length of teaching career does not seem to have any implications on the methods teachers use. Instead, personal preference appears to be the defining factor. New teachers may also prefer the methods their own teachers used, as Siiri stated doing. Also, earlier teaching experience on other levels may affect teaching methods less than one might expect, although admittedly these factors were not focused on particularly neither during the interviews or in the analysis.

The results of this study have some implications for both EFL Training Programmes and EFL teaching in Finland. The National Core Curriculum for upper secondary schools does not talk about formulaic sequences, depth of vocabulary knowledge or network building. Consequently, they play a tiny role in EFL teaching in Finland. If these skills were mentioned explicitly, more focus would have to be paid to the topic both during the teacher training programmes and when teaching the language.

Also, vocabulary testing being as common as it is, teacher training programmes might gain from explicit focus on the kinds of vocabulary tests and test items that exist, highlighting their weaknesses and strengths.

While those are the main implications of the study, I do recognise there are limitations to this study. As an exploratory study with a small sample size, the results can hardly be generalised to the whole Finnish EFL teacher body. Since most EFL teachers in Finland do vocabulary testing, further, large-scale studies on the kinds of test-items that are used and the kinds of lexical competence that teachers test would bring important data on the general tendencies of Finnish EFL instruction.

While this study does not aim to make any statistical claims, it is interesting that whatever the interviewee said about item type B, types C and D triggered the same responses. This effect, and some other responses could have been affected by the order of the questions. The teachers, who mainly work on practical matters, were also questioned what they thought about some quite abstract ideas. Preparing them for the interviews with tasks or reading materials might have resulted in quite different results.

It is often said that teaching methods change slowly, and in the light of this study the same might be true of vocabulary testing and teaching too. A study that compares the methods used by teachers with careers of differing lengths would bring information whether there is any change or development in those practices. Similarly, a study on the content of teacher training programmes with a focus on assessment and vocabulary teaching would help in improving those programmes, and as a result, could have an impact on the quality of EFL teaching in Finland.

Since teachers focus mostly on breadth of vocabulary on the first year courses of upper secondary school, stating that depth becomes important later, a further study is needed to see whether this indeed is the case with more advanced learners. Also, if there is a change in the content, it could also be studied whether this change in focus is reflected in the types of vocabulary tests the learners are given.

BIBLIOGRAPHY

Aitchison, J. (1994). *Words in the mind: An introduction to the mental lexicon* (2nd ed.). Oxford: Blackwell.

Alali, F. A. and Schmitt, N. (2012). Teaching formulaic sequences: the same as or different from teaching single words? *TESOL Journal* 3 (2), 153–180.

Albrechtsen, D., Haastrup, K., and Henriksen, B. (2008). *Vocabulary and writing in a first and second language: Processes and development*. Basingstoke: Palgrave Macmillan.

Anderson, R. C. and Freebody P. (1981). Vocabulary knowledge. in Guthrie J T. (ed.) *Comprehension and Teaching: Research Reviews.* Newark, DE: International Reading Association.77–117.

Arnaud, P. J. L. & Béjoint, H. (1993). *Vocabulary and applied linguistics (Repr.)*. Basingstoke: Macmillan.

Baddeley, A. D. (1997). *Human memory: Theory and practice* (Rev. ed.). Hove: Psychology Press.

Bogaards, P. (2001). Lexical units and the learning of foreign language vocabulary. *Studies in Second Language Acquisition* 23 (3), 321–343.

Carter, R. (2012). *Vocabulary: Applied linguistic perspectives*. Milton Park, Abingdon, Oxon: Routledge.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T. and Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin* 132 (3), 354–380. doi: //dx.doi.org/10.1037/0033-2909.132.3.354.

Chacon-Beltran, R. (2010). *Second language acquisition: insights into non-native vocabulary teaching and learning*. Bristol: Multilingual Matters.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Cambridge: Press Syndicate of the University of Cambridge.

Coxhead, A. (2018). Replication research in pedagogical approaches to formulaic sequences: Jones & Haywood (2004) and Alali & Schmitt (2012). *Language Teaching*, 51 (1), 113–123. doi:10.1017/S0261444815000221.

Cruse, D. A. (1986). *Lexical semantics*. New York: Cambridge University Press.

Crystal, D. (2012). *English as a global language* (Canto classics edition.). Cambridge: Cambridge University Press.

Daller, H., Milton, J. and Treffers-Daller, J. (eds.) (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.

DeKeyser, R. M. (2013). Age Effects in Second Language Learning: Stepping Stones Toward Better Understanding. *Language Learning*, 63 (s1), 52–67. .

Dixon, R. M. W. and Aĭkhenval′d, A. I. (. I. (2002). Word: A cross-linguistic typology. Cambridge: Cambridge University Press.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.

Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford: Oxford University Press.

Elgort, I., Perfetti, C., Rickles, B. and Stafura, J. (2015). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and event-related brain potential (ERP) indicators of learning. *Language, Cognition and Neuroscience* 30, 506–528, DOI: https://doi.org/10.1080/23273798.2014.942673.

Elley, W. B. (1989). Vocabulary Acquisition from Listening to Stories. *Reading Research Quarterly, 24* (2), 174–187. doi:10.2307/747863.

Elley, W. B. (1997). *In praise of incidental learning: lessons from some empirical findings on language acquisition*. *Report Series 4.9.* (2 Feb 2018) Retrieved from http://cela.albany .edu/reports/inpraise/inpraise.pdf

Ellis, N. C. (2010). Implicit and Explicit SLA and Their Interface. In Leow, R. P., and Sanz, C. (eds.), *Implicit and Explicit Language Learning: Conditions, Processes, and Knowledge in SLA and Bilingualism*. Washington, DC: Georgetown University Press.

Eysenck, M. W. (1982). Incidental learning and orienting tasks. In Puff, C. R. (ed.). *Handbook of research methods in human memory and cognition*. New York: Academic Press.

Federation of Foreign Language Teachers in Finland SUKOL, The (2017). https://www.sukol.fi/ajankohtaista/tilastotietoa/tilastotietoa_kielivalinnoista. (2 April 2018).

Finnish National Board of Education (2011). *The framework of the Finnish National Certificates* *of Language Proficiency 2011: Order 24/011/2011*. Helsinki: Finnish National Board of Education.

Finnish National Board of Education (2016). *National core curriculum for general upper secondary schools 2015: National core curriculum for general upper secondary education intended for young people*. Helsinki: Finnish National Board of Education.

Fitzpatrick, T. (2007). Productive vocabulary tests and the search for concurrent validity in Daller, H., Milton, J. and Treffers-Daller, J. (eds.). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press, 116–132.

Frishkoff, G. A., Perfetti, C. A. and Collins-Thompson, K. (2011). Predicting robust vocabulary growth from measures of incremental learning. *Scientific Studies of Reading*, 15 (1), 71–91.

Garnier, M. and Schmitt, N. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System*, 59, 29–44.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21 (2), 303–317.

Hinkel, E. (2006). Current perspectives on teaching the four skills. *TESOL Quarterly*, 40 (1), 109–131.

Hirsh, D. (2012). *Current perspectives in second language vocabulary research*. Bern: Peter Lang AG.

Hirsjärvi, S. and Hurme, H. (2000). *Tutkimushaastattelu: Teemahaastattelun teoria ja käytäntö*. Helsinki: Yliopistopaino.

Hirsjärvi, S., Remes, P. and E. Sajavaara (2009). *Tutki ja kirjoita* (2nd ed.). Helsinki: Tammi.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: a measurement study. *Canadian Modern Language Review*, 61 (3), 355–382.

Huckin, T. and Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in second language acquisition* 21 (2), 181–193.

Hughes, A. (2003). *Testing for language teachers* (Second edition.). Cambridge: Cambridge University Press.

Huhta, A. (2017) 13.3.2017. A lecture of the course *Kielitaidon arviointi*. University of Jyväskylä.

Huhta, A. and Tarnanen, M. (2011). Kielitaidon arviointi tutkimusvälineenä ja tutkimuksen kohteena. In Kalaja, Alanen, P., Dufva, R. (eds). *Kieltä tutkimassa: Tutkielman laatijan opas*. Helsinki: Finn Lectura, 201–220.

Hulstijn, J. H. (2012). Incidental learning in second language acquisition. *The Encyclopedia of applied linguistics*. Hoboken, N.J.: Wiley.

Härmälä, M. (2014). *Englannin kielen A-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013*. Helsinki: Kansallinen koulutuksen arviointikeskus.

Ishii, T. and Schmitt, N. (2009). Developing an integrated diagnostic test of vocabulary size and depth. *RELC Journal* 40 (1), 5–22.

Jiang, N. (2004). Semantic transfer and development in adult L2 vocabulary acquisition. In P. Bodaards and B. Laufer (eds.), *Vocabulary in a second language*. Amsterdam: John Benjamins Publishing, 101–26.

Karlsson, F. (2004). *Yleinen kielitiede* (Uud. laitoksen 2. p.). Helsinki: Yliopistopaino.

Kekki, J. (2017). *Vocabulary testing in a Finnish EFL class : a comparison of teacher and student perceptions*. A Bachelor's thesis. University of Jyväskylä, Department of Language and Communication Studies. http://urn.fi/URN:NBN:fi:jyu-201702201482 (14 March, 2018).

Kirsner, K. (1994) Vocabulary acquisition: the implicit ins and outs of explicit cognitive mediation. In Ellis, N. C. (ed.). *Implicit and explicit learning of language*. London: Academic Press, 211–282.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern language journal* 73 (4), 440–464.

Kremmel, B., and Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly*, 13, 377–392. https://doi-org.ezproxy.jyu.fi/10.1080/15434303.2016.1237516.

Kuiper, K. (2004) in Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins Pub, 37–54.

Kumpulainen T. (2016). *Opettajat ja rehtorit suomessa 2016 – Lärarna och rektorerna i Finland 2016*. Helsinki: Suomen Yliopistopaino.

Lahtikallio, V. (2016). *Finnish EFL learners' size and depth of receptive vocabulary knowledge*.

Unpublished Pro Gradu thesis. University of Jyväskylä, Department of Languages. http://urn.fi/URN:NBN:fi:jyu-201603211897.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review* 59 (4), 567–587.

Laufer, B. and Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics* 16 (3), 307–322. [online] Retrieved in http://www.lextutor.ca/vp/laufer_nation_95.pdf.

Lin, C. and Hirsh, D. (2012). in Hirsh, D. (ed.). *Current perspectives in second language vocabulary research*. Bern: Peter Lang, 117–148.

Mann, S. (2016). *The research interview: Reflective practice and reflexivity in research processes*. New York: Palgrave Macmillan.

Martinez, R. and Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics.* 33 (3), 299–320. doi: 10.1093/applin/ams010.

Matsuoka, W., and Hirsh, D. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language,* 22 (1), 56–70.

McCray, G., and Brunfaut, T. (2018). Investigating the construct measured by banked gap-fill items: Evidence from eye-tracking. *Language Testing*, 35 (1), 51-73. doi:10.1177/0265532216677105 .

Meara, P. M. (1982). Word associations in a foreign language: a report on the Birkbeck Vocabulary Project. *Nottingham Linguistic Circular,* 11, (2), 29–37.

Meara, P. M. (2009). *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam: John Benjamins Pub. Co.

Meara, P. M. and Wolter, B. (2004). V_Links: beyond vocabulary depth. In Albrechtsen, D., Haastrup, K., and Henriksen, B. (eds.). *Angles on the English-speaking world: writing and vocabulary in foreign language acquisition (vol. 4)*. Copenhagen: Museum Tusculanum Press, 35–54.

Meunier, F. (2012). Formulaic Language and Language Teaching. *Annual Review of Applied Linguistics*, 32, 111–129. doi:10.1017/S0267190512000128.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston, MA: Heinle.

Nation, I. S. P. and Webb, S. (2011). *Researching and analyzing vocabulary* (1st ed.). Boston, MA: Heinle.

Nattinger, J. R., and DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Paribakht, T.S. and Wesche, M. (1996). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In Coady, J. and Huckin, T. (eds.) *Second language vocabulary acquisition: a rationale for pedagogy*. Cambridge: Cambridge University Press, 174–200.

Perfetti, C. A. (2007). Reading ability: lexical quality to comprehension. *Scientific Studies of Reading*, 11, 357–83.

Qian, D. and Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, 21 (1), 28–52.

Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2012) Piloting vocabulary tests. In Davidson, F. and Fulcher, G. (eds.). *The Routledge handbook of language testing*. London: Routledge, 307–320.

Riazi, A. M. (2016). *The Routledge encyclopedia of research methods in applied linguistics*. London: Routledge.

Rossiter, M. J., Abbott, M. L. and Kushnir, A. (2016). L2 Vocabulary Research and Instructional Practices: Where Are the Gaps? *Teaching English as a Second or Foreign Language*. 20 (1), 1–25.

Saenger, P. (1997). *Space between words: the origins of silent reading*. Stanford: Stanford University Press.

Sahlberg, P. (2007). Education policies for raising student learning: the Finnish approach. *Journal of Education Policy*, 22 (2), 147–171. doi: 10.1080/02680930601158919

Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins Pub.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching*

Research, 12 (3), 329–363. doi:10.1177/1362168808089921.

Schmitt, N., Ng, J. W. C., and Garras, J. (2011). The word associates format: Validation evidence. *Language Testing*, 28 (1), 105–126.

Schmitt, N., Schmitt, D. and Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing* 18 (1), 55–88. doi: 10.1177/026553220101800103

Schmitt, N. and Zimmerman, C. B. (2002). Derivative Word Forms: What Do Learners Know? *Tesol Quarterly,* 36 (2), 145–171.

Seabrook, R., Brown, G. D. A. and Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology* 19 (1), 107–122. doi: 10.1002/acp.1066.

Simpson-Vlach, R. and Ellis, N. C. (2010). An academic formulas list: new methods in phraseology research. *Applied Linguistics*, 31 (4), 487–512. doi: 10.1093/applin/amp058.

Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.

Singleton, D. (2000). *Language and the lexicon: An introduction*. London: Arnold.

Singleton, D. M. and Ryan, L. (2004). *Language acquisition: The age factor* (2nd ed.). Clevedon: Multilingual Matters.

Sobel, H. S., Cepeda, N. J. and Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology* 25 (5), 763–767. doi: 10.1002/acp.1747.

Stubbs, M. (1993). British traditions in text analysis. In Tognini-Bonelli, E., Sinclair, J. M., Baker, M., and Francis, G. (eds.). *Text and Technology: In Honour of John Sinclair*. Philadelphia: John Benjamins Publishing Co., 1–35.

Sökmen, A. J. (1997). Current trends in teaching second language vocabulary. In N. Schmitt and M. McCarthy (eds.). *Vocabulary: description, acquisition and pedagogy*. Cambridge: Cambridge U.P., 237–257.

Thomas, M. (2006). Research synthesis and historiography: The case of assessment of second language proficiency. In Ortega, L., and Norris, J. M. (eds.). *Synthesizing Research on Language Learning and Teaching*. Amsterdam: John Benjamins Publishing Co., 279–

298.

Tirri, K. (2014). The last 40 years in Finnish teacher education. *Journal of Education for Teaching*, 40 (5), 600-609, doi: 10.1080/02607476.2014.956545.

Tuomi, J. and Sarajärvi, A. (2009). *Laadullinen tutkimus ja sisällönanalyysi* (2nd ed.). Helsinki: Tammi.

Wall, D. (2012). Washback. In Davidson, F. & Fulcher, G. *The Routledge handbook of language testing*. London: Routledge.

Wallace, M. J. (1982). *Teaching vocabulary*. London: Heinemann Educational Books.

Waring, Rob 2002. Scales of vocabulary knowledge in second language vocabulary assessment. Kiyo. *Occasional Papers of Notre Dame Seishin University*. [online] (11 Mar 2018) http://www.robwaring.org/papers/various/scales.htm.

Webb, S., and Chang, A. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, 37 (4), 651–675. doi:10.1017/S0272263114000606.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2004). ‘Here’s one I prepared earlier’: Formulaic language learning on television. In Schmitt, N. (ed.). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins Pub, 249–268.

Ylisirniö, M. (2012). *To be or not to be: a case study of formulaic sequences in Finnish EFL textbooks for upper secondary school*. A Master’s thesis. University of Jyväskylä, Department of Languages. http://urn.fi/URN:NBN:fi:jyu-201211062866 (18 March, 2018).

Zareva, A. (2005). Models of lexical knowledge assessment of second language learners of English at higher levels of language proficiency. *System,* 33 (4), 547–562.

Zechmeister, E.B. and Shaughnessy, J.J. (1980). When you know that you know and when you think that you know but you don’t. *Bulletin of the Psychonomic Society*, 15 (1), 41–44. doi:10.3758/BF03329756.

von Zeeland, H. and Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: the same or different from reading comprehension?. *Applied Linguistic*s, 34 (4), 457–479.

Zhong, H. (2012). Multidimensional vocabulary knowledge: development from receptive to productive use. In Hirsh, D. (ed.). *Current perspectives in second language vocabulary research*. Bern: Peter Lang AG, 22–55.

Zhong, H. and Hirsh, D. (2009). Vocabulary growth in an English as a foreign language context. *University of Sydney Papers in TESOL*, 4, 85–113. Retrieved in http://faculty.edfac.usyd.edu.au/projects/usp_in_tesol/pdf/volume04/Article04.pdf. (2 February, 2018).

APPENDICES

Appendix 1: Framework for the Interview

Haastattelurunko / Framework for the interview

Taustakysymykset / ***Background questions***

- Kauan olet ollut opettajana? */ How long for have you been teaching?*
- Mitä kursseja opetat nyt? Oletko vaihtanut hiljattain kouluastetta? / *What courses are you giving now? Have you changed the school level recently?*

Sanaston luonne / ***Nature of the vocabulary***

- Mitä sinulle tarkoittaa sanaston oppiminen? */ What is vocabulary learning to you?*
- Mitä sanasta pitää tietää, ennen kuin sen osaa? */ What one has to know about a word before one has acquired it?*
- Mitä (lukion ensimmäisen vuoden) oppilaittesi tulee tietää sanasta, että he hallitsevat sen riittävän hyvin koulutasolleen? *What do your (USS1) students have to know about a word before they know it well enough for USS1 level?*
- Mitä tarkoittavat käsitteet sanaston laajuus ja sanaston syvyys? / *What do the terms breadth and width of vocabulary mean to you?*
  - Onko jompikumpi mielestäsi tärkeämpi? / *Is either aspect more important to you?*
- Mikä kaikki lasketaan sanaksi? */ What do you count as a word?*
  - onko esimerkiksi prepositiofraasit (esim. in the park, at school) sanoja vai onko se kielioppia? */ Do you think that for example prepositional phrases (e.g. in the park, at school) are words or are they grammatical constructions?*
- Miten opettaja mieltää multi-word unitsit (give up, it's raining cats and dogs, pull a fast one, bordering on arrogance vs ?bordering on modesty), sanaston ja rakenteitten väliset suhteet? Mikä on sanastoa ja mikä kielioppia? / *How does the teacher perceive multi word units (give up, it's raining cats and dogs, pull a fast one, bordering on arrogance vs ?bordering on modesty), the relationships between vocabulary and structures? Where is the line between vocabulary and grammar?*
  - Formulaic sequences on skaala: How do you do – semanttinen prosodia. Miten käsität nämä? / *Formulaic sequences form a scale: How do you do – semantic prosody. How do you perceive these?*

Sanaston opetus ***/ Teaching vocabulary***

- Miten paljon opetat sanastoa? Mitä metodeja siihen käytät? (Opettaako sen esimerkiksi käännöksen kautta, kontekstista, kuvilla…?)

- *How much do you teach vocabulary? What methods do you use? (For example, do you use translation, context, images…?)*
- Pitäydytkö oppikirjan sanastossa? */ Do you stick to the vocabulary of the course book?*
- Missä asemassa sanasto on mielestäsi kielenoppimisessa? Kuinka tärkeää se on verrattuna esimerkiksi kielioppiin? */ What's the position of vocabulary in language learning? How example is it for example in comparison with grammar?*
- Mikä on opettajan vastuu sanaston opetus-/oppimisprosessissa verrattuna oppijan vastuuseen? Kuinka paljon opettaja pystyy ylipäätään mielestäsi tekemään? */ What is the responsibility of the teacher in the vocabulary teaching/learning process compared to the students' responsibility?*
- Mikä sanaston opettelussa on haasteellista oppilaille opettajan näkökulmasta? / *What is challenging in vocabulary learning to the students from the teacher's point of view?*
- Mikä on mielestäsi paras tapa opettaa sanastoa? */ What is the best way to teach vocabulary?*
- Toistuuko sanasto paljon omassa opetuksessa? */ Does the vocabulary get repeated in your teaching?*
- Miten opettaa formulaic languagea? */ How do you teach formulaic language?*
-

Sanakokeet / *Vocabulary tests*

- Kuinka usein pidät sanakokeita? Pidätkö ollenkaan? Jos *ei*, niin alempana kysymyksiä sitä varten. / *How often do you give vocabulary tests? Do you give them at all? If* not*, see below for questions.*
- Millaisia sanakokeita tai pikkukokeita pidät? / *What kind of vocabulary tests or short-term achievement tests do you give?*
- Mitä varten pitää sanakokeita? (Onko ne esim. arviointia varten tai ulkoista motivointia varten?) / *Why do you give vocabulary tests? (Are they for example for assessment or for external motivation?)*
    - Miten paljon vaikuttaa arviointiin? */ How much do they affect the grade?*
- Kysytkö vain viimeisimpänä opiskeltuja sanoja, vai testaako myös vanhan sanaston osaamista? / *Do you test only the most recent vocabulary or also previous vocabulary?*
- Mikäli käyttää kappalekokeita (unit testing), niin hyväksyykö synonyymeja tai kiertoilmaisuja, vai vaatiiko juuri sen, mikä kappaleessa esiintyy? */ If you give unit tests, do you accept synonyms or roundabout expressions, or do you require the word that is in the unit?*
- Mitä niissä kysyt? */ What do you ask in them?*
    - tunnistaminen, kirjoitusasu, merkitys, morfologinen käyttäytyminen, syntaktinen käyttäytyminen, ääntäminen, kollokaatiot, rekisteri… / *recognition, orthography, meaning, morphological behaviour, syntactic behaviour, pronunciation, collocations, register…*
- Minkälaisia tehtäviä käytät? / *What kind of test item types do you use?*
    - Näytä tämän jälkeen tehtävätyyppimoniste / *after this, show the test item type print-out*

    - Miksi? Miksi ei käytä jotain tyyppiä / *Why? Why do not you use a certain type?*

Jos ei pidä sanakokeita: ***/ If does not give vocabulary tests:***

- Onko pitänyt? Jos, miksi lopetti? */ Did you ever give them? If yes, why did you stop?*
- Jos ei koskaan, miksei? / *If never, why not?*
- Miten motivoi tai arvioi oppijoitten sanaston opettelua ja osaamista? / *How do you motivate the learners to study vocabulary? How do you assess their lexical competence?*

Appendix 2: Test item types

Items A to H are adopted from Nation (2001: 344-360)

A) Write T if a sentence is true. Write N if it is not true. Write X if you do not understand the sentence.

1. We cut time into minutes, hours and days. ___

2. Some children call their mother Mama. ___

B) Choose one word from the list on the right to complete the sentence. Do not use the same word twice.

faint, acute, common, direct

1. A journey straight to a place is _______
2. An illness that is very serious is ______

C) Circle the choice that best gives the meaning of the underlined word.

chronic means

1. lasting for a long time
2. dissatisfied
3. to greatly decrease

D) Circle the choice that best gives the meaning of the underlined word.

There was no response

a. movement

b. answer

c. sound

d. sign

E) Translate the underlined words into your first language.

1. You can see how the town has developed. _________

F) Translating isolated words

1. developed ___________________________
2. monivuotinen ________________________

G) Translate the whole sentence

1. You can see how the town has developed.

H) Sentence completion where morphological skills are tested

1. You can see how the town has ________________. (*kehittyä*)

I) Vocabulary Knowledge Scale (based on Paribakht and Wesche 1996: 180). In this, learners are given a word which they respond to using the following statements:

1. I haven’t seen this word before.
2. I have seen this word before, but I don’t know what it means.
3. I have seen this word before and I think it means . . . .
4. I know this word. It means . . . .
5. I can use this word in a sentence.

If the learner answers to 5, they also have to answer to 4.

Appendix 3: Original Finnish interview quotes

(1) ”Se on pitkä prosessi sitä oppii toisaalta passiivisesti pikkuhiljaa ja sitten ne muuttuu jossain vaiheessa aktiiviseksi, tai osa muuttuu aktiiviseksi osaamiseksi.” - Saana

(1)“It’s a long process, you learn passively little by little, and then it changes at some point into active [knowledge], or I mean some of it becomes active.” - Saana

(2) ”Kai se tarkottaa sitä että oppilas saa keinoja millä pärjätä elämässä vieraalla kielellä. Erilaisissa tilanteissa, minkälaista kieltä tai mitä sanastoa missäkin tarvii ja näitä tämmösiä.” - Sinikka

(2) “I guess it [=learning vocabulary] means that the student acquires means of using the language so that they can get by using a foreign language. In different situations, [knowing] what kind of language and vocabulary you need and stuff like that.” - Sinikka

(3) ”Sanaston oppiminen on sitä että osaat sillä kielellä kommunikoida mahollisimman monessa rekisterissä ja monipuolisesti.” - Ulla

(3) “Learning vocabulary means that you are able to communicate in as many registers as possible and in various ways.” - Ulla

(4) ”Se tarkottaa sitä että oppilas oppii uusia sanoja, niitten eri käyttöyhteyksiä, lausumista , oppii tunnistamaan sen sanan kun se sen kuulee, oppii kirjottamaan ja osaa käyttää sitä eri lauseyhteyksissä ja ehkä hyvä oppilas myös tajuaa, ei tietenkään heti, mutta tajuaa milloinka voi käyttää sanaa bitch hoe ja woman.” - Siiri

(4) “It means that a student learns new words, different contexts of use, pronunciation, learning to recognise the word when they hear it, learning to write it and knowing how to use it in different contexts and maybe a good student also understands, of course not immediately, but when you can use words such as ‘bitch’, ‘hoe’ and ‘woman’” - Siiri

(5) ”Tärkeetä tietenkin on että oppii miten se lausutaan, mutta kaikki niitten suoritukset mitataan toistaiseksi kirjallisesti siellä, niin varmaan se on se kirjoittaminen.” - Kirsi

(5) “Of course it is important that they learn how [the word] is pronounced, but since for the time being all assessment [in the matriculation examination] is carried out in writing, it is the writing [that is most important]”. - Kirsi

(6) ”Tarpeellisinta olis se että pystyy ite tuottamaan mutta vähintäänkin se että pystyy tunnistamaan sen sanan ja mieluiten ilman kontekstia.” - Mira

(6) "Most useful would be that the learners themselves can produce [language], but at least they should be able to recognise the word and preferably without a context". - Mira

(7) "Se tosi tärkeä, se vaikuttaa kaikkeen: ymmärtämiseen ja siihen tuottamiseenkin, että voit ymmärtää vaikka kirjoitettua tekstiä ja puhuttua kieltä ilman hirveen syvällistä kieliopin tajuamista jos on laaja sanavarasto, eli on se tosi tärkeä." -Saana

(7) "[The vocabulary] is really important, it affects everything; from understanding the language to language production: you can understand written text and spoken language even without a deep understanding of the grammar if you have a large vocabulary, so it is really important." - Saana

(8) "Ehkä se laajuus on se mikä siinä vaiheessa [on tärkeätä], se syvyys on sellanen että vasta siinä lukion aikana päästään siihen, että ne oppilaat sais huomaamaan ettei se oo pelkkä se käännös." - Tuula

(8) "Maybe it's the breadth at that stage [that is essential]; depth is something that only during the USS we get into the point where we can make the students aware of the fact that the word is not just a translation." - Tuula

(9) "Kielen opettamisen haaste on just se että mitkä sanat kuuluu yhteen ja miten sen selität että miks näin on." - Tuula

(9) "The challenge of teaching a language is precisely how to explain why certain words go together." - Tuula

(10) "Kyllähän niitä tulee siellä teksteissä kaiken aikaa, että toki niitä on siellä, ja sitten ihan konkreettisen kommunikaation kautta tulee, mutta ei niinkään erillisenä tehtävänä." - Maria

(10) "[FS] appear in the texts all the time, so yeah they do exist there, and also in concrete communication, but there's no separate exercises." *- Maria*

(11) "Keskusteltiin kun se oli laittanut Quizletiin ku siinä oli luovuttaa suomennoksessa ja se laitto *give up* ja se laitto luovuttaa ja se Quizlet sano että *give in*, et eiks tääki tarkota, ja sit mä sanoin että ei tässä kappaleessa, et ne on eri kontekstissa eri ja sit mä annoin sille kaks lausetta ja annoin sille *luovuttaa* kaks eri merkitystä, ja sit se sai ite yhistää et kumpi on kumpi, ja sitten se itte ymmärs et toinen on *antaa periksi*, ja toinen *lopettaa tekemästä* tai *ei enää jaksaa*, et koska ei se muuten aukee jos mää vaa sanon et täs on tää ja täs on tää, et sit kaks erilaista tilannetta jotka näyttää paremmin sen eron." - Nina

(11) "[The student and I] discussed because they had put 'give up' to Quizlet because the Finnish translation said 'luovuttaa' *and Quizlet said that the right answer is 'give in', so they asked that doesn't 'give up' also mean that, so I told them that not in this unit and in this context, and then I gave them two sentences and gave them two different meanings to* 'luovuttaa' *and then they could connect which one is which, and then they understood that one means '*antaa periksi' *and the other means* 'lopettaa tekemästä' *or* 'ei enää jaksaa'; *if I had just told them that this one means this and that one means that, they wouldn't get it, so these two situations show it better." - Nina*

(12) "Katoin ne ihan turhaksi, menee ensinnäkin tunnilla aikaa, sit joudun korjaamaan niitä, ja sit oppilaat ei niihin kuitenkaan lukenut, että täysin waste of time and energy, että ne voi sen 15 minuuttia tehä jotain muuta.. sit siellä on se joku joka huiskasee sen kahessa minuutissa, ja joku tuhraa sitä 12 minuuttia, niin se toinen viheltelee 10 minuuttia ja oottelee, kun ei sit taas viitti tempasta sitä paperia.. että ennemmin tekee jotain muuta sen ajan" - Sinikka

(12) "I realised they were utterly useless, firstly we lose time from the class, then I have to correct them, and the students didn't study for them, so it was a complete waste of time and energy, they could spend those 15 minutes doing something else. So someone swishes through the test in two minutes, while other tinkers 12 minutes with it, so the first is whistling for 10 minutes and just waits, because I can't just go and take the [slower student's] paper from them. So I prefer them doing something else those 15 minutes." - Sinikka

(13) "Mun mielestä on vähän epäreilua jos on vähän heikompia [oppijoita] niin että "ylläri, tässä ois nyt tää yks sieltä edellisestä kappaleesta", totta kai siellä saattaa olla sellasia sanoja jotka on ollut jo aikasemmin, niin sit se pitäis tietää ku tää on jo opeteltu jo aikasemmin, mut joo, aika vähän ne menee päällekkäin". - Nina

(13) "I think it's a little bit unfair if there are some weaker [learners] in the group, and then you're like 'Surprise, here's this word from the previous unit', but of course sometimes there may appear words that we have studied earlier and they are supposed to know them but because we've had them before, but yeah, there's little overlap between the tests." - Nina

(14) "Monelle se on se ainut kannustin että testataan, suurin syy minkä takia testaan". - Saana

(14) "For many students it's the only motivator to study, that's my main reason for testing". - Saana

(15) "Hyväksyn jos sopii haluttuun yhteyteen, ehottomasti… mutta jos se tarkottaa eri asiaa tai periaatteessa sama sana mutta ei sovi siihen niin saa vähennystä siitä. Paljon ne kiertää ja mä aina yritän sitä palkita, varsinkin heikoimpien kanssa oon yrittänyt sitä että kirjottaa sinne kaiken minkä osaa ja pystyy, niin ne pystyy tosi hyvin ne heikkoimmatkin kiertämään näitä, sit menee

tosi vaikeen kautta se sana tai fraasi, mutta ei kai se, yritän kuitenkin pyrkiä siihen että se kommunikaatio on tärkeä" - Maria

(15) "I accept it if it fits in the context but if it means something else or is the right word but not the right register or context I deduce a bit. They use a lot of roundabout language and I try to award them, especially with the weaker learners I've tried that they write everything they know and can, so even the weaker can go around these really well, sometimes the word or phrase is expressed in a very complicated way, but hey, I always strive for that communication would be the main thing." - Maria

(16) "Pidän tosi lyhyitä sanakokeita […], siin on 10-15 kohtaa ja sitte siinä voi olla jotain käännöslauseita, oon huomannut että ne jotka osaa niin se on sama onko niitä 1 vai 100 ne opettelee ne kumminkin ja sit toiset ei opettele sitäkään, sen näkee vähemmästäkin." - Ulla

(16) "I give short vocabulary tests, […] there are 10-15 items and then there might be some sentence translations, I have noticed that for those who are good it does not matter if there are 1 or 100 items, they will study them in any case, and others will not study no matter what, so you can tell it with shorter tests". - Ulla

(17) "Mä toivon että ne oppilaat pääsis pois siitä ajatusmallista että ensin käännän tän sanan jotenkin suomeksi ja sitten sanon sen." - Siiri

(17) "I hope to rid the students from the mindset that language is first translating words into Finnish and then saying them." - Siiri