FINNISH FOREIGN LANGUAGE LEARNERS'
USE OF ENGLISH COLLOCATIONS

Master's thesis

Anni Vuorinen

JYVÄSKYLÄN YLIOPISTO

Tiedekunta – Faculty Humanistinen tiedekunta	Laitos – Department Kielten laitos
Tekijä – Author Anni Vuorinen	
Työn nimi – Title Finnish foreign language learners' use of English collocations	
Oppiaine – Subject Englannin kieli	Työn laji – Level Pro gradu –tutkielma
Aika – Month and year Toukokuu 2013	Sivumäärä – Number of pages 63 sivua

Tiivistelmä – Abstract
Tämän tutkimuksen tarkoituksena oli tutkia suomalaisten englannin oppijoiden käyttämiä verbistä ja sen objektina olevasta substantiivista muodostuvia kollokaatioita, tarkemmin sanoen kollokaatioiden osuutta kaikista oppijoiden tuottamista verbi – substantiiviobjektipareista sekä niiden sisältämiä virheitä ja negatiivista äidinkielen vaikutusta. Tutkimuksessa kollokaatioita tutkittiin kolmen eri taitotason kirjallisista teksteistä. Vertailun vuoksi tutkimukseen otettiin kollokaatioiden lisäksi mukaan myös ns. tavalliset, vapaat verbin ja substantiiviobjektin yhdistelmät (free combinations) sekä idiomit.

Sanaston merkitys vieraan kielen oppimiselle on korostunut viime vuosikymmeninä, ja tämä on havaittavissa myös sanastoa käsittelevien tutkimusten määrän kasvussa. On yleisesti tunnustettu, että merkittävä osa englannin kielen sanastosta koostuu erilaisista usean yksittäisen sanan muodostamista sanakeskittymistä, joita tutkijat ovat määritelleet ja luokitelleet eri tavoin. Kollokaatioita, jotka useimpien määritelmien mukaan lasketaan kuuluviksi em. sanakeskittymiin, ei kuitenkaan ole vielä kovin paljon tutkittu, etenkään suomalaisten kielenoppijoiden osalta.

Tutkimuksen aineistona oli 90 aikuisten, suomea äidinkielenään puhuvien kielenoppijoiden kirjoittamaa englanninkielistä tekstiä, jotka oli kirjoitettu osana Yleistä kielitutkintoa (YKI). Teksteistä 30 oli kirjoitettu osana perustason kielikoetta, 30 keskitason kokeessa ja 30 ylimmän tason kokeessa. Tutkimuksen tarkoituksena oli esitellä kollokaatioita ja niissä esiintyviä virheitä eri taitotasoilla, ei varsinaisesti vertailla tasoja toisiinsa, minkä vuoksi tutkimus ei sisällä tilastollisia merkitsevyyksiä, vaan tulokset on esitelty prosenttilukuina. Tutkimus on luonteeltaan laadullinen ja kuvaileva.

Kyky käyttää kollokaatioita ja idiomeja nähdään yleisesti osana hyvää vieraan kielen taitoa. Tämän vuoksi olettamuksena oli, että mitä korkeampi kielitaito tutkimukseen osallistuneilla kielenoppijoilla oli, sitä enemmän he käyttäisivät teksteissään kollokaatioita ja idiomeja, ja sitä vähemmän niissä ilmenisi virheitä. Myös äidinkielen vaikutuksen oletettiin vähenevän taitotason noustessa. Tutkimuksen tulosten mukaan kollokaatioiden suhteellinen osuus ei kuitenkaan noussut kielitaidon parantuessa. Päinvastoin, eniten kollokaatioita oli perustason teksteissä. Keskitasolla niitä oli selvästi vähemmän, ja ylimmällä tasolla suuntaus jatkui. Sen sijaan virheiden määrä väheni odotetusti taitotason noustessa. Odotusten mukaisesti myös niiden virheiden osuus, joiden taustalla tulkittiin olevan äidinkielinen ilmaus, väheni kielitaidon parantuessa. Ylimmällä tasolla äidinkielestä johtuvia virheitä ei ollut enää lainkaan.

Tutkimuksen tulokset olivat osittain ristiriidassa aiempien tutkimusten ja teorioiden kanssa, minkä vuoksi aihetta olisikin syytä tutkia lisää tulevaisuudessa.

Asiasanat – Keywords   foreign language learning, vocabulary, formulaic sequences, collocations
Säilytyspaikka – Depository Kielten laitos
Muita tietoja – Additional information

TABLE OF CONTENTS

1. INTRODUCTION

This study examines English vocabulary. The reason why vocabulary was chosen to be the topic of the study was simply its importance to communication. Vocabulary is an essential part of all languages as without it, it is practically impossible to communicate at all. It is, therefore, only natural that learning target language vocabulary constitutes a major part of foreign language (later: FL) learning, too. This time the focus is on written English-language vocabulary produced by adult foreign language learners whose mother tongue is Finnish. The data were collected within The Finnish National Foreign Language Certificate and texts were included from three different EFL skill levels.

The vocabulary of a language is a complex thing. Instead of being a static phenomenon that is easy to define, it is constantly changing according to the needs of its speakers. New words are being added and old ones are being left behind all the time as the world keeps on changing. In addition, contrary to popular beliefs, vocabulary is more than just single words. Instead, a surprisingly large proportion of e.g. English vocabulary is formed by different types of formulaic sequences, i.e. lexical units consisting of more than a single word. Estimations vary, but their proportion has been evaluated to be at least 50 percent of the vocabulary used by a native speaker (Schmitt and Carter 2004: 1). As the language proficiency of a native speaker is, at least to some extent, the goal of all foreign language learning, the use of formulaic sequences should have a central role also in FL learning and teaching.

This study examines Finnish FL learners' use of one subcategory of English formulaic sequences, collocations, in short written tasks. The data consisted of 90 texts written by adult learners of English representing three different skill levels: the basic level, the intermediate level and the advanced level. The texts were originally collected within The Finnish National Foreign Language Certificate, a language proficiency testing system for adult language learners which is available also in other foreign languages. The study is qualitative and includes descriptive analysis. As the following chapters will reveal, defining the linguistic phenomenon at issue is anything but simple. In short, the term collocation refers to words that are linked together either based on frequency alone or because of some sort of semantic restriction that is involved. For practical reasons, in this study, the focus has been narrowed down to collocations consisting of a verb and a noun in the object position (V + N). Although the importance of vocabulary, as well as that of useful word

groups, has been recognised also earlier on, both areas have been faced by an increasing interest among scholars during the past few decades, as analysing the research data has been facilitated considerably by the development of computer-aided corpus linguistics (Boers et al. 2006). However, even though studies on vocabulary and different multi-word units are nowadays easier to find, ones focusing on collocations are not that numerous, not to mention studies which focus on their production by non-native speakers (Nesselhauf 2003: 224). Therefore, it goes without saying that, so far, also Finnish FL learners' use of English collocations has been studied only a little. This is another reason why this was chosen as the focus of the study. Even though vocabulary and especially collocations turned out to be quite a complex topic to examine, the fact that this area had been studied so little guaranteed that the results would provide some interesting new information, which helped to overcome the difficulties during the research process.

Before moving on to the present study in more detail, it is useful to discuss its position in the field of foreign language learning research. The first part of the theoretical framework introduces the key vocabulary terminology involved in the study. A great deal of the next chapter is dedicated to the complexity of the linguistic phenomenon of formulaic sequences and collocations. During the research process, it turned out that there is no generally accepted definition or categorisation of these two areas of language. For this reason, it was considered important to introduce a few different alternatives before presenting the choices made in this study. The second part, in turn, focuses on foreign language vocabulary teaching and learning. It briefly discusses how the role of vocabulary in foreign language learning has changed over the years and presents the famous Lexical Approach by Lewis (1993) that underlines the importance of different types of multi-word units. The role of the learner's mother tongue to his/her foreign language learning is also discussed. Finally, at the end of the background section, before moving on to the details of this study, there is a chapter that presents some interesting and relevant earlier research related to the topic.

2. VOCABULARY

The main focus of the present study is on collocations, a specific aspect of vocabulary. In other words, we are dealing with words. But what exactly is a word and how are

collocations connected to them? The basic concepts involved in the present study are introduced in the first part of the background section, starting with the widely used term *word* and moving on to the more specific concepts of *formulaic sequences* and *collocations*. In addition, the relationship between vocabulary and grammar is also briefly discussed.

2.1. Word – a complex phenomenon

When asked to define a 'word', most people would probably find the answer quite obvious - a word is a single set of letters separated from other sets by a small space on both sides. There are many factors supporting this kind of thinking, for example dictionaries, which, according to Moon (1997: 40), present vocabulary "as a series of headwords or individual lexical items". Although the above approach to lexicon is, without a doubt, practical, exploring vocabulary a little deeper reveals the complexity of the phenomenon since, as Moon further points out, the unit to which the term 'word' refers is in many ways arbitrary.

It is safe to assume that all language experts would agree with Thornbury (2002: 12) in that "a word is a more complex phenomenon than at first it might appear". According to him, the clear definition of 'word' is made difficult by the huge variability of their nature. Words can vary in their function, i.e. the meaning of a word can be grammatical or its role can be more of an information carrier. In addition, one word can have numerous different forms. New words can be formed by adding elements to single words (e.g. *understand – **mis**understand*), or by combining them to one another (e.g. *police* and *man – policeman*). In addition, a number of words can be joined together to form combinations that function similarly to single words, and words may have the tendency to co-occur with certain other words. As if this was not complicated enough, Thornbury continues the list of the characteristics of words by presenting the possibility of different words having similar meanings, and how similarity in meaning does not guarantee that different words can be used in the same situations or for the same effects, thus referring to the main focus of this study, collocations. Let us, however, continue with the complex nature of words a little more before taking a closer look at them.

As the previous paragraph presents, it is not at all simple to decide what to include in the category of 'words', a problem which Thornbury (2002: 2-3) illustrates with the help of some of the more complex vocabulary items: Should

homonyms be counted as one word, although they have several different meanings (e.g. *skate*, which can refer to a type of fish or to gliding on ice)? How about all the different inflectional forms that can be derived from a single source word (e.g. *horse – horses*, *run - running*)? Should we handle compound words as one or two words (e.g. *high school* or *railway station*), and does it matter whether or not there is a hyphen between the constituent words (e.g. *eye shadow* and *eye-opener*)? What about idioms (e.g. *to be back to square one*)? Or phrasal verbs, the parts of which may be located far from each other in a sentence (e.g. *She **let** her whole family **down**.*)? The list could go on and on. Due to the complexity of defining a 'word', scholars have introduced other terms, such as a lexeme and a word family, to clarify the picture.

However, the purpose of this chapter is not to solve the problems concerning the definition of 'word', but to briefly introduce the complexity of the phenomenon. Instead, the main focus is on the vocabulary categories relevant to the present study, i.e. those of formulaic sequences and collocations, which will be discussed next. Before moving on, it is, nevertheless, worth pointing out that in this study, the terms 'word' and 'vocabulary' are understood in their broader senses. In other words, units consisting of single vocabulary items are not the only ones included.

2.2. Formulaic sequences

A remarkable proportion of vocabulary consists of units formed by more than one word. It is, of course, impossible to give an exact figure, and the results vary e.g. according to the definition used and the type of language studied, but they can be said to form roughly about half of the vocabulary used by a native speaker (Schmitt and Carter 2004: 1). However, although the existence of the units has been generally acknowledged, the terminology related to the linguistic phenomenon at issue is far from being unanimous. Over the years, researchers have used several different terms to refer to formulaic language in its different forms. Wray (2002: 9), for example, found over fifty examples of the variety of terms, some of them being 'chunks', 'collocations', 'conventionalised forms', 'formulaic speech', 'formulas', 'holophrases', 'multiword units', 'prefabricated routines' and 'ready-made utterances'. In addition to the different choices of terminology, the following paragraphs present some examples of how leading researchers of the field define and

categorize the linguistic phenomenon in question, two problems which scholars have solved differently as well. The problems of choosing the most suitable terminology to refer to the phenomenon are acknowledged by the researchers themselves. The situation is well described by Moon (1998: 2) as follows: "There is no generally agreed common vocabulary. Different terms are sometimes used to describe identical or very similar kinds of unit; at the same time, a single term may be used to denote very different phenomena."

Moon (1998: 2) herself uses *fixed expression* as a general umbrella term to refer to a variety of phrasal lexemes, phraseological units and multi-word lexical items, i.e. units of two or more words that are holistic, in which she includes "frozen collocations, grammatically ill-formed collocations, proverbs, routine formulae, sayings, similes" and idioms, although admitting that many of them are not fixed after all. To narrow down the focus, Moon (1998: 2-3) has chosen to exclude e.g. all phrasal verbs and foreign phrases from her study, even though they would otherwise be included in her definition of the phenomenon examined.

Schmitt and Carter (2004) also discuss suitable terms to best refer to formulaic language. They have rejected the term *formula*, since it, according to them, often refers to formulaic strings that have "idiosyncratic conditions of use" and, therefore, is not the best possible term to refer to the overall phenomenon. The term *lexical phrase* by Nattinger and DeCarrico (1992) was not accepted either, as it, according to Schmitt and Carter, unnecessarily for a cover term, connects formulaic strings with functional language use. Finally, Schmitt and Carter (2004: 4) end up having two equally suitable candidates: *phrasal lexical item* and *phrasal lexeme*. Still, eventually, their choice is *formulaic sequence* (FS) and its definition by Wray (2002: 9), which defines a formulaic sequence as follows:

> 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.

Read and Nation (2004: 24-25) have also chosen to use the term formulaic sequence, but challenge the above definition by Wray (2002: 9). According to them, if the focus of attention are sequences which are "stored and retrieved whole from memory at the time of use", researchers face a challenging task, as different individuals do not automatically store and retrieve the same sequence identically, nor does the same individual on all occasions. In addition, they suggest that only a small minority of

sequences are always formulaic. Therefore, they present that the criteria used to determine whether a sequence is formulaic or not should take into account the "features that are present in each particular use of a possible sequence" (Read and Nation 2004: 25). As a conclusion, Read and Nation (2004: 25-27) suggest that, as formulaic sequences cover a huge variety of linguistic phenomena, the best alternative is to modify the definition of the term to fit the aims of each individual research.

Sinclair (1991: 109-114) approaches the phenomenon of formulaic sequences through the concepts of 'open-choice principle' and 'idiom principle'. In his view, language conveys meaning according to these two different principles, neither of which is enough to cover all language on its own. The open-choice principle sees language as "the result of a very large number of complex choices" the only restraint being their grammaticalness (Sinclair 1991: 109). He calls this 'the normal way of seeing and describing language', typical to e.g. grammars. The idiom principle, on the other hand, takes into account the unrandomness of co-occurring words, as, like the existence of collocations suggests, grammaticalness is not the only factor affecting our choice of words. According to Sinclair (1991: 110), language is affected by e.g. the world around us. In other words, things that co-occur in the real world are more likely to co-occur in language, too. Another thing he sees influencing language is register. However, as Sinclair further explains, even these two combined with the open-choice principle are not enough. The main idea in Sinclair's idiom principle is that language contains numerous "semi-preconstructed phrases" which look like they are formed by several pieces but which at the moment of use actually behave as if they were single words. Sinclair (1991: 110) explains this partly by the fact that people naturally tend to economise on the effort they make and by demanding nature of communicating with someone in real time, but reminds that the phenomenon deserves much more attention than it has received in the field of linguistics which, in Sinclair's (1991: 110) view, is dominated by the open-choice principle.

Even though Sinclair (1991: 110-111, 115) presents two-word units such as 'of course' as the most simple form of the idiom principle, he underlines that the phenomenon is more complex and more common than has traditionally been acknowledged. In fact, taking into account all realizations of the principle, including collocation, Sinclair (1991: 112) states that the idiom principle is "at least as important as grammar in the explanation of how meaning arises in text".

Antunović (2007: 28-29), who studied Croatian translation of Swedish lexical phraseological units, uses a classification where collocations, idioms and formulas are all separate linguistic categories, and defines them as items that for various reasons are not considered compositional, but rather holistic units. In her study, she further defines collocations as multi-word lexical units that habitually co-occur, and which, in contrast to idioms, consist of lexical units that are all also semantic constituents of the sequence at issue. Formulas, on the other hand, she sees as units which are partly similar to idioms in that some of their lexical constituents may not always reflect their original meaning in the sequence, but which, according to Antunović (2007: 28-29), are syntactically more complex, "sentence-like phraseological units, such as proverbs, slogans, sayings, winged words, similes, commandments and maxims, etc."

As the above examples indicate, in addition to using a variety of cover terms for the overall phenomenon, different scholars also define formulaic language in different ways. Similarly to the difficulties of defining a word, a remarkable proportion of the difficulties in defining formulaic sequences is due to their diversity, as Schmitt and Carter (2004: 3) point out, since formulaic sequences can vary, for example, in length, fixedness, and the purpose they are used for. Although the exact definition of formulaic sequences varies from researcher to researcher to fit the needs of specific studies, some generalisations can be made about the characteristics typically present in the phenomenon, even though they do not necessarily apply to each individual formulaic sequence.

Moon (1997: 44) presents three criteria with the help of which holistic multi-word items (here: formulaic sequences) can be distinguished from other lexical items consisting of more than one word, i.e. the concepts of institutionalisation, fixedness, and non-compositionality. By institutionalisation she refers to the extent a formulaic sequence is conventional in the language examined, in other words, whether or not it recurs or the language community regularly considers it as a unit. Fixedness, on the other hand, stands for the frozenness of the sequence of words studied, i.e. does the sequence have the capacity to inflect as a whole or when divided into pieces, and whether or not it can vary e.g. in terms of its component words or their mutual order. Non-compositionality, in turn, is explained by Moon (1997: 44) as the characteristic of formulaic sequences to have unitary meanings that cannot be straightforwardly interpreted from the meanings of their constituent words. All in all, it is important to see, as Moon further reminds, that none of these features

are absolutes, but rather variables which can be present in a sequence in different degrees, and that each formulaic sequence consists of an individual combination of them.

As the earlier discussion on the terminology and definitions used by different researchers presents, formulaic sequences can be divided into several different subcategories, including, for example, idioms, proverbs, sayings, phrasal verbs, compound nouns, and fixed and semi-fixed expressions. Since scholars use different terms for the overall phenomenon, in addition to defining it differently, there has been disagreement in terms of the further classification as well. Some researchers argue that even single words and morphemes can be seen as formulaic sequences, whereas some draw the line e.g. to the expressions the meaning of which can be derived from that of their constituents, in other words leaving out e.g. idioms (Schmitt and Carter 2004: 4; Jiang and Nekrasova 2007). In most cases, idioms are, however, included. Moon (1997: 44-47), on the other hand, uses "multi-word item" as an umbrella term to compounds, phrasal verbs, idioms, fixed phrases and prefabs, but underlines at the same time that overlaps between the different categories are inevitable.

In the present study, the term *formulaic sequence* has been chosen as the term referring to the overall phenomenon of lexical units consisting of more than a single unit. The term has been chosen over e.g. the term 'multi-word unit' or 'chunk', as it expresses better the non-arbitrariness involved. In addition, nowadays, it is the most established term used. The definition used in the present study leaves out the multi-word lexical units formed based on purely grammatical aspects, such as e.g. some verb tenses. Moreover, the definition of the term has been modified of that provided by Wray (2002: 9) to the direction of Read and Nation (2004: 24-25) as, in the case of the data examined in this study, the holistic nature of the sequences at the time of storage in and retrieval from memory can no longer be examined. The chosen term can, however, be seen broader than for example 'formulaic speech', which could be interpreted as referring only to spoken language, or 'collocation', which, in the present study, forms just one of the subcategories of the phenomenon.

Although the terminology, definitions and categorizations concerning the field of formulaic language may vary, different forms of collocation are in most cases included in the linguistic phenomenon in question. This is the case also in the present study. Collocations, the main focus of attention and, according to the definifition and

classification used in this study, a subcategory of formulaic sequences, will be discussed in more detail next.

2.3. Collocations

Similarly to formulaic sequences, also the definition of collocation differs from one scholar to another. Roughly, the points of view can be divided into two groups, one of which sees collocations as a realization of co-occurrence of words on the basis of frequency alone, whereas the other requires that some form of semantic restriction must be present. Often the two approaches are combined, and the examined collocations are categorized further according to the degree of restriction involved.

According to Moon (1998:26), collocations typically refer to two or more words that co-occur repeatedly or statistically significantly and which may or may not be linked semantically. She further describes collocations as "the surface, lexical evidence that words do not combine randomly but follow rules, principles, and real-world motivations". Moon (1998: 26-27) divides collocations into three different groups. The first group she describes as the simplest kind as the basis of the collocation lies in semantics: words belonging to the same semantic field often co-occur the same way as the things they refer to do in the real world (e.g. *school – pupil – teacher – learn*). In her opinion (Moon 1998: 27), this kind of collocation can be useful, as the collocates may be of help to the language user in case there is polysemous vocabulary involved and the topic and context of the language use is not otherwise clear. The present study does not focus on this kind of collocation. In the second group, Moon (1998: 27) includes collocations formed by words that require being associated with a word that is not only semantically linked but also belongs to a lexicogrammatically limited group. The examples she gives include the word *rancid* which typically co-occurs with the words *butter* or *fat* and words referring to foods that contain either of the two. Another type of collocation belonging to the same group is a word that does not contain a specific meaning without the collocation of particular words, such as *face the truth/facts/problem* as opposed to the typical meaning of the word *face,* and e.g. verbs that require "certain kinds of subject or object: 'animate', 'liquid', 'vehicle', and so on: for example, the verb *drink* normally requires a human subject and a liquid as object" (Moon 1998: 27). Moon's third group of collocations are syntactic collocations formed by e.g. verbs, adjectives or their nominalizations which need to complemented with e.g. a certain particle. According to Moon (1998: 27), this group of collocations is not very

far from other kinds of word combinations that occur repeatedly in text, meaning that they have been formed according to grammatical rules and are used frequently, but do not necessarily share the independent and holistic nature she otherwise connects to collocations.

Nattinger and DeCarrico (1992: 36-37) also provide their definition of collocations, which they distinguish from lexical phrases, such as *how do you do?* and *for example,* their main focus of attention. To them, collocations are "strings of specific lexical items, such as *rancid butter* or *curry flavor*" which co-occur more frequently than can be explained by pure chance. What separates them from lexical phrases, which Nattinger and DeCarrico (1992: 36-37) consider a subcategory of collocations, is that, in contrast to lexical phrases, collocations as such do not have specific pragmatic functions.

The Oxford Collocations Dictionary for Students of English (2002: vii) emphasizes the role of idiomaticity in its definition of collocations, and defines collocation simply as "the way words combine in a language to produce natural-sounding speech and writing", as opposed to unidiomatic but otherwise grammatical combinations of words. This aspect is one of the focuses of attention also in this study. Unlike Moon (1998), the dictionary does not divide collocations into three but two different categories: *word collocations* and *category collocations*. In the dictionary, word collocations refer to collocations in which none of the specific collocating words can be replaced by their synonyms without destroying the collocation, whereas category collocations mean collocations where a certain word can collocate with "any word from a readily definable set", e.g. that of nationalities (The Oxford Collocations Dictionary for Students of English (2002: ix).

In spite of the above division, or in addition to it, more attention is paid in the dictionary to the way collocations vary in terms of their fixedness. According to The Oxford Collocations Dictionary for Students of English (2002: vii), the fixedness of word combinations varies from totally free, in other words fully predictable combinations formed purely on grammatical terms, to totally fixed and idiomatic, i.e. opaque idioms, with collocation covering everything between the two. The dictionary uses the verb *see* as an example: *see a man* is a totally free combination and, therefore, not a collocation, nor is the idiom *not see the wood for the trees*. In contrast, *see a film* can be called a 'weak' collocation, *see a doctor* a 'medium strength' collocation and *see danger/reason/point* a 'stronger' collocation (The Oxford Collocations Dictionary for

Students of English 2002: vii). The dictionary, thus, shares the idea of further classification based on the restriction involved that was presented already earlier.

The Oxford Collocations Dictionary for Students of English (2002: vii) underlines the importance of English collocations to language learners by reminding that the phenomenon is present everywhere in natural language use, both in spoken and written English. The dictionary points out that using the correct collocations makes the language sound a lot more like that of a native speaker even though there would not be any causes of misunderstandings involved. Moreover, it emphasizes that, in addition to making language sound more native-like, the correct use of collocations also makes it less ambiguous, and, similarly to Moon (1998), supports the statement with the numerous, partly overlapping meanings common to many of the English words, the correct one of which is always determined by the context, i.e. the collocating words.

3. FOREIGN LANGUAGE VOCABULARY TEACHING AND LEARNING

The previous chapter shed some light on the terminology used to refer to the linguistic phenomenon examined in this study and the various perspectives researchers have taken on the issue. This chapter concentrates on the expanding role of vocabulary – and, thus, also that of collocations - in teaching and learning foreign languages. Traditionally, the focus has been on grammar but, in recent decades, studies have shown that, in fact, a lot of learners' foreign language proficiency relies on the size of their vocabulary and their ability to use it correctly. Let us first take a closer look at how the relationship between grammar and vocabulary in FL learning has been seen in the past.

Traditionally, foreign language teaching has considered grammar and vocabulary separate areas of language, and, to some extent, this division remains even today. The traditional view, which underlined the importance of grammar at the expense of vocabulary has, however, recently been challenged, as the more modern approaches to foreign language learning put more emphasis on the role of vocabulary in increasing language proficiency than earlier ones. After all, nowadays, the focus is on communication skills.

In the past, grammatical structures were made the priority of foreign language teaching for example in approaches like the Direct Method and audiolingualism (Thornbury 2002: 14). As the focus of these methods was on

grammar, and learning the structures could have been distracted by too many additional linguistic elements, teaching vocabulary was often kept limited, Thornbury (2002: 14) explains, and continues that the vocabulary taught was mainly simple and chosen for its suitability in terms of the grammatical structure being taught. It was not until the rise of the communicative approach to foreign language teaching in the 1970s, he claims, that the importance of vocabulary and its role in carrying meaning started to be fully acknowledged, which in turn led to vocabulary being taught as such, and not just as a by-product of teaching grammar.

Even though vocabulary has recently gained more ground in foreign language teaching, the truth is, as Thornbury (2002: 14) presents, that learning grammar often dominates foreign language learning even today. He explains this by the perceived productivity of grammar, as one grammatical rule can create a great variety of sentences, as opposed to the characteristics of vocabulary items to simply add information.

However, as Thornbury (2002: 4, 14) mentions, the traditional juxtaposition between grammar and vocabulary has recently become blurred, as the findings of corpus linguistics have underlined the importance of the interdependence of the two, emphasising the role of frequent words and multi-word items in foreign language learning. The developments arising from the findings of corpus linguistics include the lexical syllabus, which Thornbury defines as a syllabus that, in contrast to the traditional grammar-based syllabuses, is based on words appearing with high frequency both in spoken and in written English. In addition to the development of the lexical syllabus, corpus studies have given unprecedented rise to the acknowledgement of the important role of different multi-word units in language acquisition process and in generating fluency (Thornbury 2002: 14). As Thornbury (2002: 14) further explains, the above developments have played a key role in the recognition of the importance of vocabulary in language learning, which can be seen in the increasing amount of attention given to e.g. frequent words and collocations even in coursebooks following a grammatical syllabus.

In addition to being challenged by the ideology of a lexical syllabus, the traditional grammar-based syllabus has been faced by the development of a lexical approach to language teaching. While resembling the lexical syllabus in that it underlines the importance of vocabulary and individual high-frequency words rather than grammatical structures, a lexical approach to language teaching also emphasises the role of multi-word units, as the basic idea behind the approach is that the lexicon

of a language always consists of a greater number of items than there are individual words in that language (Thornbury 2002: 112; Lewis 1997: 7). Moreover, a lexical view of language focuses on meanings, which, the approach argues, are mainly encoded in vocabulary items (Thornbury 2002: 112). According to Thornbury (2002: 112), the principles of the approach are based on the idea that syllabuses should be designed around useful meanings, the most frequent of which are conveyed by high-frequency words. Furthermore, the fact that words have a tendency to co-occur with certain other words explains the importance of learning frequent co-occurrences, i.e. frequent multi-word combinations such as collocations. Such a syllabus based on meanings following the above principles of a lexical approach is called a semantic syllabus. (Thornbury 2002: 112).

As far as this study is concerned, a particularly important aspect of the development of a lexical approach to language is the acknowledgement of the significance of different multi-word units both to the process of language acquisition and to fluency. Scholars now widely agree that a remarkable proportion of early language learning is based on learning unanalysed multi-word combinations, or chunks, as single units that are only later on analysed into their component words, Thornbury (2002: 14) explains, and adds that using these unanalysed chunks creates fluency not only in the beginning of language studies but also throughout the language learning process by reducing the processing time needed compared to forming all language from scratch at the moment of use, thus agreeing with Sinclair's idiom principle discussed earlier.

Perhaps the most famous theory based on the principles of a lexical approach to language teaching is the Lexical Approach by Michael Lewis (1993), which will be discussed briefly in the next section.

3.1. The Lexical Approach

The Lexical Approach by Lewis (1993) is perhaps the most famous theory underlining the importance of formulaic sequences in foreign language learning. Although the theory does not involve explicit teaching of these multiword units to help learners store them in memory, it does encourage teachers to raise learners' awareness of the sequences in order to make the acquisition process more efficient both inside and outside the classroom (Boers et al. 2006).

Lewis' Lexical Approach understands language as a phenomenon that, contrary to traditional views, does not consist of grammar and vocabulary, but rather of prefabricated multi-word chunks, which can be combined to produce longer extracts of coherent language (Lewis 1997: 7). As Lewis (1997: 7-11) explains, the approach divides chunks into four basic categories only one of which contains single words, the others being collocations, fixed expressions and semi-fixed expressions. Although the category of single words is the most familiar one and contains a far greater number of items than the others, Lewis (1997: 8) emphasises that, in the end, a remarkable proportion of the lexicon is formed by the other three categories.

While recognising that it is fairly common that teaching materials present single words, e.g. adjectives, with the prepositions needed (e.g. *relevant to*), Lewis (1997: 9) points out that the Lexical Approach goes further than that by preferring longer chunks that, in addition to being possible, are also highly likely in English (e.g. *relevant to our discussion / problem / needs*). He argues that these kinds of complete, fully contextualised phrases are the form in which the mental lexicon is stored in mind, and that, earlier, in order to facilitate learning, teaching has wrongly had the tendency to break them down into too small chunks. In addition, Lewis (1997: 12) also criticises traditional language teaching of treating all sentences similarly, with no reference to their usefulness in real language use.

Although acknowledging the importance of single words, the key areas of interest in the Lexical Approach are collocations, and fixed and semi-fixed expressions, i.e. fully fixed phrases and phrases that are frame-like and contain 'slots' that allow a limited amount of variance from their fillers (Lewis 1997: 9). Even though some expressions of these types, e.g. greetings, have been recognised in language teaching for a long time, Lewis (1997: 10-11) claims that they have often been treated as somewhat unimportant in terms of learners' needs and can, in many cases, be easily criticised as dated or unrepresentative of the language used by native speakers in reality. However, modern research has shown, as Lewis (1997: 11) points out, that the language we use is not as original as we think, and that prefabricated expressions provide a large proportion of both our spoken and our written communication.

According to Lewis (1997: 11) the most significant category in the Lexical Approach is not that of fully fixed expressions, but the varied group of semi-fixed expressions. According to him, it is this part of the lexicon that has proven the traditional distinction of language to vocabulary and grammar to be clearly

oversimplified. Without denying the importance of grammar, Lewis (1997: 11, 14) argues that such a division is artificial as, in addition to "'fixed' vocabulary and 'generative' grammar", language contains items that form a continuous continuum of fixedness between the two. Collocations, the main focus of this study, form a large part of this so-called continuum.

The Lexical Approach puts the main focus on communication, which, in Lewis' (1997: 15) terms, should be "at the heart of language and language learning". As communication consists of conveying meanings, he further argues that the focus is naturally on vocabulary, the most important part of language carrying meaning. Lewis (1997: 16) continues that in conveying meaning, making a grammatical error rarely causes as much damage as lexical errors, which are likely to make understanding more difficult, prevent it completely, or, in some cases, even offend the other person we are communicating with. The relationship between grammar and lexis is explained by Lewis (1997: 15) as follows:

> Grammatical knowledge permits the creative re-combination of lexis in novel and imaginative ways, but it cannot begin to be useful in that role until the learner has a sufficiently large mental lexicon to which grammatical knowledge can be applied.

3.2. FL learning and teaching and formulaic sequences

The important role of formulaic sequences in the language of native speakers should guarantee them a solid position in FL learning and teaching. As mentioned also earlier, estimations on the exact proportion of formulaic language vary, but it has been evaluated to form at least 50 per cent of the language used by a native speaker (Schmitt and Carter 2004:1). Moon (1997: 48) estimates the amount of different multi-word items to be "many thousands", but, at the same time, acknowledges the impossibility of giving an exact figure, as their category does not have clear boundaries and it keeps on developing as the language itself develops.

Despite the fact that they are a natural part of the language of a native speaker, formulaic sequences are not always presented as such in FL classrooms. On the contrary, according to Moon (1997: 57), in teaching and learning a foreign language, they are often seen problematic. Carter (1987: 136) brings up the risks involved in this kind of approach to the phenomenon and argues that concentrating only on problems may lead to a situation where all idiomaticity and fixed expressions are given a problematic status. Moreover, as Moon (1997: 58) continues, the fact that

many formulaic sequences, especially metaphorical ones, are "marked, infrequent, and generally considered 'difficult' " can make the teacher present them to language learners as only receptive vocabulary that does not need to be paid as much attention to as some other lexical items, or worse - even set them aside completely. According to her, this may, to some extent, be explained by the fact that a language learner probably does not very often run across a certain individual formulaic sequence. Unfortunately, as Moon (1997: 58) further points out, this, together with the possible communicative errors caused by semantically or stylistically inappropriately used formulaic sequences, may lead to a situation, where both the teacher and learners avoid these kind of lexical items altogether. The above situation may be true especially in written language, as it does not provide the speaker with the same possibilities to repeat and re-form his/her message in case of a communicative error as spoken discourse.

Even though some individual formulaic sequences may be rare in terms of their relative frequency in a language, this is not the case for all of them. Moon (1997: 63) emphasises that some of them are, in turn, very frequent, which makes them very useful for FL learners as well. Nevertheless, as she continues, even a sequence that is rare still has an important role in language and it is, therefore, worth acquiring. This is true especially when reaching for native-like language skills, since FL speakers' ability to use and interpret formulaic sequences correctly is generally seen as a sign of language proficiency (Moon 1997: 58).

When dealing with elements consisting of more than a single linguistic element, it is clear that producing and using them correctly is more complex than that of single words, too. The next chapter discusses errors in terms of formulaic sequences and, in particular, collocations.

3.3. Errors and formulaic sequences

The reasons behind FL learners' errors are not always easy to find. Swan (1997: 161-162) presents the traditional model of categorising FL errors, where the errors are divided into two groups depending on the cause: "'interlingual' confusions, caused by interference or transfer from the mother tongue, and 'intralingual' confusions, caused by complexities in the second language itself". However, this is not the only possible categorisation. When needed, errors can be further categorised in a number of ways, depending on the point of view.

Errors in the use of formulaic sequences can be classified in different ways, too. Moon (1997: 58-60) divides them roughly to three different categories, i.e. formal, pragmatic and stylistic errors. According to her, formal errors may be caused by e.g. situations where a learner does not recognise that a certain expression is non-compositional and, incorrectly, tries to compose the formulaic sequence piece by piece relying only on the grammatical rules of the language. To pragmatic errors, she includes errors caused by e.g. the use of otherwise correctly formed formulaic sequence in an inappropriate discoursal context or in a situation where the learner has misunderstood the context in some way, whereas the source of a stylistic error, on the other hand, may be the "use of an excessively marked multi-word item – very rare, dated or overinformal – or in an inappropriate genre" (Moon 1997:60). In addition to the above three categories, Moon (1997: 58-59) brings up the group of lexical errors, errors caused by e.g. translating idioms word for word from one language to another, trying to make use of partly similar formulaic sequences in L1, or not knowing the syntactic "rules" concerning the use of formulaic sequences. Errors related to L1 will be discussed in more detail in the next section.

3.4. The two-sided role of L1 in FL vocabulary learning

Language affects our perception of the world. It is probably not exaggeration to claim that a person's mother tongue (later: L1), as well as all the other languages (s)he has learned later in life, have a huge effect even on the formation of his/her identity. Considering this, the idea that earlier languages influence a person's further language learning does not sound very surprising. Being the first language to be acquired, the mother tongue can easily be seen as the basis for all later language learning. Schmitt and McCarthy (1997: 2) see a learner's mother tongue as one of the most influental factors affecting FL vocabulary learning. According to them, the mother tongue can, depending on the distance between the two languages, either ease or hinder the learning process, as it "will determine whether a majority of L2 (here: FL) words are easy or difficult, and whether whole new knowledge systems (new alphabets, new sounds, and sound combinations, new syntactic notions like articles, phrasal verbs, or case endings) have to be mastered". Swan (1997: 156) is of the same opinion. He claims that L1 can either "support, fail to support or actively hinder" the person trying to learn or use FL vocabulary. According to him, there are three possible stages when this may happen: when acquiring new vocabulary, when

trying to recall and use vocabulary that has already been learnt or when trying to form a more complex lexical unit which has not been learnt as a whole, but in smaller pieces.

However, as it was said earlier, classifying FL learners' errors is not always simple. Even though Swan (1997: 161-162) presents the already mentioned traditional model of categorising FL errors into linterlingual and intralingual confusions, he also further points out that not all errors can be classified as easily. In his opinion, contrary to what has commonly been done, errors do not always have to be analysed as either interlingual or intralingual, but possibly also as a combination of the two, since a learner's L1 can have a considerable effect on how (s)he perceives the possible ease or difficulty of the FL. This is where Swan (1997: 163), similarly to Schmitt and McCarthy above, underlines the role of 'language distance'. According to him, in addition to the actual language distance, i.e. whether or not the two languages are related, and if so, to what extent, also the learner's perception of the distance affects the possible transfer between the two languages, which, in turn, has an effect on the degree of support or hinder it causes. In the present study, the two languages in question, Finnish and English, are not related, which is, therefore, supposed to decrease the amount of L1 transfer to be found in the data.

In the beginning of his/her FL learning process, with experience of only one language, the mother tongue, it is natural that a learner tries to make use of its rules also in FL, regardless of the possible language distance. Swan (1997: 166) calls this "the learner's equivalence hypothesis*"*, which, in its simplest version, allows the learner to think that the words of the two languages work semantically and grammatically in the same way, even though they look different. Another version of the hypothesis, which Swan describes as more reasonable than the first one, could be *"'Regard everything as the same unless you have a good reason not to'"*. According to Swan (1997: 167), this kind of thinking is natural, even indispensable in FL learning, and present especially in the beginning of the learning process.

Trying to make use of the rules of one language in communicating in another will, however, inevitably lead to problems. Formulaic sequences are in this sense especially problematic, since, as Swan (1997: 180) points out, compared to other aspects of language, formulaic sequences are particularly difficult to form and understand when relying on the help of the learner's mother tongue. According to him, it is practically impossible to choose a correct sequence or construct one by direct translation from L1, as "attempts to match the idiomatic quality of mother-

tongue formulae usually lead to error, and sometimes to absurd results*". (Swan 1997: 178). Moon (1997: 58) is of the same opinion, reminding that even in cases where the two languages in question, i.e. the learner's mother tongue and the target language, contain corresponding sequences, ones that do not differ at any level are rare to find.

There is, however, also another side of the story. Even tough the focus has been on errors, Swan (1997: 167) also underlines mother tongue's role in avoiding them and states that, in addition to errors, L1 can be held responsible also for a lot of correctness in the language produced by a learner. According to him, the effects of the mother tongue should not, therefore, be seen only negative, as having a L1 makes it possible for us to learn more languages without having to conceptualise the entire world all over again, as we did in our infancy with the mother tongue (Swan 1997: 168).

3.5. Previous studies on foreign language learning, formulaic sequences and collocations

Although FL vocabulary and formulaic sequences can be said to have faced a new wave of interest among scholars quite recently, both have been studied also earlier on. Koprowski (2005) studied the nature and usefulness of lexical phrases in English coursebooks by comparing three different contemporary ELT coursebooks, which all represented mainstream general English coursebooks and were designed for intermediate learners. Koprowski examined the lexical syllabus of each book by counting the lexical phrases found and dividing them into several subcategories: collocations, phrasal verbs, binomials, idioms, compound nouns, and fixed and semi-fixed expressions. In addition, he studied the usefulness of each lexical phrase with the help of a computerised corpus of over 330 million English words, with special attention paid to their frequency and range in the English language. According to the results, all three books contained many lexical phrases, and all subcategories were present in each book, although with different percentages. The general usefulness of the lexical phrases found, however, was rated very low. Many of the lexical phrases could not be found in the corpus at all. In addition, it was found that the more lexical phrases there were in a book, the less useful they were. What was especially interesting was that less than 1% of the phrases were shared by any two of the

investigated books and none of them by all three books. As a conclusion, the lexical phrases found in the three coursebooks seem to have been chosen quite arbitrarily.

Although the present study excludes teaching materials and Koprowski (2005) focuses only on monolingual English coursebooks, i.e. teaching material designed for FL learners with varying mother tongues, the above study still provides interesting background information on the position of formulaic sequences in modern-day FL teaching in general. Even though it is impossible to make any exact statements on the way formulaic sequences are chosen to and presented in the bilingual coursebooks designed for Finnish learners of English based on Koprowski's findings, it can be assumed that paying more attention to their general usefulness would have a positive effect on learning outcomes and, therefore, also make them more common in the language spoken/written by FL learners. This, on the other hand, according to the following study, has a positive effect on the perception of learners' language proficiency.

Boers et al. (2006) examined how the use of formulaic sequences affects English L2 learners' perceived oral proficiency and whether or not the fact that special attention is paid to formulaic sequences in teaching helps learners to add them to their active vocabulary. The participants were 32 college students. According to the results, the participants belonging to the experimental group, which received more explicit teaching on formulaic sequences, were evaluated more proficient and more native-like than the control group. Moreover, the number of formulaic sequences used by the students belonging to the experimental group was higher than that of the control group. The researchers, therefore, concluded that the use of formulaic sequences is connected to the perceived language proficiency of a L2 learner and that the more formulaic sequences the learner uses, the more native-like (s)he sounds. Both conclusions, thus, agree with e.g. Thornbury (2002) and Moon (1997). The experimental group's higher use of formulaic sequences compared to that of the control group, on the other hand, was interpreted as proof for the hypothesis that emphasising the noticing of these multiword units in teaching can increase learners' active use of them.

Jiang and Nekrasova (2007) studied the representation and processing of formulaic sequences by proficient non-native and native speakers of English. The study involved lists of formulaic and nonformulaic phrases, as well as nongrammatical phrases, from which the participants were asked to identify the correct formulaic sequences. The underlying goal was to find out whether or not

formulaic sequences are stored in and retrieved from memory holistically as single units, which the researchers aimed to find out by comparing the reaction times the participants needed to identify them from the nonformulaic or ungrammatical ones. According to the results, both native and non-native speakers of English responded to the grammatical items faster than to the ungrammatical ones. In addition, the participants made fewer mistakes with the grammatical formulaic sequences than with the grammatical but non-formulaic items. The results were, therefore, consistent with those of several previous studies done on the same research field, as they, too, suggest that formulas, like idioms, are stored and processed holistically as single units by both native and non-native speakers. However, it is worth reminding that the holistic nature of formulaic sequences is also one of their most controversial features (see e.g. Read and Nation (2004): 25-27) and that, for example, in this study, it is not seen as a characteristic needed to be present in each sequence in order to be considered formulaic. Nevertheless, as Jiang and Nekrasova's (2007) findings support the holistic nature of English formulaic sequences also when processed by non-native speakers, it can be assumed that presenting them as single units from the beginning might have its advantages in FL teaching.

Although studies on vocabulary and even those on different types of formulaic sequences are quite numerous, studies where the focus is collocations are more difficult to find, especially ones concentrating on the production of collocations by non-native speakers (Nesselhauf 2003: 224). In addition, as Nesselhauf (2003: 224) explains, the rare studies that can be found on the subject are often unsatisfactorily done, or they do not provide a clear definition of the kind of collocation they are investigating. Nesselhauf (2003) herself conducted a study which had a relatively big influence on this study. She examined English verb – object-noun collocations in free written production produced by advanced learners of English with German as their mother tongue. The collocations included were chosen on the basis of their phraseological features, instead of focusing only on their frequency, which is the case also in this study, as it will be explained later on. Moreover, similarly to what was done in the present study, she, too, further classified the combinations according to the degree and nature of restriction involved, resulting in the three categories of combinations used also in this study, i.e. free combinations, collocations and idioms (Nesselhauf 2003: 225-226).

While being similar to Nesselhauf's (2003) study in many ways, the data analysis of the present study was not as complex. Nesselhauf (2003) first

manually extracted all verb – object-noun combinations from the texts, then classified them into the three classes of combinations, and finally evaluated whether or not they were correct and acceptable according to the rules of the English language. Moreover, she also classified the mistakes found in the combinations and examined the role of the learners' L1 (German) in them. All these steps were present also in this study, but, instead of combining dictionaries, corpus analysis and native-speaker judgements, they were carried out in a more simplified manner. In addition, unlike in Nesselhauf's (2003) study, in this study the collocations were not further classified according to the degree of restriction, nor were the combinations categorised using as many degrees of acceptability. The reseach method used in the present study will be explained in more detail later on.

According to Nesselhauf's (2003: 237-238) results, producing collocations causes difficulties even for advanced-level learners of English. She, therefore, calls for explicit teaching of them, instead of settling for a brief introduction of the linguistic phenomenon. As, in reality, not all collocations can be explicitly taught, the focus should be on the ones that are acceptable and frequent in English in general, or in a specific register useful to the learner (Nesselhauf 2003: 238). In addition, as her results suggested that the collocations that are not shared by the learner's L1 and the target language are the ones that cause more difficulties, she argues that the role of L1 should be more acknowledged in FL classrooms. Another problematic feature in collocations was, according to Nesselhauf's (2003: 238) results, the restriction involved. However, unlike one could imagine, it was not the most restricted collocations that were the most difficult ones. Instead, the ones causing most problems were the less restricted collocations such as to *perform a task* that involve some restriction but, nevertheless, are close to being combined simply based on the semantic properties of their constituent parts.

4. AIMS AND RESEARCH QUESTIONS

The aim this study was to investigate Finnish foreign language learners' use of one category of English formulaic sequences, i.e. collocations, in short written tasks. As the term collocation subsumes a large number of subcategories, there was a need to narrow down the focus and the study has, therefore, deliberately been limited to cover only collocations consisting of a verb and a noun in the object position (later:

V+N, similarly to the already mentioned study by Nesselhauf (2003)). In order to present the use and nature of V + N collocations more clearly, other types of V + N combinations, in other words free V + N combinations and idioms containing a V + N combination, were included in the study as well.

The underlying goal was to outline Finnish foreign language learners' use of English collocations, an area which has not been studied that much earlier. Although, based on the chapter presenting the results of the study, it becomes evident that the three different EFL skill levels were, to some extent, compared to each other, this was not the number one purpose of the study. Instead, the different skill levels were included in order to give a more comprehensive account on the phenomenon than would have otherwise been possible. The fact that the study does not focus on the comparison of the three different skill levels as such is also the reason why it does not contain more detailed statistical comparisons between the levels, but, instead, settles for rough percentages.

To find out how common collocations are in Finnish EFL learners' texts, the V+N collocations found in the data were studied in terms of their proportion of all V+N combinations found. Moreover, attention was paid on their correctness and on the connection between the amount of mistakes in the collocations and the level of restriction involved, i.e. whether or not it mattered in terms of mistakes if the combination in question was a free combination, a collocation or an idiom. Finally, the role of the learners' mother tongue in the possible mistakes found in the V+N combinations was briefly examined.

The research questions of the present study are:

1. How do Finnish EFL learners belonging to different EFL writing skill levels use English V + N collocations in written tasks?
2. What is the amount of collocations used (related to all V+N combinations)?
3. What is the amount of mistakes found in the collocations?
4. Does the restriction involved have an effect on the amount of errors found in the V + N combinations?
5. What is the role of L1 in the errors found?

5. DATA AND METHOD

This study is qualitative and it includes descriptive analysis. The data consist of 90 written tasks collected within The Finnish National Foreign Language Certificate, a language profiency testing system designed for adult language learners which, in addition to English, is available also in a number of other foreign languages. The Certificate tests are divided into three levels: Basic, Intermediate and Advanced, all of which measure test-takers' language skills in five different categories: listening comprehension, reading comprehension, speaking, writing, and structures and vocabulary. The data used in the present study result from English writing tasks.

Provided that all five subtests have been passed, the Certificate provides test-takers with an overall grade between below 1 and 6, below 1 being the lowest and 6 the highest possible grade. Basic level test covers grades up to 2, Intermediate up to 4 and Advanced level the highest grades. In the present study the data consist of all three test levels, as the 90 written texts contain 30 texts from each level. As it was impossible to find out the grades given for each individual text, and there was too much variation within each test level, the three skill levels included in the study in question had to be based on the overall writing grades given to the test-takers, i.e. the grades that cover all writing tasks included in the test. As a result, the three skill levels included are the informants given grade 2, 4 and 6 as an overall grade for writing. In order to ensure that the data really represented three different skill levels, bearing in mind the possible variation of individual texts within each grade, it was decided not to include three consecutive grades.

Even though the tasks are different at each test level, they measure roughly the same communicative functions. As the ideal was that the only variable in the comparison in question would be the writing skill level of the informants, an effort was made to find tasks as much alike as possible. As a result, the data from all three levels consist of texts with given instructions, as opposed to writing about a free topic. In addition, at all levels the tasks include stating one's own opinion on a matter. At the basic level, the task was to write an informal message or a formal letter as a response to a message in a paper. In the intermediate level task, the informants were asked to write an informal letter as a response to a Letter to the Editor stating their own opinion on the matter in question, and in the advanced level, the task was to informally comment an article.

The data were analysed manually. First, all V+N combinations found in the texts were listed. After this, the combinations were categorised according to the level of restriction involved, or the lack of it, resulting in three categories: free combinations, collocations and idioms, partly following Nesselhauf's example. The definitions and the categorisation were based on those in her study. Free combinations cover verb-object noun combinations that have been formed purely based on their semantic suitability, whereas idioms include V+N combinations where neither of the two words could be replaced by its synonym without the expression becoming incorrect. Collocations, in turn, cover the combinations in which the amount of restriction is between that of free combinations and idioms. The collocations were examined using Oxford Collocations Dictionary for Students of English (2002) and Collins Cobuild English Dictionary for Advanced Learners (2001). Second, the number of free combinations, collocations and idioms were compared to that of all V+N combinations at each level to find out their individual proportions. Third, all V+N combinations were divided into correct and incorrect ones, and the incorrect ones were later categorised further based on the type of error found. A combination was classified incorrect if it contained one or several mistakes in one or more than one of the following elements: the verb or the noun in question (e.g. *said a letter* instead of **send** *a letter*, or *contact the chef* instead of *contact the* **manager**), or the possible articles (e.g. *find home* instead of *find* **a** *home*), prepositions (e.g. *ask help* instead of *ask* **for** *help*) and other determiners, such as possessive pronouns (e.g. *call you landlord* instead of *call* **your** *landlord*), attached to the combinations. In contrast, possible mistakes found in, for example, adjectives and other modifiers connected to the V + object N combinations were not included (e.g. *have a quite big problem* for *have* **quite a big** *problem*). Mistakes concerning the tense or the conjugation of the verb were also excluded from this study, as they, too, were not seen as mistakes in the V + object N combination. At this point, it is worth mentioning that drawing the line was extremely difficult and some of the errors and linguistic elements that were excluded from this study may be included in others. For practical reasons, however, a decision had to be made.

In detail, partly following the categorisation by Nesselhauf (2003), the mistakes found in the incorrect V + object N combinations were classified into eight categories: spelling mistakes (which include possible typing mistakes when the original hand-written texts were transcribed into electric form), article mistakes, preposition mistakes, mistakes concerning the choice of verb, mistakes concerning

the choice of noun, determiner mistakes (e.g. mistakes in possessive pronouns), mistakes in the choice of singular or plural form of noun, and mistakes that could not be clearly defined and categorised, the category of which is marked by question mark (?). To some extent, the categorisation followed that of Nesselhauf (2003), but some modifications were made in order to not make the analysis too complicated for the needs of this study. The eight categories were the ones that covered the mistakes involved, in addition to being simple enough to handle with the available resources. In all categories, except that of spelling mistakes, the category containing mistakes in the choice of singular of plural form, and the incomprehensible mistakes marked with ?, wrong forms included, in addition to the wrong choice of word, also the lack of needed forms and the use of excess forms. The classification of the errors was complicated, as well, since it was impossible to know whether or not e.g. the incorrect form *you* instead of *your* was in fact an error in the choice of determiner and not simply a spelling error. Once again, the line had to be drawn somewhere, and the cases where the lack of one or two letters, or using one or two excess letters, resulted in a recognisable English form, were not considered spelling mistakes.

After classifying the errors, the proportion of correct combinations was compared to that of incorrect ones within each category and the results of each skill level were studied with regard to those of the other levels. Finally, at the end, the focus was shifted purely into the incorrect combinations and the mistakes were examined in terms of whether or not they could be explained by mother tongue influence, in this case whether or not the existence of a similar expression in Finnish (or the lack of it) could be the reason behind the incorrect combination. This part of the analysis is done more or less superficially, purely based on the researcher's own native language skills, and it should, therefore, be viewed only as a rough outline.

6. FINNISH LEARNERS OF ENGLISH AND THEIR USE OF COLLOCATIONS

The correct formation and use of collocations and idioms are generally connected to a certain level of language proficiency. Thus, the underlying assumptions in this study were that the higher the overall grade for writing, i.e. the higher the test level, the higher the number – or proportion - of collocations and idioms found in the texts,

and the lower the number of mistakes found in them. Moreover, it was assumed that higher language skills would correlate with less negative L1 influence.

This chapter presents the results of the study. The results are presented according to the three skill levels involved in this study, in other words, the chapter begins with the results of the basic level, after which it will move on to the results of the intermediate level as well as those of the advanced level. Possible differences between the results of the levels are presented as they occur. Finally, at the end of the chapter, some attention is paid also to the participants' mother tongue and its role in the mistakes found. Let us start with the results of the basic level, in other words the overall writing grade 2.

6.1. Basic level

The results of the basic level refer to the results of the 30 texts written by the participants given overall grade 2 in the basic-level English writing test of The Finnish National Foreign Language Certificate. The texts chosen were all written as responses to other texts and in all cases the task included further instructions on the type of text to be written. Most of the texts were informal, but in order to gather 30 texts a few formal texts had to be included as well. The length of the texts varied a lot. The shortest text contained only 30 words and the longest one 163 words altogether, the average number of words being 88 words per text. For practical reasons, each group of letters separated from others by space was counted as a word, ignoring the complexity of defining a word in this case.

As the group of participants represented the lowest skill level involved in this study, the underlying assumption was that, compared to the overall number of combinations found, the proportion of collocations and idioms would be smaller and the proportion of incorrect combinations higher than at the intermediate and advanced levels, the results of which will be presented later on. Table 1 on the next page presents the overall number of verb – object noun (from now on: V+ object N) combinations found in the basic level texts, as well as their distribution into the three categories of free combinations, collocations and idioms.

Table 1. Distribution of V+ object N combinations in the basic-level texts.

	Number	Percentage
Free combinations	59	53%
Collocations	53	47%
Idioms	0	0%
Total	112	

As Table 1 above shows, the 30 basic-level texts contained 112 V + object N combinations altogether. When compared to the overall amount of words in the texts, the basic-level texts included one V + object N combination per 24 words. 59 of the combinations, 53% of them, were classified as free combinations. The category of collocations covered 53 combinations, which correspond to the percentage of 47% of all combinations found. Not surprisingly, considering the level of language proficiency in question, no idioms were found in the basic-level texts. The group of examples (1) below provides some examples of the basic-level combinations found:

(1) Free combinations

own a house   *handle a thing*

know the situation   *help people*

tell about a problem   *make marks*

talk about a situation   *buy clothes*

Collocations

get a good night sleep   *give a warning*

ring the bell   *open the door*

call the police   *find a way*

answer questions   *read a story*

In addition to the different types of V + object N combinations, the focus of this study was also on the proportions of correct and incorrect combinations and their possible variation between different skill levels. Table 2 below shows the overall number and percentage of correct and incorrect combinations in the basic-level texts,

as well as those in each of the three categories of free combinations, collocations and idioms.

Table 2. Correct vs. incorrect combinations in the basic-level texts.

	Correct		Incorrect	
	Number	Percentage	Number	Percentage
Free combinations (N=59)	26	44%	33	56%
Collocations (N=53)	28	53%	25	47%
Idioms (N=0)	-	-	-	-
Total (N=112)	54	48%	58	52%

As explained also earlier, a combination was classified incorrect if it contained one or several mistakes in one or more than one of the following elements: the verb or the noun in question (e.g. *said a letter* instead of **send a letter**, or *contact the chef* instead of *contact the **manager***), or the possible articles (e.g. *find home* instead of *find **a** home*), prepositions (e.g. *ask help* instead of *ask **for** help*) and other determiners, such as possessive pronouns (e.g. *call you landlord* instead of *call **your** landlord*), attached to the combinations. In contrast, possible mistakes found in, for example, adjectives and other modifiers connected to the V + object N combinations were not included (e.g. *have a quite big problem* for *have **quite a big** problem*). Mistakes concerning the tense or the conjugation of the verb were also excluded from this study, as they, too, were not seen as mistakes in the V + object N combination.

At the basic level, 54 of the 112 V + object N combinations were correct, in other words contained none of the mistake types described above. This represents 48% of the combinations found. Accordingly, 58 combinations, which corresponds 52% of all combinations, were judged incorrect. It is generally considered that the correct formation and use of collocations and idioms requires higher language skills than those of free combinations. In this case, however, the results did not match the expectations. Instead, when the correctness of the combinations was examined according to the three categories of free combinations, collocations and idioms, the proportion of incorrect combinations was higher in the category of free combinations, with 56% of the combinations being incorrect, than in that of collocations, where 47% of the combinations were incorrect. The proportions of correct combinations were accordingly 44% of free combinations and 53% of collocations. As it was mentioned earlier, and as it can be seen also in Table 2, there were no idioms to analyse in the basic level texts. The following examples

(2) present some of the incorrect free combinations and collocations in the basic-level texts:

(2) Incorrect free combinations

need some piece and quiet (correct: *need some **peace** and quiet*)

care of somebody's opinion (correct: *care **for/about** somebody's opinion*)

phone to the owner (correct: *phone the owner*)

think life (correct: *think **about/of** life*)

dream about the house (correct: *dream **of** a house*)

get hobby (correct: *get **a** hobby*)

try ggreenment (correct: *try **to agree***)

Incorrect collocations

make a sugestion (correct: *make a **suggestion***)

do noises (correct: ***make** noises*)

hope somebody good luck (correct: ***wish** somebody good luck*)

keep noise (correct: ***make** noise*)

reed papers (correct: ***read** papers*)

look at the TV (correct: ***watch** TV*)

enjoy of one's life (correct: *enjoy one's life*)

As the chapter dealing with the methodology of the study already presented, next the mistakes found in the incorrect V + object N combinations were classified into eight categories: spelling mistakes (which include possible typing mistakes when the original hand-written texts were transcribed into electric form), article mistakes, preposition mistakes, mistakes concerning the choice of verb, mistakes concerning the choice of noun, determiner mistakes (e.g. mistakes in possessive pronouns), mistakes in the choice of singular or plural form of noun, and mistakes that could not be clearly defined and categorised, the category of which is marked by question mark (?). In all categories, except that of spelling mistakes, the category containing mistakes in the choice of singular of plural form, and the incomprehensible mistakes marked with ?, wrong forms included, in addition to the wrong choice of word, also the lack of needed forms and the use of excess forms. The incorrect combinations (3)

below give a few concrete examples of the mistakes classified into the eight categories mentioned:

(3) Spelling mistake
*need some piece and quiet (correct: need some **peace** and quiet)
*make a sugestion (correct: make a **suggestion**)

Article mistake
*find home (correct: find **a** home)
*give somebody a advice (correct: give somebody **a piece of** advice)

Preposition mistake
*ask help (correct: ask **for** help)
*care of somebody's problems (correct: care **for/about** sb's problems)

Verb mistake
*said a letter (correct: **send** a letter)
*look at the TV (correct: **watch** TV)

Noun mistake
*contact the chef (correct: contact the **manager**)
*try ggreenment (correct: try **to agree**)

Determiner mistake
*call you landlord (correct: call **your** landlord)
*read yours letter (correct: read **your** letter)

Sing./plur. mistake
*make a rules (correct: make **a rule**/make **rules**)

? mistake
*call boy&girl (correct: call **them** boy and girl (?), difficult to figure out what the writer was trying to say)
*go their busnes (correct: **do their business**, although the combination does not fit the context in this form either)

All V + object N combinations judged incorrect may have contained one or more than one mistake, and each mistake was counted. Thus, the number of mistakes exceeds the number of incorrect combinations. The numbers and percentages of different types of mistakes in the basic-level texts are presented in Table 3 below.

Table 3. Types of mistakes in the basic-level texts.

	Free combinations		Collocations		Idioms		**Total**	
	N	%	N	%	N	%	N	%
Spelling	5	11%	6	20%	-	-	11	15%
Article	13	29%	7	23%	-	-	20	27%
Preposition	11	24%	3	10%	-	-	14	19%
Verb	5	11%	7	23%	-	-	12	16%
Noun	7	16%	3	10%	-	-	10	13%
Determiner	2	4%	2	7%	-	-	4	5%
Sing./plur.	1	2%	1	3%	-	-	2	3%
?	1	2%	1	3%	-	-	2	3%
Total (N=75)	45	60%	30	40%	-	-	75	

In Table 3 above, the mistakes are classified both in terms of the three categories of free combinations, collocations and idioms, and in terms of the different types of mistakes described earlier. 60% of the 75 mistakes in the basic-level texts were in free combinations, whereas collocations contained only 40% of them. This is no longer surprising after the earlier finding according to which, in the basic level texts, the proportion of incorrect free combinations was higher than that of collocations. When looking at the overall percentages of different types of mistakes on the far right-hand side of Table 3, it does not come as a surprise that the two highest percentages belong to the categories of article and preposition mistakes. After all, it is a commonly known fact that those two aspects of English often cause problems for Finnish-speaking language learners, as their Finnish has a different system for conveying the same meanings and functions of language. In fact, the two categories covered as much as 46% of all mistakes in basic-level texts, which means that almost every other mistake was connected to either an article or a preposition. Other major mistake categories were those of verb mistakes, spelling mistakes and noun mistakes, with percentages of 16%, 15% and 13% of all mistakes. Even at the basic level, with overall grade 2 in writing, there were only a few mistakes concerning determiners

(5%) and the choice of singular or plural form of noun (3%), as well as incomprehensible mistakes marked with the question mark (?) (3%).

When the mistakes were examined according to the different types of V + object N combinations, in other words mistakes in free combinations and mistakes in collocations separately, the big picture remained the same as in the overall percentages. The top five categories with the highest percentages of mistakes were the same. However, there were differences in their mutual order. In free combinations, the two highest percentages could be found in the categories of article and preposition mistakes, similarly to the overall results, followed by noun mistakes with 16%, and spelling and verb mistakes, both with 11% of all mistakes. However, in the category of collocations, the largest percentage of 23% was shared by article mistakes and verb mistakes. The third highest percentage belonged to spelling mistakes, with 20% of all mistakes, followed by preposition and noun mistakes with the percentage of 10% each. The three lowest percentages belonged to determiner mistakes, singular/plural form mistakes and ? mistakes in both free combinations and collocations, with percentages varying from 2% to 7%.

Although in most cases it is difficult to find a reason behind the changes in the percentages, the rise in the percentage of verb mistakes, when moving from free combinations to the category of collocations, seems logical. After all, it is the choice of verb that makes a combination a collocation, as the noun is often chosen by its meaning. As certain nouns require to be used with certain verbs, the result is a collocation. It is, therefore, only natural that the choice of verb has caused problems in the case of collocates, and, as a result, the percentage of verb mistakes rises from 11% to 23% when the nouns in question could no longer be connected to any semantically possible verbs. The changes in the percentages of the other mistake types can be explained only by the change in the category of verb mistakes and the small sample of the study in question. After all, 30 texts per test level is quite a small sample, and, as Table 3 shows, even if the percentages changed a little when moving from free combinations to collocations and vice versa, the numbers of mistakes behind the percentages do not show very big changes. Let us take a look at the results of the intermediate level next.

6.2. Intermediate level

In this study, texts representing the intermediate level refer to texts written by the participants given overall grade 4 in the writing test. At this level, the texts chosen were all informal letters where the participants were asked to express their own opinion on a matter discussed in the test material, and the task included further instructions. Looking at the intermediate-level texts, even before actually counting the words used, it became clear that these 30 texts were considerably longer than their basic-level equivalents. After counting the number of words they included, the assumption was proved right: the average length of the intermediate-level texts was 157 words, the shortest containing 78 words and the longest as much as 225 words. Compared to the average number of words in the basic-level texts, 88 words, the average length of texts almost doubled when moving on to the intermediate level. As a consequence, the intermediate-level texts also provided more V + object N combinations to analyse. Table 4 below presents the overall number of combinations, as well as the proportions of free combinations, collocations and idioms.

Table 4. Distribution of V+ object N combinations in the intermediate-level texts.

	Number	Percentage
Free combinations	146	63%
Collocations	83	36%
Idioms	1	0%
Total	230	

The intermediate-level texts contained 230 V + object N combinations altogether, which is more than twice as much as could be found in the basic-level texts. When comparing the number of combinations to the overall number of words used in the texts, the intermediate-level texts included one V + object N combination per 20 words. This shows that, in addition to the rise in the number of combinations largely explained by the increase in the length of texts, the intermediate-level participants also used V + object N combinations more frequently than the lower-level participants.

Although the overall number of combinations more than doubled, an even bigger rise could be seen in the number of free combinations: In contrast to the 59 free combinations provided by the basic–level texts, overall grade 4 texts included

as much as 146 free combinations, which represents 63% of all V + object N combinations found at this level and an increase of 10% compared to their proportion of the basic-level combinations. Keeping in mind the underlying assumption of participants with higher target language proficiency providing a higher percentage of collocations and idioms, the rise in the proportion of free combinations came as a surprise. Although also the number of collocations rose remarkably when moving from the basic level to the intermediate level, from 53 to 83 collocations, their proportion of all combinations decreased from 47% to only 36%, contrary to what was expected. The number of idioms found remained extremely low: the 230 intermediate-level V + object N combinations included only one idiom. See the examples (4) below for a few examples of the intermediate-level combinations:

(4) Free combinations

understand the frustration	*have space*
allow smoking	*get resources*
think about one's options	*ignore the ban*
damage life	*give room*

Collocations

make changes	*see sense*
invest money	*inhale smoke*
have customers	*increase appetite*
spend time	*cause cancer*

Idioms
*see the bright side

Even though the change in the proportion of collocations did not match the expectations, the percentages of correct and incorrect combinations reflected the higher target-language proficiency of the intermediate-level participants, when compared to those at the basic level.

Table 5. Correct vs. incorrect combinations in the intermediate-level texts.

	Correct		Incorrect	
	Number	Percentage	Number	Percentage
Free combinations (N=146)	112	77%	34	23%
Collocations (N=83)	68	82%	15	18%
Idioms (N=1)	0	0%	1	100%
Total (N=230)	180	78%	50	22%

As Table 5 above shows, when compared to the corresponding figures of the basic-level combinations presented earlier in Table 2, the intermediate-level combinations provided a considerable rise in the proportion of correct combinations. 112 of the 146 free combinations were correct, resulting in a percentage of 77%, the corresponding percentage at the basic level being 44%. Quite a similar rise could be seen in the category of collocations, of which as much as 68 out of 83, 82% of them, were correct at the intermediate level, as opposed to the 53% at the basic level. The percentages of incorrect combinations decreased accordingly, resulting in 23% of combinations being incorrect in the category of free combinations and 18% in that of collocations. The only idiom found was incorrect, causing the striking percentage of 100% in the section of incorrect idioms in Table 5 above. Once again, one has to keep in mind the number of combinations behind the percentages before jumping into conclusions. The examples (5) below provide a few concrete examples of the incorrect combinations found at this level:

(5) Incorrect free combinations

*re consider one's choises (correct: **reconsider** one's **choices**)

*put a law (correct: **pass** a law)

*have restaurant (correct: have **a** restaurant)

*chooce between one's health and income (correct: **choose** between...)

*affect on business (correct: **affect** business)

*get intension (correct: get **attention**)

*need a rules (correct: need **rules/a rule**)

Incorrect collocations

*do decisions (correct: **make** decisions)

*spend tim (correct: spend **time**)

*have good time (correct: have **a good time**)

take a paint (correct: **have a pint**)

have astma (correct: have **asthma**)

make decisitions (correct: make **decisions**)

become a law (correct: become law)

Incorrect idioms

see the bright side (correct: **look on** the bright side)

Table 6 below presents the overall number of mistakes found in the incorrect intermediate-level V + object N combinations, as well as their division into the three categories of free combinations, collocations and idioms, similarly to what Table 3 earlier showed of the basic-level mistakes. In accordance to Table 3 presenting the corresponding figures at the basic level, Table 6 below includes also the distribution of the intermediate-level mistakes into the eight different mistake categories presented earlier, both as numbers and as percentages of the overall number of mistakes.

Table 6. Types of mistakes in the intermediate-level texts.

	Free combinations		Collocations		Idioms		**Total**	
	N	%	N	%	N	%	N	%
Spelling	10	29%	10	59%	0	0%	20	38%
Article	7	20%	4	24%	0	0%	11	21%
Preposition	5	14%	0	0%	0	0%	5	9%
Verb	4	11%	3	18%	1	100%	8	15%
Noun	3	9%	0	0%	0	0	3	6%
Determiner	1	3%	0	0%	0	0	1	2%
Sing./plur.	3	9%	0	0%	0	0	3	6%
?	2	6%	0	0%	0	0	2	4%
Total (N=53)	35	66%	17	32%	1	2%	53	

As Table 6 shows, there were 53 mistakes altogether in the intermediate-level incorrect V + object N combinations at the intermediate level, which is less than the 75 mistakes found at the basic level. At this point, this did not come as a surprise, as the lower number of mistakes could already be expected by the lower percentage of incorrect combinations presented by Table 5. The distribution of mistakes into the different categories of combinations remained roughly the same as at the basic level:

66% of the mistakes were in free combinations and 32% of them in collocations, whereas the corresponding percentages were 60% and 40% at the basic level. The mistake in the only idiom found represents 2% of all intermediate-level mistakes.

What is interesting in Table 6, however, is the distribution of mistakes, when compared to that of the basic level (shown in Table 3 earlier). Let us first take a look at the overall percentages of mistakes on the right-hand side of Table 6. At the basic level, the two categories with the highest percentages were article and preposition mistakes, covering almost half of all mistakes (46%). At the intermediate level, the percentages were considerably lower in both categories, even though they both still belonged to the top four mistake categories. Article mistakes now covered only 21% of all mistakes and preposition mistakes as little as 9%, compared to the percentages of 27% and 19% at the basic level. Instead, as much as 38% of intermediate-level mistakes were spelling mistakes, which is a remarkable rise when compared to their proportion of 15% at the basic level. The percentage of mistakes concerning the choice of verb remained almost the same as at the basic level, 15% as opposed to the earlier 16%, leaving only a few mistakes, and low percentages, to the rest of the mistake categories: noun mistakes and sing./plur. mistakes covered both 6% of all intermediate-level mistakes, incomprehensible mistakes 4% and determiner mistakes 2%.

When the mistakes were examined according to the different combination categories, once again, the big picture naturally stayed the same as with the overall mistakes. Spelling mistakes formed the biggest group of mistakes both in free combinations, with the percentage of 29%, and in collocations, with as much as 59%. The second largest group of mistakes were article mistakes, which represented 20% of mistakes in free combinations and 24% of them in the category of collocations. While the percentages of mistakes in free combinations followed roughly the same order as the percentages of all intermediate-level mistakes, with the exception of the proportion of preposition mistakes exceeding that of verb mistakes, as opposed to their reverse order in the overall column, the low number of mistakes in collocations, and especially their narrow range, sticks out when looking at the column in question. After the already mentioned percentages of spelling and article mistakes, and the proportion of 18% of verb mistakes, the rest of the mistake categories were left empty. Quite surprisingly, there were no mistakes concerning prepositions, the choice of noun, determiners, sing./plur. or ? in collocations. The only idiom contained a mistake in the choice of verb, thus resulting in the possibly

misleading percentage of 100% in the section of verb mistakes in the column of idioms. Before taking a look at the role of L1 in the mistakes found, let us examine the results of the last of the three skill levels, the advanced level.

6.3. Advanced level

The third level of proficiency in the study in question, the advanced level, refers to the texts written by the participants given overall grade 6 in the advanced-level writing test, in other words the highest possible grade to be given in the National Language Certificate Test. The texts were all informal letters or messages, where the participants, similarly to the task at the intermediate level, had to express their own opinions on a subject discussed in the test material. At this level, too, the task included further instructions for the writers. Similarly to the increase in the length of texts when moving from the basic level to the intermediate level, the advanced-level texts also contained more words than their lower-level equivalents. The shortest text at the advanced level contained 125 words and the longest 311 words, resulting in the average of 193 words per text, which is significantly more than the average length of text at the two lower levels: 88 words at the basic level and 157 words at the intermediate level. The underlying expectation was that the higher level of target-language proficiency would also be seen in the results, that is to say that, compared to the overall number of V + object N combinations at this level, the proportions of collocations and idioms would rise, as would also the percentage of correct combinations. Table 7 below shows the overall number of advanced-level combinations, as well as their distribution into free combinations, collocations and idioms.

Table 7. Distribution of all V+ object N combinations in the advanced-level texts.

	Number	Percentage
Free combinations	194	64%
Collocations	104	34%
Idioms	4	1%
Total	302	

The advanced-level texts contained 302 V + object N combinations in all. As it was expected, the number of combinations was considerably higher than at the lower test

levels. At the basic level, the overall number of combinations was 112 and at the intermediate level 230. Similarly to the rise from the basic level to the intermediate level, the most obvious reason behind the rise seems to be the length of the texts in question, as once again the higher target-language proficiency resulted in longer texts. However, the increase in the number of combinations cannot be explained simply by the longer texts, as, just like when moving from the basic level to the intermediate level, also in this case the participants with higher language skills used V + object N combinations more frequently in their texts than the participants at the lower levels. At this level, there was one V + object N combination per 19 words on average, which was more frequently than the 20 and 24 words per combination at the intermediate and basic levels. Although, once again, the number of V + object N combinations rose considerably when moving to a higher skill level, a similar rise could not be seen in the proportions of collocations and idioms, unlike what the higher level of language proficiency was supposed to cause. On the contrary, the percentage of collocations even decreased a little from the corresponding percentage of 36% at the intermediate level, resulting in a percentage of 34%. There were 4 idioms within advanced level V + object N combinations, which is more than the one idiom found at the intermediate level. As it was earlier presented, the basic level texts did not contain any idioms at all. However, when presented as percentages, the four idioms represent only 1% of all advanced-level combinations. Thus, contrary to what was expected, even though the number of collocations and idioms found in the advanced-level texts was higher than the corresponding figures at the lower levels, when they are compared to the overall number of V + object N combinations found, their proportions remained roughly the same as at the intermediate level. The following group of examples (6) contains a few examples of the combinations found in the advanced-level texts:

(6) Free combinations

have a possibility   *compete for awareness*

express one's fears   *need content*

promote freedom   *educate people*

need merits   *offer a way*

Collocations

raise awareness   *get a grasp*

show potential   *offer an insight*

bring up children   *have rights*

make judgements   *see evidence*

Idioms

have a part to play   *serve time*

make a difference

Table 8 below shows the proportions of correct and incorrect V + object N combinations in the advanced-level texts. The percentages are presented both as an overall percentage and individually for each of the three combination categories of free combinations, collocations and idioms. The expectation was that the higher level of language proficiency would be seen as higher percentages of correct combinations than at the two lower levels presented earlier.

Table 8. Correct vs. incorrect combinations in the advanced-level texts.

	Correct		Incorrect	
	Number	Percentage	Number	Percentage
Free combinations (N=194)	167	86%	27	14%
Collocations (N=104)	89	86%	15	14%
Idioms (N=4)	3	75%	1	25%
Total (N=302)	259	86%	43	14%

In this sense, the expectations were met. At the advanced level, as much as 86% of the 302 V + object N combinations were correct, which is more than the 78% at the intermediate level and significantly more than the 48% at the basic level. Not surprisingly, the proportion of correct combinations rose also in each of the three subcategories. In the category of free combinations, 86% of the combinations were correct at the advanced level, as opposed to the 77% at the intermediate level and the 44% at the basic level. Similarly, 86% of the advanced-level collocations were correct, the corresponding percentages at the lower test levels being 82% and 53%. In the category of idioms, the small number of combinations, or, in the case of the basic level, the lack of them, complicates the comparison between the three test

levels. It does not seem reliable to draw too many conclusions based on the one intermediate-level idiom and its four advanced-level equivalents, even though the percentages of 100% being incorrect at the intermediate level and 25% at the advanced level are presented in the tables similarly to the other percentages. The examples (7) below reveal some of the incorrect combinations found at this level:

(7) Incorrect free combinations

farm tomatoes (correct: **grow** tomatoes*)

concern free speech* (correct: *concern **freedom of speech*)

*make justice (correct: ***do** justice*)

ponder upon matters* (correct: *ponder **on/over matters*)

qustion the credebility* (correct: ***question the **credibility***)

approve for an imprisonment* (correct: *approve **of an imprisonment*)

sneal past the security controls* (correct: ***sneak past the security controls*)

Incorrect collocations

*change world (correct: *change **the** world*)

take one's contribution* (correct: ***make one's contribution*)

bring negative consequences* (correct: ***have negative consequences*)

*loose power (correct: ***lose** power*)

*give opportynity (correct: *give **an opportunity***)

voice they concern* (correct: *voice **their concern*)

*open door (correct: *open **a/the** door*)

Incorrect idioms

ake things at face value* (correct: ***take things at face value*)

Table 9 on the next page presents the number of mistakes found in the incorrect combinations at the advanced level, as well as their distribution into the already familiar types of mistakes. The distribution of mistakes and the overall figures are presented both as numbers and as percentages.

Table 9. Types of mistakes in the advanced-level texts.

	Free combinations		Collocations		Idioms		**Total**	
	N	%	N	%	N	%	N	%
Spelling	11	39%	7	41%	1	100%	19	41%
Article	3	11%	4	24%	0	0%	7	15%
Preposition	5	18%	1	6%	0	0%	6	13%
Verb	6	21%	3	18%	0	0%	9	20%
Noun	3	11%	1	6%	0	0%	4	9%
Determiner	0	0%	1	6%	0	0%	1	2%
Sing./plur.	0	0%	0	0%	0	0%	0	0%
?	0	0%	0	0%	0	0%	0	0%
Total (N=46)	28	61%	17	37%	1	2%	46	

As Table 9 above shows, the 43 incorrect V + object N combinations contained 46 mistakes altogether. Similarly to the two earlier test levels, also at the advanced level the number of mistakes exceeded that of incorrect combinations due to the fact that a single incorrect combination may have contained more than one mistake. At the two highest test levels, the intermediate level and the advanced level, the difference between the number of mistakes and the number of incorrect combinations was no longer as big as at the basic level, thus reflecting the higher target language proficiency of the test-takers and the smaller number of mistakes found in the combinations.

The right-hand side of Table 9 presents the overall percentages of all mistake types at the advanced level. Similarly to the intermediate level, also at the advanced level, the highest percentage belonged to spelling mistakes, which, with the proportion of 41%, left the rest of the categories far behind. In fact, the proportion of spelling mistakes was even a little higher than the 38% at the intermediate level. The second largest group of mistakes is verb mistakes, with the percentage of 20%, followed by article mistakes (15%), preposition mistakes (13%), noun mistakes (9%) and determiner mistakes (2%). The advanced-level V + object N combinations did not contain any sing./plur. mistakes or ? mistakes at all.

The top four mistake categories were the same as at the two lower skill levels, in other words mistakes concerning spelling, choice of verb, articles and prepositions provided the highest percentages at every skill level involved in this study, even though their mutual order varied between the levels. In this sense, the results of the advanced level were in accordance with the results of the two lower

levels presented earlier. The higher target language proficiency of the advanced-level test takers could be seen, in addition to the fewer number of mistakes, also in the higher percentage of spelling mistakes, some of which were without a doubt simply careless mistakes instead of being signs of inadequete language skills. Moreover, the higher target language skills were reflected also through the more narrow range of mistakes found. Sing./plur. mistakes and ? mistakes can be categorised as mistake types that are so basic that they are more common to beginners. Thus, their lower percentages at the advanced level, in this case even the lack of them, did not come as a total surprise.

Table 9 also shows the proportions of the different types of mistakes within each of the three categories of V + object N combinations – free combinations, collocations and idioms. As free combinations formed as much as 61% of all advanced-level V + object N combinations, the results of this category did not differ that much from the overall results of the level presented above. The largest category of mistakes was spelling mistakes with the percentage of 39% of all mistakes found in the advanced-level free combinations. Mistakes concerning the choice of verb formed 21% and preposition mistakes 18%, followed by the categories of article mistakes and mistakes concerning the choice of noun, both covering 11% of all mistakes. Three mistake categories, those of determiner mistakes, sing./plur. mistakes and ? mistakes were left empty.

Also in the category of collocations, spelling mistakes formed the largest group of mistakes with the percentage of 41% of the 17 mistakes found in advanced-level collocations. The second largest group were article mistakes, covering 24% of the mistakes, and the third largest group were mistakes concerning the choice of verb, with the percentage of 18%. Preposition mistakes, mistakes concerning the choice of noun and determiner mistakes covered 6% each, which in this case, the overall number of mistakes in collocations being only 17, refers to only 1 mistake in each category. Sing./plur. mistakes and ? mistakes were not represented at all in the category of collocations. At the advanced level, the results of the category of collocations resemble those of the intermediate level in that spelling mistakes, article mistakes and mistakes in the choice of verb formed the top three types of mistakes when it comes to the percentages. At the intermediate level, the rest of the mistake categories were left empty. As the same types of mistakes covered either one mistake or none at all also at the advanced level, the results of the category of collocations are nearly alike at the two levels in question.

At the advanced level, similarly to the two lower skill levels, the results of the category of idioms were clear and simple. As it was earlier presented, none of the basic-level V + object N combinations were idioms. The intermediate level combinations contained one incorrectly formed idiom, which was the case also at the advanced level. This time, unlike the intermediate level where the cause of mistake was choice of verb, the idiom contained a spelling mistake. In both cases, although the tables in question rightly provide the striking percentage of 100% in the two mistake categories, the results have to be read with caution. As the percentages refer to only single mistakes, generalisations such as all mistakes in the idioms formed by intermediate of advanced-level writers belonging to these to categories would be poorly grounded.

6.4. L1 influence

In addition to the mistakes found in the V + object N combinations at the three skill levels and their distribution into the different types of mistakes, the focus of this study was to examine the role of L1 in the mistakes found, in other words, whether the participants' mother tongue, in this case Finnish, seemed to have had a role in the production of the erroneous combinations. This part of the analysis relies entirely on the language skills of the researcher. Due to practical reasons, other sources of information have not been used and, thus, the results should be seen only as a rough outline. In any case, it is interesting to find out, even if only approximately, how big a proportion of mistakes were due to e.g. a falsely similar-sounding form in the mother tongue. If not anything else, if the results show that the forms of the mother tongue are behind the erroneous target-language forms, at least this calls for their acknowledgement in the teaching process.

The possible L1 influence in the incorrect V + object N combinations was studied simply by counting all the incorrect combinations and trying to come up with the reason behind the mistake or mistakes behind the incorrect form(s). In case the possible reason could have been a form in the Finnish language, this was also supposed to be the cause and the incorrect combination was marked down as containing negative L1 influence. For the purposes of this study, this was done even though it may not have been the real cause of mistake. In this part of analysis, the data were not analysed according to the three different types of combinations. Instead, the results are presented simply according to the three skill levels. As it is

commonly acknowledged that learners tend to rely on the rules of their mother tongue more in the beginning of the learning process than later on when they have higher target language proficiency, the assumption was that the role of L1 would diminish when moving from the basic level to the higher skill levels. Table 10 below shows the number and percentage of incorrect V + object N combinations in which the cause of mistake could have been a form in Finnish language.

Table 10. Number and percentage of incorrect combinations with negative L1 influence.

	L1 influence	
	Number	Percentage
Basic level (N=58)	16	28%
Intermediate level (N=50)	5	10%
Advanced level (N=43)	0	-
Total (N=151)	21	14%

As Table 10 presents, the assumption was proved right and the proportion of incorrect V + object N combinations where the cause of mistake could have been a form in the participants' mother tongue really did diminish as the level of target-language proficiency rose. In total, at all levels, there were 151 incorrect V + object N combinations, 21 of which, 14%, were marked as containing negative L1 influence as the source of mistake.

At the basic level, the role of negative L1 influence was the biggest. At this level, as much as 16 out of 58, 28%, of the incorrect combinations could be linked to a form in the Finnish language. Some examples (8) of the combinations in question are presented below.

> (8)*phone to the owner* (correct: *phone the owner*; Finnish *soittaa omistajalle*, the inflectional end *–lle* often corresponds to the preposition *to*)
>
> *call to the police* (correct: *call the police*; Finnish *soittaa poliisille*, see the explanation above)
>
> *enjoy of one's life* (correct: *enjoy one's life*; Finnish: *nauttia elämästään*, the inflectional end *–stä* often corresponds to the English *of*)
>
> *keep noise* (correct: *make noise*; Finnish *pitää meteliä*, the verb *pitää* has many meanings, some of which translate as *keep*)

look at TV (correct: *watch TV*; Finnish: *katsoa/katsella televisiota*, unlike in English, in Finnish both the equivalent of 'look' at and that of 'watch' can be used when referring to watching TV)

As many as 7 of the 16 L1-influenced combinations involved either the verb *phone* or the verb *call* incorrectly followed by the preposition *to*. This is clearly due to the corresponding verb *soittaa* in Finnish, after which the following object noun has to contain an inflectional end equivalent to the English preposition *to*. These, in addition to the other preposition mistakes, formed the biggest group of L1 influenced mistakes. This does not come as a surprise. After all, preposition mistakes were earlier in this paper proved to be one of the largest mistake categories at all skill levels.

Another big category where the role of L1 was clear was that of verb mistakes, with 6 combinations. The link to the participants' mother tongue could easily be seen in the incorrect choices of verb, some of which are presented above. The verbs had been either translated straight from Finnish or the participants had chosen the wrong translation for a Finnish verb that has several equivalents in English, each of which collocates with different nouns.

At the intermediate level, the percentage of L1 influenced mistakes is already significantly lower. Only 5 incorrect combinations out of 50, in other words only 10% of them, could be traced to a form in the participants' mother tongue.

do decisions (correct: *make decisions*; Finnish *tehdä päätöksiä*, both 'do' and 'make' translate as 'tehdä' in Finnish)

take a paint (correct: *have a pint*; Finnish *ottaa tuoppi*, in Finnish, the verb used in the expression is the equivalent of 'take', the spelling mistake in the noun is probably due to the phonetic form of the word 'pint')

see the bright side (correct: *look on the bright side*; Finnish *nähdä valoisampi puoli*, the Finnish verb 'nähdä' in the corresponding expression translates as 'see' in English)

Finally, at the advanced level, there was not a single incorrect combination that could be straightforwardly linked to a form in the Finnish language. As a conclusion, at this

point, the participants' language proficiency has apparently reached the point where they no longer rely on their mother tongue when interacting in the target language.

Chapter 6 presented the findings of the study in detail. As the results cover quite a few pages, it felt useful to discuss them further, for example in terms of earlier research, in the next chapter.

7. DISCUSSION

The purpose of this study was to examine the use of English collocations by Finnish foreign language learners. The focus was narrowed down to combinations consisting of a verb and a noun in the object position (V + object N). To be able to present the use of collocations more thoroughly, the study included also free combinations, i.e. V+ N combinations formed purely based on the semantic properties of the words without any other restrictions, and idioms. In addition, the three types of V + object N combinations included samples from three groups of participants, each representing a different skill level. This, too, was done to give a wider perspective on the issue.

Based on the theoretical background information on collocations and the earlier research done in the field of different types of formulaic sequences, there were several underlying assumptions concerning the results. As the correct use and formation of collocations and idioms is generally considered to require quite a high skill level in the target language, it was assumed that the group of participants with the highest target language proficiency, i.e. the advanced-level participants, would provide more collocations and idioms than the two lower levels. Moreover, there was a presumption that their collocations and idioms would contain less mistakes and that negative L1 influence would be the cause of their mistakes more rarely than in the case of basic and intermediate-level participants. Although many of the findings were consistent with the assumptions, there were also some surprises. Let us first take a look at the results supporting the presumptions.

Firstly, the higher the skill level, the longer the texts were. The average length of texts became considerably longer as the participants' language proficiency got higher. Even though there was a lot of variation concerning the length of text within each group of informants, the difference in the average length

was even bigger than it was anticipated. For example, when moving from the basic level to the intermediate level, the average number of words in a text almost doubled. Although the length of texts was not one of the main subjects of study in this case, it supports the presumption that the participants with higher skill level would provide more material to analyse, as the longer the texts were the more V + object N combinations they contained. The advanced-level texts contained more combinations than their intermediate-level equivalents in which the combinations were more numerous than at the basic level. In addition to the number of combinations, also their density increased as the skill level rose. The average number of words in a text per one V + object N combination was smaller at the advanced level than at the intermediate level, the basic level being the one with the most words per one combination. It is easy to see why the informants with higher proficiency in English provided longer texts than the ones in the beginning of their language learning process. As the skill level in the target language becomes higher, it is easier to express oneself in the foreign language and to produce more text. Also, the tendency of the number of V + object N combinations to get higher as the texts became longer seems natural. The more text the informants produced, the more they had to use different linguistic elements to create the text. However, the result concerning the change in the number of V + object N combinations compared to the overall number of words is more complex. Unfortunately, this study does not provide an explicit explanation to why the informants with higher target language skills used V + object N combinations more frequently than their counterparts. Relying on common sense, one can only assume that the structure of language also develops as the skill level becomes higher and the more proficient informants no longer used as simple sentences as the ones with lower target language skills, thus resulting in the higher frequency of all types of V + object N combinations put together.

Secondly, according to the results, higher target language proficiency correlated with fewer mistakes. As presumed, the percentage of correct combinations rose and that of incorrect ones diminished as the skill level got higher. This was true both when examining all types of combinations together and in each of the three types of combinations, i.e. free combinations, collocations and idioms, individually. Although, in some categories, the number of combinations was small – sometimes even nonexistent – all figures agree with the earlier assumptions. Moreover, the number of mistakes in the incorrect combinations decreased as the participants' language proficiency became higher. This corresponds with the lower number of

incorrect combinations, but as the mistakes were counted separately and one incorrect combination could contain several mistakes, it is worth mentioning. At all levels, the number of mistakes exceeded the number of incorrect combinations. As the difference between the figures decreased as the skill level became higher, it can be concluded that the higher the skill level, the lower the average number of mistakes per one incorrect combination. In a sense, the incorrect combinations in the advanced-level texts were "less incorrect" than at the two lower levels. The reason behind the number of mistakes becoming lower as the target-language skills became higher seems only logical and does not require any further explanations at this point.

   In addition to the results supporting the assumptions made based on earlier research, there were several areas where the results of this study were a little surprising. Let us start with the proportions of each of the three different types of V + object N combinations involved in the study. As it has been stated also earlier in this study, the correct use of collocations and especially that of idioms is commonly associated with fairly good English skills. For that reason, it was assumed that the basic-level informants would have produced mainly free V + object N combinations and only a small number of collocations and idioms, the proportions of which would become higher with each skill level. Partly, this was true. The majority of the combinations formed by the basic-level informants were free combinations. However, there were nearly as many collocations, the difference being only a few percentages. This, in addition to the fact that, at the basic level, there were a higher percentage of correct combinations in the category of collocations than in that of free combinations, disagrees with the assumptions. Moreover, even though the number of collocations was higher at the intermediate level than at the basic level, their proportion of all combinations diminished. For some part, the same was true when moving from the intermediate level to the advanced level. Again, the number increased, but the proportions remained almost the same as at the intermediate level. In this sense, the assumptions were proven wrong. However, the results corresponded with the presumptions in that the proportions of correct combinations became higher with each skill level both in the category of free combinations and in that of collocations. The number of idioms remained extremely low throughout the three skill levels. With only five idioms altogether, it is inappropriate to draw any major conclusions, even though their number increased with each skill level, as did also their correctness. Anyway, it did not come as a surprise that there were no idioms in

the basic-level texts, and that their number remained low also at the two higher skill levels. After all, the correct use of idioms requires high language skills.

The study also examined the types of mistakes found in the incorrect V + object N combinations. The mistakes were divided into eight different categories and, in addition to the overall figures concerning the three types of combinations altogether, results were given also for each of them separately. Considering the structural differences between Finnish and English, it did not come as a surprise that throughout the skill levels a remarkable proportion of mistakes were related to either articles or prepositions. As the Finnish language does not contain articles or separate prepositions, but is instead based on a system of inflectional endings, these two parts of English often cause problems for Finnish language learners. In addition to these two types of mistakes, at each skill level, the top four categories included also spelling mistakes and mistakes concerning the verb of the combination, even though their mutual order varied between the skill levels. Similarly to mistakes related to articles and prepositions, also spelling mistakes and mistakes involving the verb are easy to understand. The spelling of a foreign language is a natural cause of mistake for all language learners, regardless of their mother tongue. However, as the pronunciation of Finnish is extremely reliable to the spelling system of the language, the differences between the pronunciation and spelling of English can easily be seen as a problematic area for Finnish people. As for the mistakes concerning the verbs of the combinations, their high percentages in the category of collocations seem natural. After all, it is the choice of verb that makes a combination a collocation, as the noun is often the one chosen based on its semantic property. Certain nouns require to be used with certain verbs, which, combined to the characteristics of English verbs to have several nearly synonyms, is a definite cause of problems for Finnish learners of English.

Even though many aspects of the results concerning the categorisation of mistakes seem logical, some of them were a little surprising. Firstly, the remarkable decrease in the proportions of article and preposition mistakes when moving from the basic level to the intermediate level, even though both still remained in the top four mistake categories. According to the results, it can be concluded that a considerable number of article and preposition mistakes can be avoided by reaching the intermediate level, although they are fairly common at this level, too. Another result worth pointing out at the intermediate level is the increase in the proportion of spelling mistakes. Perhaps, at this level, higher language skills

enable language learners to try to express themselves more by using a larger vocabulary, which results in more spelling mistakes. Of course, the decrease in the categories of article and preposition mistakes affects the figures, too. The highest skill level, the advanced level, provided some interesting results as well. Interestingly, at this level, there were no longer any mistakes concerning the choice of singular or plural form of the noun in question, or any incomprehensible mistakes categorised in the ? category. The advanced-level informants have, without a doubt, reached a skill level that is beyond as simple types of mistakes as these. Consequently, there was a high increase in the proportion of spelling mistakes, when moving from the intermediate level to the advanced level.

The study also briefly examined the role of the informants' mother tongue, in this case Finnish, in the incorrectly formed V + object N combinations they had formed. Based on earlier research done on L1 influence, the assumption was that the informants would rely less on the rules of their mother tongue as their target language skills improved. In other words, it was presumed that in the advanced level, the structures of Finnish language would be the cause of fewer mistakes than at the two lower skill levels. The assumption was proven right. The number of incorrect combinations where the participants' mother tongue Finnish could be clearly identified as the most likely cause of mistake diminished as the participants' skill level got higher. At the basic level, more than a quarter of all mistakes could be traced back to the participants' L1, whereas at the intermediate level, the corresponding figure was 10%. Finally, when moving on to the advanced level, negative L1 influence could no longer be found at all. Similarly to the earlier findings supporting the assumptions made based on earlier research, also the tendency of negative L1 influence to diminish with increasing foreign language proficiency is only natural. As presented in the theoretical background section, it is common for foreign language learners to rely on the rules of their mother tongue at the beginning of their language learning process. When their skills improve and they become familiar with the structures of the target language, the learners no longer need to depend so much on their L1. As a consequence, at the advanced level, the informants apparently had already reached a level where they could communicate in English without a need to translate their thoughts from Finnish, or at least they were capable of translating their thoughts using purely English structures and did not fall for word-for-word translations.

The results support the earlier studies done on formulaic sequences and collocations (see Chapter 3.5.) in that the learners with higher English skills were more proficient users of collocations and idioms than the ones with a lower skill level. In other words, the higher the skill level of the informants, the fewer mistakes their collocations and idioms contained. Although the skill level of the informants of this study was not defined based on their correct use of collocations and idioms, indirectly the results support the theories according to which that is connected to a higher language proficiency. However, unlike Boers et al. (2006) who suggest that the perceived language proficiency is connected to a higher number of formulaic sequences, the results of this study showed that especially the proportion of collocations did not increase as the skill level became higher. Instead, the proportion was at its highest at the basic level, and in fact became lower as the skill level rose. At this point, it is worth reminding, however, that the formulaic sequences examined by Boers et al. (2006) included a wider range of formulaic multi-word units than this study.

The results also agree with those of Nesselhauf (2003) in that collocations were proven problematic also for learners with higher English skills. There was, nevertheless, a clear decrease in the number of mistakes in the collocations especially when moving from the basic level to the intermediate level and also when moving on to the advanced level. Still, without a doubt, both areas of English deserve to be fully acknowledged in Finnish EFL classrooms. Similarly to the results by Nesselhauf (2003), also the results of this study showed that, in addition to supporting the learning process, the learners' mother tongue sometimes has a negative effect on their use and formation of English collocations. It is, therefore, easy to agree with her in that the role of mother tongue should be more acknowledged in foreign language classrooms. Similar structures are easy to learn and remember, but also partly similar structures should be addressed to prevent confusion. However, as Nesselhauf (2003: 238) points out, it is simply impossible to explicitly teach all collocations. It is, therefore, important to pay more attention on the usefulness of the expressions, i.e. their acceptability and frequency both in English in general and in the registers that are especially important for the learners in question (Nesselhauf 2003: 238; Koprowski 2005).

8. CONCLUSION

The study provided some interesting information concerning Finnish FL learners' use of English collocations, an area which has not been studied that much. Based on the results of this study, the use of collocations is not as straightforward a phenomenon as one might assume. There were several points where the results agreed with the assumptions made based on the theoretical background information and earlier research, but that is not the whole picture. As a conclusion, it can be said that higher target language skills have an effect on the collocations used by language learners. It seems that the higher skills improve the quality of collocations and idioms, resulting in fewer mistakes and less negative L1 influence. However, unlike presumed, when compared to the overall number of all V + object N combinations found, the proportion of collocations does not increase with language skills.

The results also showed that there are a few categories of mistakes that are common for Finnish learners of English regardless of their skill level. At all levels, the top four mistake categories included articles, prepositions, verbs and spelling, even though their mutual order varied. According to the results, although these four areas of English have been known as problematic for Finnish people for a long time, they still call for extra attention in language classrooms. The results of the study also revealed that, although the results were quite similar at the basic and intermediate levels, when reaching the advanced skill level, the learners' English skills are high enough for them to no longer produce some of the more simple types of mistakes, such as mistakes concerning the choice of singular or plural forms of nouns.

Similarly to all research, this study had several limitations. First and foremost, the sample of the study was relatively small. However, an effort was made to give as wide a perspective on the issue as it was possible with this amount of data. As a result, the data included texts from three different skill levels. Consequently, the number of texts per skill level was only 30. This, in turn, had its effect on the percentages counted. With a small sample, a few individual combinations or mistakes seem like a lot more as a percentage, which should be kept in mind when examining the results of this study. Despite this, the percentages were given in order to make the results easier to read. Another major limitation concerns the part dealing with the L1 influence. The analysis was purely based on the researher's own native language skills, which inevitably influences the reliability of the results. For the

purposes of this study, the analysis did, however, give some interesting extra information and, if not anything else, provides an interesting topic for a more thorough study in the future. Moreover, as the role of mother tongue in foreign language learning is not only negative, but also positive (an area excluded from this study), the relationship of the two languages provides quite a few interesting topics for further research. In addition, since the present study does not provide any information on the way collocations and other formulaic sequences have been addressed by the participants' teacher, nor on their role in the teaching materials used, these two and their possible effects on the amount and quality of formulaic sequences used would form an interesting topic for further research, too.

As a conclusion, while giving a lot of information on the way Finnish learners of English use English collocations in short written tasks, and how their mother tongue Finnish affects the mistakes they make, the study also revealed a need for further studies on several areas related to the issue. After all, collocations and other formulaic sequences are an essential part of the language used by native speakers of English and, as such, they deserve to be fully acknowledged in learning English as a foreign language, too.

.

9. BIBLIOGRAPHY

Antunović, G. 2007. Croatian translators' take on Swedish collocations and idioms. In Nenonen, M. and Niemi, S. (eds.). *Collocations and Idioms 1. Papers from the First Nordic Conference on Syntactic Freezes. Joensuu, May 19-20, 2006.* Joensuu: Joensuu University Press, 27-40

Boers, F., Eyckmans, J, Kappel, J., Stengers, H. and Demecheleer, M. 2006. Formulaic sequences and perceived oral proficiency: putting a Lexical Approach to the test. *Language Teaching Research* 10,3, 245-261.

Carter, R. 1987. *Vocabulary.* London: Allen and Unwin.

Collins Cobuild English Dictionary for Advanced Learners 2001. Glasgow: HarperCollins Publishers.

Jiang, N. and Nekrasova, T. 2007. The Processing of Formulaic Sequences by Second Language Speakers. *The Modern Language Journal* 91 (3), 433-445.

Koprowski, M. 2005. Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT Journal* 59 (4) October.

Lewis, M. 1993. *The Lexical Approach. The state of ELT and a way forward*. LTP.

Lewis, M. 1997 (reprinted 1998, 2001). *Implementing the Lexical Approach*. Hove: LTP.

Moon, R. 1998. *Fixed Expressions and Idioms in English.* New York: Oxford University Press Inc.

Moon, R. 1997. Vocabulary connections: multi-word items in English. In Schmitt, N. and McCarthy, M. (eds.). *Vocabulary: Description, Acquisition and Pedagogy.* Cambridge: Cambridge University Press.

Nattinger, J. R. and DeCarrico, J. S. 1992. *Lexical Phrases and Language Teaching.* Oxford: Oxford University Press.

Nesselhauf, N. 2003. The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics* [online] 24 (2), 223-242. (5 August 2008)

Oxford Collocations Dictionary for Students of English 2002. Oxford: Oxford University Press.

Read, J. And Nation, P. 2004. Measurement of formulaic sequences. In Schmitt N. (ed.). *Formulaic Sequences: Acquisition, processing and use*. Amsterdam: John Benjamins.

Schmitt, N. and Carter, R. 2004. Formulaic sequences in action. In Schmitt N. (ed.). *Formulaic Sequences: Acquisition, processing and use*. Amsterdam: John Benjamins.

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

Schmitt, N. and McCarthy, M. (eds.) 1997. *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Swan, M. 1997. The influence of the mother tongue on second language vocabulary acquisition and use. In Schmitt, N. and McCarthy, M. (eds.). *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.

Thornbury, S. 2002. *How to teach vocabulary*. Longman.

Wray, A. 2002. *Formulaic Language and the Lexicon.* Cambridge: CUP.