Tampereen yliopisto

Kati Laitinen

‘THE DREAM ONE WOULD LIKE TO COME TRUE’

A corpus study on how the word *one* is utilized by L2 English speakers

Faculty of Information Technology and Communication Sciences

MA Thesis

April 2022

ABSTRACT

Kati Laitinen: 'The dream one would like to come true': A corpus study on how the word *one* is utilized by L2 English speakers
Master's thesis
Tampere University
Master's Programme in English Language and Literature
April 2022

This thesis examines the use of the English word *one* by L2 speakers of English. The aim of the study is to see if there are differences in the usage of the word *one* by L2 English speakers and if their L1 interferes with L2 English. The material consisted of the International Corpus of Learner English (ICLE), but the material that was analysed in this study was limited to three subcorpora of ICLE. The three subcorpora included texts that were written in L2 English by L1 speakers of Finnish, Swedish, and Spanish.

This study focuses solely on the English word *one* because it can function as a member of different word classes depending on the context, and it is possible that there are challenges in the use of the word in different contexts that are connected with the (non)occurrence of similar constructions in the first language of the English learner. Based on the material, different categories were created in order to analyse the usage of the word *one* in different contexts. The different categories were the following: pronouns, determiners, numerals, the *one of* -structure, the genitival expression *one's*, idiomatic expressions, errors, and irrelevant instances in the data.

The results indicate that L1 speakers of Finnish and Swedish use the word *one* in L2 English rather similarly and they seem to be competent in L2 English as they do not make many errors. Nevertheless, L1 speakers of Spanish stand out in the results and especially when compared to L1 speakers of Finnish and Swedish. L1 speakers of Spanish make more mistakes in L2 English, and they use the word *one* considerably more as regards some categories and respectively, considerably less in some categories when compared to Finnish and Swedish subcorpora. The way L1 speakers of Spanish use the word *one* in L2 English indicates that they are not that competent in L2 English which would explain the differences between the Finnish, Swedish, and Spanish subcorpora. However, there were also indications that the L1 interfered with L2 English, especially as regards the category of errors in the material.

Since this study focused solely on the word *one* in L2 English, this topic would require more research. For instance, L2 English material could be compared to L1 English material and the research could focus on a larger topic.

Keywords: corpus linguistics, L2 English, L1 influence

TIIVISTELMÄ

Kati Laitinen: 'The dream one would like to come true': A corpus study on how the word *one* is utilized by L2 English speakers
Pro gradu -tutkielma
Tampereen yliopisto
Englannin kielen ja kirjallisuuden maisteriopinnot
Huhtikuu 2022

Tämä pro gradu -tutkielma tarkastelee, miten L2 englannin puhujat käyttävät englannin *one*-sanaa. Tutkimuksen tarkoituksena on selvittää kuinka L2 englannin puhujat käyttävät *one*-sanaa, onko heidän kesken sanan käytössä eroja ja häiritseekö heidän L1 kielensä L2 englantia. Tutkimuksen materiaali koostui ICLE-korpuksesta, mutta materiaalista analysoitiin vain kolme alakorpusta. Nämä kolme alakorpusta sisälsivät englanninkielisiä tekstejä, jotka olivat kirjoittaneet äidinkieleltään joko suomen, ruotsin tai espanjan kielen puhujat.

Tämä tutkimus keskittyy pelkästään *one*-sanan tarkasteluun, koska se voi toimia eri lauseenjäsenenä eri konteksteissa ja rakenteissa. Materiaalin perusteella määriteltiin eri lauseenjäsen- ja rakennekategoriat, joiden avulla *one*-sanaa analysoitiin eri konteksteissa. Tarkasteltavat kategoriat olivat seuraavat: pronominit, determinantit, numeraalit, *one of*-rakenne, genetiivinen ilmaisu *one's*, idiomaattiset ilmaisut, virheet ja epäolennaiset tapaukset materiaalissa.

Tulokset osoittavat, että suomea ja ruotsia L1-kielenä puhuvat käyttävät englannin *one*-sanaa L2 englannissa melko samalla lailla ja he vaikuttavat tältä osin olevan päteviä L2 englannin käytössä, koska he eivät tee juurikaan virheitä. Kuitenkin espanjaa L1-kielenä puhuvat erottuvat tutkimuksen tuloksissa varsinkin verrattaessa suomea ja ruotsia L1-kielenä puhuviin. Espanjaa L1-kielenä puhuvat tekevät enemmän virheitä L2 englantia käyttäessä ja he käyttävät englannin *one*-sanaa selkeästi enemmän joissain kategorioissa ja vastaavasti selkeästi vähemmän joissain kategorioissa verrattuna suomen ja ruotsin alakorpukseen. Se, miten espanjaa L1-kielenä puhuvat käyttävät *one*-sanaa L2 englannissa osoittaa, että tämän sanan käyttötapojen suhteen he eivät ole niin päteviä L2 englannissa, mikä selittäisi erot alakorpusten välillä. Tutkimuksen tulokset viittaavat myös siihen, että L1-kieli sekaantui L2 englantiin ja tämä oli erityisesti nähtävissä virhekategoriassa.

Koska tämä tutkimus keskittyi pelkästään englannin *one*-sanaan L2 englannissa, tämä tutkimusaihe vaatisi enemmän tutkimusta. Aihetta voisi tutkia esimerkiksi niin, että englannin L2 materiaalia verrattaisiin vastaavaan L1-englannin materiaaliin ja tutkimuksen aihe voisi olla laajempi.

Avainsanat: korpuslingvistiikka, L2 englanti, L1 vaikutus

TABLE OF CONTENTS

1 INTRODUCTION 1

2 THEORETICAL FRAMEWORK 4

2.1 Language Acquisition 4
2.2 L1 Interference 6
2.3 Corpus Linguistics 11
2.4 Corpus-driven Approach 13
2.5 Pattern Grammar 14
2.6 Previous Studies 17
2.6.1 Housen (2002) 18
2.6.2 Westergaard (2003) 19
2.6.3 Erarslan and Hol (2014) 20
2.6.4 Axelsson and Hahn (2001) 22

3 MATERIALS AND METHOD 24

3.1 ICLE Corpus 24
3.1.1 The Finnish Subcorpus 26
3.1.2 The Swedish Subcorpus 27
3.1.3 The Spanish Subcorpus 27
3.2 Method in the Analysis 27
3.3 Different Categories 28
3.3.1 Pronouns 30
3.3.2 Determiners 31
3.3.3 Numerals 32
3.3.4 The *One of*-structure 33
3.3.5 The genitival expression *One's* 34
3.3.6 Idiomatic Expressions 35
3.3.7 Errors 36
3.3.8 Irrelevant instances in the data 37

4 ANALYSIS 39

4.1 Pronouns 39
4.2 Pronoun subjects 41
4.3 Determiners 42
4.4 Numerals 43
4.5 The *One of* -structure 45
4.6 The genitival expression *One's* 46
4.7 Idiomatic Expressions 47
4.8 Errors 48
4.9 Irrelevant hits 49

5 DISCUSSION 51

6 CONCLUSIONS 61

REFERENCES 63

1 INTRODUCTION

It is natural to rely on prior knowledge when learning something new. For instance, when learning a new language, it is part of the learning process trying to connect the new elements to the prior linguistic and other knowledge that someone might already have (Ringbom 2007, 1). In this regard, it seems natural that language learners may transfer some elements of their first language (L1) to their second language (L2) during the learning process.

L1 interference on L2 is a relevant and important research topic because the more there is knowledge on how L1 can affect learning L2, the easier it will be to understand what helps in the learning process and what respectively stands in the way of the learning process. Cook (2016, 17) claims that the elements that L1 and L2 have in common help in the language learning process and correspondingly, the elements that L1 and L2 do not share can hinder the L2 learning process.

This study focuses on studying and analysing different patterns that include the word *one* as a part of the pattern. Moon (1998, 28) presents Sinclair's (1987, in Moon 1998) *Idiom Principle* which includes two principles underlying language. These two principles are *the open choice principle* and *the idiom principle*. The open choice principle means that language is seen as a text that includes a large number of complex choices that have been made in order to produce language. Moreover, every point where a choice is being made to complete a unit there are several choices "and the only restraint is grammaticalness" (ibid.). However, as regards the idiom principle the choices that can be made include a great number of semi-preconstructed phrases which form single choices. In addition, these single choices may appear to be analysable further into segments. Therefore, Moon (1998, 28) states that at a point in text where the open choice

model would allow several possible choices, the idiom principle constrains the choices due to the predictable semantic restraints that result from the topic or the context. According to Moon (1998, 29), Sinclair has illustrated the difference between the open choice principle and the idiom principle with the example *of course*. According to the open choice principle, the combination *of course* would include two free choices, the first slot being fulfilled by *of* and the other slot by *course*, whereas the idiom principle suggests that the combination *of course* is a single choice which happens to occupy two slots. In other words, the idiom principle considers *of course* as a whole and as a pattern and respectively, the open choice principle sees *of course* as a random choice of two words which does not have a specific meaning unlike the idiom principle, which considers *of course* as a pattern with a specific meaning.

The principles and reasons for identifying a pattern or a structure with a specific meaning is a crucial subject regarding the present study as the contexts in the material where the word *one* occurs include different patterns but also just random word choices which are not regarded as patterns with specific meanings in context. The present study presents nine different categories for the word *one*. Some of the categories are considered as patterns or structures which have specific meanings but some of the categories represent situations where the word *one* is just a random word choice. Some of the categories represent most of the instances where the word *one* occurs in the material and some of the categories are relatively small, especially when compared to other categories. Nonetheless, all the present study's categories are relevant and necessary in order to illustrate different scenarios and contexts where the word one can occur, can be utilized, and how its meaning changes according to the context where it is found.

To sum up the topic and focus of the present study, this thesis aims to provide answers for the following research questions:

1) How are the uses of the word *one* different when it is used by L2 English speakers whose L1 is either Finnish, Swedish or Spanish?
2) How does L1 interfere with L2 English when it comes to the usage of the word *one*?

The material of the present study consists of the ICLE corpus, which is in other words the International Corpus of Learner English. Nevertheless, the whole corpus is not being studied in the present study but only three of its subcorpora are taken into consideration. These three subcorpora are the Finnish, Swedish, and Spanish subcorpora. The idea behind the ICLE corpus is that it contains texts that are produced by L2 speakers of English and that the participants have different L1 backgrounds so that the material can be used to study differences in L2 English.

In order to illustrate the differences between the different categories and their portion compared to other categories in the material, tables including numbers and percentages of different subcorpora are provided in the analysis chapter. The results are presented and explained in more detail in the Chapter 4. In addition, possible reasons for the results and differences that are found in the material are presented and discussed in the Chapter 5.

2 THEORETICAL FRAMEWORK

The central topics and terms that comprise the theoretical framework of this study include language acquisition, L1 interference, corpus linguistics, corpus-driven approach, pattern grammar, and previous studies that are somewhat similar regarding the research topic with the present study. All the terms and concepts are explained in more detail in the paragraphs below and the theory related to these terms will be covered from different point of views that are considered being relevant regarding the present study. After the introduction of the general topics and concepts of the thesis, the final part of the background section discusses the ways in which similar research topics have been conducted earlier, what has been detected from these studies, and why these studies are relevant with regards to the present study.

2.1 Language Acquisition

In this section, the term *language acquisition* and the theory related to it is explained and discussed from different perspectives. As a general concept, language acquisition involves a variety of different point of views regarding the linguistic research field. Thus, it is necessary to explain the term more accurately.

According to Biber et al. (1998, 172), research regarding language acquisition and development includes three major areas that researchers have mostly concentrated on. These three areas are first language acquisition of very young children, later language development, e.g., the acquisition of literacy skills by students at several different ages, and second language acquisition by both children and adults. Levey (2019, 11) defines language as “the means for human communication through the use of spoken words, written symbols, or sign language”. However, language acquisition refers to the way a

person learns a language. “This term is appropriate when it comes down to explaining how this acquisition proceeds in general without focusing on a specific language”, according to Königs (2014, 17). Furthermore, it is noted that the acquisition of a first language and the acquisition of additional languages can be distinguished from each other. The same observation is made by Jordens and Lalleman (1996, 3) who state that second language acquisition is furthermore also distinguished from foreign language acquisition.

Acquiring a first language is somewhat a special situation because the human brain and the memory have not yet learned any language and are therefore blank before learning the first language (Königs 2014, 18). Nevertheless, when a child acquires their first language, “they learn the concept *language* which leads to understanding the importance that language has for interacting and cooperating with other human beings” (ibid.). Studies have shown that first language acquisition is determined by a specific processing of mental resources and the physiological predisposition as well. Therefore, it can be stated that “first language acquisition generally proceeds at the same time as the development of the brain and memory” (ibid.).

Since first language acquisition occurs simultaneously with the development of the brain and memory, it can be concluded that learning additional languages differs considerably from learning one’s first language. Königs (2014, 18) notes that “[i]n contrast to these forms of first language acquisition, we always have to state the fact that for all the other forms of acquisition of a second or further language the acquisition process takes place against the background of the first language acquisition”. In other words, this means that as a child acquires their first language, they learn the special characteristics of their first language and have their first language as a base for other

languages they might learn later on. Ringbom (2007, 5) notes that "[t]he search for similarities is an essential process in learning". Moreover, it is natural to establish an association between a new task and what already exists in one's mind. Furthermore, Cook (2016, 17) states that L1 acquisition and L2 learning may be similar in many instances, but there is still a crucial difference between them because in L2 learning the L1 is always present in the background in L2 learning process.

Additionally, the acquisition of a foreign language can be differentiated in general as to whether the process in question underlies an instructional control or not. When it comes to the terminology, "this has led to the distinction between foreign language learning and second language acquisition" (Königs 2014, 18). Ellis (2000, 3) argues that the word *second* in the term second language acquisition "can refer to any language that is learned subsequent to the mother tongue". In addition, the term *L2 acquisition* refers to the way in which people learn an additional language inside or outside of a classroom. Jordens and Lalleman (1996, 4) point out that the difference between the terms *learning,* and *acquisition* is not that consistent in the linguistic field and that many researchers use the term *instructed second language acquisition* when they mean learning of a language through formal instruction. Furthermore, "[f]oreign language acquisition (FLA) is considered to be the learning of a second (third, fourth, etc.) language outside of its own language area" (ibid.).

2.2 L1 Interference

This section introduces the concept *L1 interference*, what it means in practical terms and what sort of theories have been proposed considering L1 interference on L2. It is relevant to note that in this study L1 refers to the language that one has learned as their first

language and L2 refers to a language that one learns or has learned after learning their L1. This means that L1 and L2 are not acquired at the same time and, therefore the acquisition processes of L1 and L2 are considered being different from each other.

As already noted in the section 2.1, Brown (1998, 137) has stated that "[o]ne such factor that distinguishes second language acquisition from first language acquisition is the fact that the second language learner comes to the task of acquisition already knowing a language". Thus, it is crucial to note that when it comes to learning or acquiring any additional language to the L1, the learner already has a conception of what language is and utilizes their L1 in the learning process. Furthermore, some studies have pointed out that there is a reason to assume that the L1 of the learner does play a role in the L2 acquisition, even though, it is not clear what exactly the role is in the acquisition.

Already in the 1950s, Lado (1957) presented his Contrastive Analysis Hypothesis (CAH) and stated in it that it would be assumed that a student who comes in contact with a foreign language will think that some of its features would be quite easy to understand and some would be extremely hard. In other words, those features or elements that are similar with one's L1 would be considered easy to acquire and those that would be different from one's L1 would be considered hard (Lado 1957, 2). Cook (2016, 17) argues that one of the crucial elements of L2 learning is "the transfer of aspects of the first language to the second language". Thus, Cook (ibid.) basically makes the same observation that Lado did already in the 1950s, which is the idea that L1 helps in learning the L2 especially when the two languages in question have elements in common. Respectively, when the languages in question differ from each other, the L1 hinders learning the L2 as noted already in the *Introduction*.

The term *language transfer* can be used in the early stages of language learning when someone who is learning another language relies on their L1 during the learning process (Cook 2016, 17). As noted already in the previous paragraph, language transfer can either help in the learning process or complicate it. According to Ellis (2000, 51), *language transfer* can be divided into *positive transfer* and *negative transfer*. Ringbom (2007, 30-31) notes that *positive transfer* means the similarities between L1 and L2, which supports L2 learning. Nevertheless, *negative transfer* refers to the differences between L1 and L2, which are transferred from the L1 to the L2 in the early stages of learning process. According to Krashen (1981, 67), "[t]he L1 may 'substitute' for the acquired L2 as an utterance initiator when the performer has to produce in the target language but has not acquired enough of the L2 to do this". Krashen (ibid.) continues and argues that "[f]irst language influence may therefore be an indication of low acquisition. If so, it can be eliminated or at least reduced by natural intake and language use".

Moreover, Ellis (2000, 51) argues that "L1 transfer can also result in *avoidance*". This means that L2 learners may avoid e.g., some construction in the L2 because their L1 does not contain the equivalent construction. For instance, studies have shown that Japanese and Chinese learners of English tend to avoid the use of relative clauses when using English due to the fact that their L1 do not recognize an equivalent structure (ibid.). Another case in point where L1 transfer can be found, is the *overuse* of some structures in L2. One example of this tendency is Chinese learners of English who overuse expressions of regret when they use English in a situation where they are apologizing to someone. This Chinese custom is "in accordance with the norms of their mother tongue" (Ellis 2000, 52).

In his book, Hawkins (2019) presents an overview of the findings relating to L1 interference in L2 learning. Hawkins (2019, 288) names five problems that L2 learners of a certain language need to solve when studying another language. The first two problems are segmentation and categorisation of morphemes of input and then the remaining three problems have to do with solving the syntax, semantics, and context problems. All in all, these five problems are related to establishing a mental grammar for the target language. Nevertheless, one must be careful when it comes to making any claims about the effects of a speaker's L1 on their knowledge of an L2 on the grounds of similarity between the properties of L1 and what is accepted in the L2. Hawkins (2019, 155) gives an example where "Spanish determines the (a)telicity of dynamic, durative transitive verbs like *make* through the presence or absence of an overt determiner in the accompanying object: *made a cake* (telic) vs *made cakes* (atelic)". Hawkins (ibid.) continues that "[t]he Slavic languages determine (a)telicity in such cases through the presence of a verbal prefix: *s-jela tort* 'ate the cake' (telic) vs *jela tort* 'ate cake' (atelic)". Cook (2016, 17) mentions another case in point; "Spanish speakers may leave out the subject of the sentence when speaking English, saying 'Is raining' rather than 'It's raining', while French speakers do not. The explanation is that subjects may be omitted in Spanish, but they may not be left out in French". Therefore, L1 interference on L2 depends on the languages themselves and their properties as well.

How exactly L1 influences L2, and what role input plays in learning e.g., the acquisition of movement properties in L2 are naturally questions, which scholars have attempted to examine in greater detail. Hawkins (2019, 93) notes that there seems to be a trade-off between a learner transferring the movement properties of categories that exist in the L1 to the L2 and a learner who identifies the movement properties of L2 categories

from samples of the language in question. According to Hawkins (2019, 99), there are three syntactic rules that have to do with how the sentence structure is learned in L2 and one of the syntactic rules is called *agree* and a case in point is the subject-verb agreement that exists in English. This syntactic rule means that a subject and a verb must be in agreement with each other in a sentence. For instance, if the subject is a third person pronoun *she* the verb needs to have the affix *-s* at the end of the verb in present tense in order to be grammatically correct.

According to Hawkins (2019, 151), "L1 might play a role in the way that L2 speakers deal with the challenge of acquiring phrasal and sentential meaning in the target L2". For instance, Hawkins (2019, 151) mentions that Spanish and Japanese learners of English tend to allow verb-framed clause structures in order to indicate manner of motion along a path. However, allowing these verb-framed sentence structures might be something that basically all L2 learners do and that it would not be dependent on their L1. In other words, it could simply be a part of the L2 learning process and development in general. Nonetheless, Hawkins (ibid.) states that L2 learners from different L1 backgrounds should be compared in order to take a stand regarding L1 influence on L2. In addition, he continues that in order to compare and study the background effects of L1, one group should have L1 which has a property similar to the L2 and another group whose L1 would be different in that regard to the L2 in question. If there is clearly different behaviour between these two groups and the learners are at the same stage of general development of knowledge of the target L2, there would be good grounds for suggesting that L1 does play a role in guiding the development.

When it comes to learning the grammar of L2, Hawkins (2019, 297) states that some of the influences that L1 may have on L2 are short-lived and other influences

are lasting. According to Schwartz and Sprouse (1996), there is a hypothesis which is the initial state of L2 learning. That state is the L1 grammar as a whole but without the forms of L1. Therefore, the instances where L1 and L2 differ from each other, there is a change towards the L2 that happens with encounters with input in working memory which collide with the L1-transferred property. In addition, Hawkins (ibid.) argues that "[r]estructuring can occur quickly, can take time or may never occur at all if the evidence required from input is ambiguous, obscure or unavailable". Furthermore, Hawkins (ibid.) also states that regarding Schwartz and Sprouse's (1996) proposal, it contains a sudden change from the L1 to the L2 property. Nevertheless, it seems that L2 development happens gradually with long periods of optionality.

2.3 Corpus Linguistics

As the empirical section of the present thesis involves the analysis of data from learner corpora, it is necessary to first elaborate what is meant by the term *corpus linguistics* and how it is applied in the field of linguistics. To begin with, the concept *corpus* comes from Latin, and it means a body. Nevertheless, when it comes to linguistics, a corpus means a collection of texts, and therefore it could be referred to as a "body" of language, which is stored in an electronic database. In general, corpora include texts that contain thousands or millions of words that can be analysed electronically (Baker et al. 2006, 48).

According to McEnery and Hardie (2012, 1), corpus linguistics is not especially about the study of any specific aspect of language but rather it is a field that focuses on a set of procedures or methods for studying language. However, one school of corpus linguistics does not agree with defining corpus linguistics as a methodology (ibid.). According to Sinclair (1991, 4), corpus linguistics is about investigating language

by observing great amounts of naturally-occurring and electronically-stored discourse. Corpus linguistics also includes using a software which selects, sorts, matches, counts, and calculates the data according to what is being studied. Sinclair (1991, 16-17) continues that the material in different corpuses ranges from formal to informal and that one of the major reasons for using a corpus to study language is that with the help of a corpus it is possible to identify what is central and typical in a certain language and its usage. It is said that most of the corpora try to cover a specific period of time and they should be as large as possible in order to be good tools to analyse languages.

Granger et al. (2002, 4) note that even though corpus linguistics has only existed since the late 1980s it has become an important link between the two previously disparate fields of corpus linguistics and second/foreign language research. With the help of corpus linguistics' main principles, tools and methods, corpus linguistics aims to provide enhanced descriptions of learner language which can be utilized for several purposes in second or foreign language learning research and furthermore, to improve foreign language teaching. In addition, corpus linguistics is best defined as a linguistic methodology which is based on the use of electronic collections of naturally occurring texts, namely corpora. According to Granger et al. (2002, 4), corpus linguistics is not a new branch of linguistics and neither it is considered as a new theory of language. However, it is stated that corpus linguistics is a very useful and powerful methodology due to the evidence it uses. Therefore, corpus linguistics has the power and potential to change perspectives on language as it already has contributed to the discovery of new facts when it comes to studying languages (ibid.).

2.4 Corpus-driven Approach

As regards the corpus linguistics and how it is applied in the linguistic research field, there are two different approaches which are based on corpora. The first one is *corpus-based approach* and the second one is *corpus-driven approach*. The idea of this section is to explain and clarify what these two different approaches specifically refer to and how they differ from each other. Furthermore, it is relevant to state reasons how and why the corpus-driven approach is the approach that is used in the present study.

According to Tognini-Bonelli (2001, 84), in corpus linguistics, corpus-driven approach means that "the linguist uses a corpus beyond the selection of examples to support a linguistic argument or to validate a theoretical statement". When using a corpus-driven approach a linguist is committed to the integrity of the data as a whole and therefore, the descriptions attempt to be extensive with respect to the corpus evidence. In a corpus-driven approach, the theoretical statements reflect and are clearly consistent with the evidence that is found in the corpus. It has been stated that many of these kinds of statements could not be accessible by any other means than analysing a corpus. Moreover, Tognini-Bonelli (2001, 84) has noted that when using the corpus-driven approach the researcher is expected to study the data without prior assumptions or expectations.

The differences between the corpus-driven and corpus-based approaches have to do with the aims, prior knowledge etc. As McEnery and Hardie (2012, 5-6) note, the corpus-based approach uses the data of a corpus to explore a theory or a hypothesis in order to confirm it, disprove it or refine it, whereas Tognini-Bonelli (2001, 84) state that the corpus-driven approach is considered to include its own theory of language in the corpus itself. In addition, Sinclair (1991, 4) states that the data that is the basis for a corpus-driven approach is authentic, is not selected on linguistic grounds, there is lots of

data, it is systematically organized, and it is not annotated in terms of existing theories. Moreover, Biber et al. (1998, 9-10) stated that the corpus-based approach to a study "should be seen as a complementary approach" and that "research questions for corpus-based studies often grow out of other kinds of investigations". Even though there are differences between these two approaches, it should be noted that the difference between the corpus-driven approach and the corpus-based approach is not necessarily that straightforward in every circumstance.

The corpus-driven approach was chosen for the current study because the material is the ICLE corpus, in other words International Corpus of Learner English, which contains a great amount of data. In addition, the data is authentic, and it is not annotated according to any existing theories. Moreover, the current study is not based on any previous study and the goal of the current study does not include complementing any previous studies. Consequently, the corpus-driven approach is the most suitable approach for the present study.

2.5 Pattern Grammar

As the pronoun *one* occurs in many different patterns in English, it is relevant to include the concept of *pattern grammar* in the theoretical background of the present study. Therefore, the point of this section is to explain and elaborate what is being referred to when the term *pattern grammar* is used. It should be noted that pattern grammar is in general considered as an orientation rather than just one person's thoughts on the matter, even though this chapter relies largely on the observations and views of Hunston and Francis (2000).

The concept of *pattern grammar* has been defined as all the structures and words that are frequently associated with a particular word and which contribute to its meaning (Hunston and Francis 2000, 37). Furthermore, if a combination of words appears fairly frequently, and if the combination in question is dependent on a specific word choice and moreover, there is also an explicit meaning that is associated with the combination, then there is a pattern which can be identified. There are different types of patterns and, for instance, Hornby (1975) categorises different patterns according to the part of speech. Therefore, e.g., verb patterns, noun patterns, and adjective patterns can be recognized as different examples of patterns.

In order to investigate a word pattern, the data should be organized according to the concordance lines which are usually in alphabetical order (Hunston & Francis 2000, 37). The data can be sorted according to the word that is being examined either to the right or to the left. This means that if the data is sorted to the right, then that word that follows the word that is being examined is looked at. For instance, it is typical to sort the data to the right when examining a verb because most verbs have complementation patterns that follow them and not precede them. Even though a pattern of a specific word usually consists of the elements that follow the word, it is also possible to include elements that precede the word to the pattern of a word (Hunston & Francis 2000, 51). Nevertheless, it could be interesting to sort verbs to the left as well because then it would be possible to examine whether the verb appears in the passive or in the infinitive form, if there are any specific modals which are used with the verb, and what kind of subjects it usually has etc.

Moreover, Hunston and Francis (2000, 43) state that one word can have a great number of patterns, but it is also possible for a pattern to be associated with a variety

of different words. However, it is important to consider which elements are included in a pattern and which are not. According to Hunston and Francis (2000, 49), some elements need to be excluded on the grounds that these elements in question can appear with almost any word of the same class. Hunston and Francis (ibid.) give the two following sentences as examples:

(a) I have come to the decision that the earring discovery scene does not work
(b) … to counteract a decision that required a committee's questioning of a witness

Hunston and Francis (2000, 49) argue that the sentence (a) "is an appositive or defining clause", which qualifies the word decision whereas the sentence (b) is just a typical relative clause. Hunston and Francis (ibid.) state that "[a] defining that-clause is considered to be part of the complementation pattern of a noun; there are only a few hundred nouns that are typically followed by such a clause". In addition, ordinary relative clauses can qualify basically any noun and therefore, they are not classified as part of their complementation patterns.

As noted earlier, corpus linguistics might seem like an appealing research field due to the simplicity of exploring data with the help of automatic processing of the texts. In addition, Hunston and Francis (2000, 67) note that it might seem easy to identify a pattern that is associated with a particular word because "the computer would simply have to recognise that a word is often followed by a particular preposition, by a to-infinitive or an 'ing-form', for example, and a pattern would have been identified". Even though using corpus linguistics in identifying patterns of particular words seems quite straightforward, it is not necessarily such a simple and straightforward task because there

are several clear complications that could occur. One challenge with the automatic processing of the texts is that "the computer would have to recognise the different forms that a pattern can take", (ibid.).

2.6 Previous Studies

In general, Biber et al. (1998, 1) have stated that language studies have been divided into two main research areas which are studies of structure and studies of use. They further point out that "[m]any studies of language use focus on a particular linguistic structure, investigating the ways in which seemingly similar structures occur in different contexts and serve different functions" (ibid.). As noted earlier, corpus linguistics has been used in several studies to study languages in general, but it has been especially useful to study what is typical for a specific language, how languages differ from each other, and how standard language differs from spoken language, to name a few research topics (Sinclair 1991, 16-17). Moreover, what is special to corpus linguistics is that it enables the research topic to be restricted to a small and specific research area.

In this section, four previous studies, their research topics and what was discovered in these particular studies are described as examples of specific studies relevant to the present study in that they focus on different restricted research areas in the field of linguistics. For instance, Housen (2002) focused on the acquisition of the English verb system considering L2 acquisition. Another point of view is given by Erarslan and Hol (2014) who studied L1 interference on L2 but from the point of views of vocabulary, simple present tense, and the use of prepositions. Additionally, Westergaard (2003) studied Norwegian children who were also learners of English and how quickly they learned that the verb is not in the second position in a sentence in English unlike in their

L1 Norwegian. Moreover, Axelsson and Hahn (2001) studied how non-native speakers of English use the progressive verb form in English.

2.6.1 Housen (2002)

L2 acquisition of the English verb system has been studied, for instance, by Housen (2002). In Housen's (2002) study, the focus was on the formal and functional development of selected verb categories using a cross-sectional comparison of their distributional patterns in the interlanguage speech. The participants in Housen's (2002) study were 46 young L2-learners from two first language (L1) backgrounds. These two backgrounds were Dutch and French and four different stages in proficiency in L2 acquisition were presented by the participants of the study. The data for the study's analysis was from the Corpus of Young Learner Interlanguage (CYLIL) which consists of English oral interlanguage data elicited from European School pupils. The data included material from participants who were at distinct stages of development and had different L1 backgrounds. These L1 backgrounds in question were Dutch, French, Greek and Italian.

It was discovered in the study that the acquisition of the forms and functions of the verb is a complicated process even for young L2 learners, at least regarding the English language. Housen (2002, 108) states that this acquisition seems to follow "a route similar to that observed for older learners in naturalistic language learning contexts". Moreover, they continue with the following statement: "attempts to explain this process point to a complex interaction between general mechanisms of language processing, universal perceptual-conceptual predispositions, L1-based predispositions and the quantitative tendencies of the input language" (ibid.). Nonetheless, longitudinal case

studies have suggested that there might considerable individual variation when it comes to the route of development even in the cases where the learners share the same proficiency level and L1 background (ibid.).

2.6.2 Westergaard (2003)

According to Hawkins (2019, 294), it is not typical for L2 acquisition to change from one state of knowledge to another instantly but rather it happens gradually over time, as already noted at the end of the chapter *2.2 L1 Interference*. A case in point is a study by Westergaard (2003) who studied how transfer, markedness, and the importance of input cues in the acquisition of word order in English by Norwegian-speaking children. The children that participated in the study were 7-12 years old. The main goal of the study was to see how Norwegian speaking children acquire the basic SVO (subject, verb, and object) word order that English has. Norwegian is a Germanic verb second language (V2) which means that the verb is always in the second position in all types of main clauses. On the contrary, the word order SVO is considered as the basic word order. Westergaard (2003, 78) states that the two languages under investigation, namely Norwegian and English, are not similar as regards the verb movement because "Norwegian is assumed to have verb movement to C (complementizer) in all types of main clauses (even subject-initial ones, which are superficially SVO)". Therefore, V2 word order occurs in topicalized structures, questions, and sentences with adverbials. On the other hand, English allows changes in word order with the help of auxiliaries. Westergaard (2003, 78) illustrates the differences in word order between Norwegian and English with the following example sentence:

a. *I går spilte Peter piano hele dagen.*

yesterday played Peter piano all day

b. “Yesterday Peter played the piano all day.”

The study revealed that the children used the V2 word order largely when writing in English or using English in general. Westergaard (2003, 85) states that there was a significant amount of L1 transfer of V2 word order into the children’s English and it wasn’t only found at the early stages of learning English. As a matter of fact, Westergaard (2003, 85) notes that “this feature seems to persist in some children even after many years of English instruction”. During observation period of the study, which was from the fourth to the seventh grade, the proportion of use of non-V2 constructions gradually increased during the time that participants were observed. Initially, the percentage was 2%, then 14%, then 38%, and lastly 61%, with no apparent sudden shift from V2 to non-V2 construction. Therefore, it can be stated that the children, whose L1 is a V2 language (in this case Norwegian), had to unlearn the V2 word order rule when learning English, which follows the SVO word order. Since the study shows that the percentage grows from 2% at the beginning of the observation period to 61% at the end of the observation period, the study’s results support the claim that there is no sudden change from V2 to non-V2 construction, but it happens gradually over time.

2.6.3 Erarslan and Hol (2014)

Erarslan and Hol (2014) studied how L1 interferes with L2 English regarding transfer on the vocabulary, simple present tense, and the use of prepositions. Erarslan and Hol (2014)

studied Turkish EFL (English as a Foreign Language) learners and how much L1 transfer to L2 English there was. The study included that the participants took a translation test and also performed a translation task. The degree of L1 transfer on the earlier mentioned three topics was evaluated separately and the students' level of English was also taken into consideration. In the study, there were three different levels of English which were elementary, pre-intermediate and intermediate level. The participants of the study were randomly selected based on their level of English.

The results showed that on the elementary level the vocabulary transfer was 35%, on the pre-intermediate level it was 21.90% and on the intermediate level it was 28.33%. As regards the preposition transfer, it was 41% on elementary level, 21.49% on pre-intermediate level and 21.33% on intermediate level. As for the tense transfer, on the elementary level it was 19%, on the pre-intermediate level it was 19.87% and on the intermediate level it was 21.23%. The study revealed that there was most L1 transfer when it came to the use of prepositions and vocabulary following it. In addition, the participants were influenced by their L1 in cases or situations where they had to guess the meaning of a specific word, phrase, or sentence. On the contrary, the influence of L1 was smallest in the use of the simple present tense. The study revealed that the L2 errors that the participants made were traceable to their L1 and therefore, it can be concluded that the L1 Turkish definitely interfered with the L2 English. Since the participants were influenced by their L1 in contexts where they did not know the meaning of a word or a sentence, it can be concluded that L2 is affected by one's L1 especially if one does not have a great knowledge of the L2.

2.6.4 Axelsson and Hahn (2001)

As regards the use of the progressive in English, Axelsson and Hahn (2001) studied how non-native speakers of English use the progressive form when they use English language. In their study, Axelsson and Hahn (2001) studied L1 speakers of Swedish and German. The material of the study consisted of two samples of advanced learners' essays, the sample of native German speakers' argumentative essays was part of the International Corpus of Learner English (ICLE) and the sample of native Swedish speakers' essays was from the Uppsala Student English corpus (USE). Axelsson and Hahn (2001, 16) observed that "[n]either German nor Swedish has a grammatical means of expressing progressive aspect, so it would be natural to hypothesize that learners with German or Swedish as their first language neglect to use the progressive (instead of overusing it, as has been claimed)".

As for the results of Axelsson and Hahn's (2001) study, they note that there were differences in the content matter and general approach to the task, which was the argumentative essay, between the native Swedish and native German participants. These differences turned out to be surprisingly significant considering comparing the data between Swedish and German participants and therefore, it is important to note that the differences might have linguistic consequences which could have an impact on the results of the study.

The results indicate that based on the German and Swedish students' argumentative essays there might a tendency to slightly overuse the progressive in English. However, the two participant groups were basically similar in using the progressive and therefore, there were no significant differences between the groups. Moreover, the results reveal that both German and Swedish students understand well

when to use the progressive and the non-progressive. Nevertheless, the study also revealed that most students did not quite understand the rules that apply to the use of the progressive and the non-progressive. Thus, it can be stated that even though the students might not understand or remember the rules of the progressive and non-progressive, they still know how and when to use them correctly, at least almost. There were some instances produced by the Swedish students that were considered ambiguous because in some situations it was possible to choose either an informal or a more formal expression. As Swedish learners are used to hearing more colloquial and informal language, it can be argued that it might be one of the reasons why Swedish students tend to choose the more informal option because they are more familiar with it. In conclusion, unlike Axelsson and Hahn (2001) expected there was not a significant overuse of the progressive by native German and Swedish speakers in their English essays.

3 MATERIALS AND METHOD

This chapter provides detailed information about the material and method that were utilized in the present study. The material of this study consisted of the second version of the ICLE (International Corpus of Learner English) corpus. However, the whole ICLE corpus was not utilized but instead three different subcorpora of the ICLE corpus were. These three different subcorpora included the materials that L1 Finnish, Swedish, and Spanish speakers had written for the ICLE corpus project as L2 English speakers. Detailed information about the ICLE corpus as well as the Finnish, Swedish, and Spanish subcorpora of the ICLE corpus will be given in this chapter. Moreover, this chapter provides information about how the data was organized and analysed and what tools were utilized in the process. As noted in the section 2.4, the method that was chosen for the current study was the corpus-driven approach.

3.1 ICLE Corpus

The material of the current study consists of the second version of the ICLE (International Corpus of Learner English) corpus. "The whole ICLE corpus consists of 4.5 million words in total, which is about one million words more than the first version of this corpus" (Granger et al. 2009, 1). In addition, the second version of the corpus contains material outside of Europe, unlike the first version. For instance, data from China, Japan and South Africa were included in the second version, as well as two new subcorpora from Europe which included Norway and Turkey (ibid.). All in all, it can be stated that the ICLE corpus consisted of L2 speakers of English from different countries and their written texts in English. Nevertheless, as regards the present study, the whole ICLE corpus was not

utilized. Instead, only the texts that L1 speakers of Finnish, Swedish, and Spanish had written in English were studied.

Already at the beginning of the ICLE corpus project, the requirements for the corpus design criteria were clear. Granger et al. (2009, 3) mentioned the following requirements:

- learners: young adults (university undergraduates); advanced proficiency level; learners of English as a Foreign Language (EFL) rather than as a Second Language (ESL);
- language: academic writing (mainly argumentative); 200,000 words per subcorpus.

The main goal of the ICLE corpus project was to collect learner English productions "that shared a large number of task variables, notably in terms of medium (writing), genre (academic essay), field (general English rather than English for Specific Purposes) and length (between 500 and 1,000 words)" (Granger et al. 2009, 4). Nevertheless, "it depended on the national coordinators of the project what the topics and other task variables e.g., timing, exam conditions, and use of reference tools were" (ibid.). When it comes to the genre of the written texts, the majority (91%) of the ICLE data were argumentative essays. However, since it might be difficult to collect this type of material it was noted that 25% of the collected data could be literary essays, if necessary. Therefore, the proportion of the argumentative essays in the data ranged from 79% (e.g. Spanish corpus) to 100% (e.g. Chinese corpus) (ibid.).

The essential information about the data regarding the present study has to do with the Spanish, Swedish, and Finnish subcorpora. As noted already, the Spanish corpus' proportion of argumentative essays was only 79% whereas that of the Swedish corpus' was 85% and the Finnish corpus' 92% (Granger et al. 2009, 5). In addition, all the essays

in the data had approximately 617 words, although there were great differences between the subcorpora, for instance the Tswana subcorpus had an average of 384 words and the Dutch subcorpus had an average of 893 words (ibid.).

3.1.1 The Finnish Subcorpus

As noted already, the data of the present study consists of the three subcorpora of the ICLE corpus, namely Finnish, Swedish, and Spanish. Nevertheless, it is necessary to note that the data of the present study was organized according to the participants' L1, which were Finnish, Swedish, and Spanish. However, the Finnish subcorpus includes both L1 speakers of Finnish and Swedish because there are Swedish-speaking Finns in Finland.

All in all, "the Finnish subcorpus included three different groups of speakers which were categorised as speakers of Finnish, speakers of Swedish from Finland and bilinguals" (Granger et al. 2009, 30-31). Since the L1 of Swedish-speaking Finns is Swedish, the data that was collected from Swedish-speaking Finns, and was named as Swedish from Finland, is analysed with the data that was collected from L1 Swedish speakers in Sweden. As for the group of bilinguals (people who know Finnish and Swedish equally well) in the subcorpus, this group of speakers is not taken into consideration in the study because the goal of the research is to see whether one's L1 affects one's L2 in some way. Therefore, the texts that the bilingual group had written are not analysed in the present study.

The total number of essays in the Finnish subcorpus is 390 and the number of words is 274,628 in total (Granger et al. 2009, 30-31). Nevertheless, since the part that is only taken into consideration in the analysis as regards the Finnish subcorpus is the group

of people whose L1 is Finnish, the total number of essays is then 229 and the total number of words is 170,976 in total (ibid.).

3.1.2 The Swedish Subcorpus

Most of the Swedish subcorpus data comes from the University of Lund but some of the data was also collected at the University of Vaxjo and the University of Göteborg (Granger et al. 2009, 37). Nonetheless, since there are Finns whose L1 is Swedish, namely the group of Swedish from Finland in the Finnish subcorpus, their texts are analysed together with L1 Swedish speakers from Sweden. The Swedish subcorpus itself includes 355 essays in total and the total number of words in the corpus is 200,033 (ibid.). The Swedish from Finland in the Finnish subcorpus includes 123 essays and 80,230 words in total (ibid., 30). Since the data of the group of Swedish from Finland is analysed with the rest of the L1 Swedish material, the total number of essays is then 478 and the total number of words is 280,263.

3.1.3 The Spanish Subcorpus

The total number of words in the Spanish subcorpus is 198,131 and the number of essays is 251 in total. Basically, the whole data was collected at the Universidad Complutense de Madrid with the exception of one batch, which was collected at the University of Alcala (Granger et al. 2009, 37).

3.2 Method in the Analysis

The method that was utilized in the analysis of the current study was the corpus-driven approach, which was explained in more detail in the section 2.4.

In order to analyse the data, it had to be organized. The Finnish, Swedish, and Spanish subcorpora were analysed separately. Since the research topic was the word *one*, the data was organized according to the word *one* and the instances where it occurred in the data were organized in alphabetical order. In order to organize the data according to the word *one* and all its instances in alphabetical order, the AntConc software was utilized. With the help of AntConc, the data could easily be organized and therefore, also analysed. The analysis included analysing all the instances where the word *one* occurred and categorizing these instances into different categories. Based on the material that was analysed, the following categories were chosen for the study. Present study's categories for the word *one* are the following:

- pronoun
- determiner
- numeral
- *one of* -structure
- *one's* genitive
- idiomatic expression
- error
- irrelevant

These categories and what instances they include are covered with examples in more detail in the following paragraphs.

3.3 Different Categories

This section presents the different categories for the word *one* which were chosen based on the ICLE corpus material. The section also explains on what grounds these categories were chosen for conducting the analysis. Each of the categories includes information in more detail about the category, and also some examples of the instances in which the word *one* occurred in the material.

Even though several different categories were created and there were several examples which were found in the material that fit into these categories, it is not that straightforward to come up with different and clear categories for a particular word and its function in order to analyse the word in question. For instance, *Macmillan Dictionary* and *Collins English Dictionary* had different points of view about how the word *one* should be categorised. Therefore, there were differences e.g., between these two dictionaries and it turned out that they had categorised the different usages of the word *one* differently and the categorisation was based on different grounds. For example, *Collins English Dictionary* has the following example:

a. We ask why peace should have an apparent chance in the **one** territory and not the other. (*Collins English Dictionary*)

The *Collins English Dictionary* categorises the instance of the word *one* above as an adjective, whereas the *Macmillan Dictionary* categorises the word *one* in the example b below as a determiner.

b. He grew roses on **one** side of his garden, and vegetables on the other. (*Macmillan Dictionary*)

Hence, some sources categorise some of the instances of the word *one* as members of word class differently. Therefore, some dictionaries do not share the same point of view on how to categorise particular words and their functions in different contexts.

3.3.1 Pronouns

The category for pronouns includes all the instances of the word *one* where it functions in general as a pronoun. Nonetheless, the generic subject *one* is analysed separately in the analysis even though it is included in the pronoun category.

1. Who can guarantee that a public defender can defend you as successfully as a top lawyer, if you can not afford **one**? (Finnish subcorpus, FIJO1018.txt)

In example 1, the word *one* functions as a pronoun, more specifically as an object pronoun. In this example, the word *one* refers anaphorically to the "the top lawyer" in the main clause.

2. Since it is very difficult to change the way adults think, **one** would have to start educating youngsters at an early age, the younger the better. (Swedish subcorpus, SWUL1017.txt)

3. The dream **one** would like to come true. (Swedish subcorpus, SWUL3017.txt)

Examples 2 and 3 include instances of the pronoun subject, namely the generic subject, *one*. In this case, the word *one* does not refer to any previous word in the context and therefore, it functions as a generic subject in the sentences.

4. You cannot make everybody substitute petrol for unleader fuel, but you can change your old car for another **one** with catalytic converter and try to spread that idea to your friends. (Spanish subcorpus, SPM03022.txt)

Example 4 is an instance where the word *one* is a part of the structure *another one* and it functions as a pronoun, substituting the word *car*. There were also similar instances where

the word *one* was a part of a structure but functioned as a pronoun like, for example *each one*.

5. He was the **one** who made a fortune selling alcohol during the prohibitory law. (Finnish subcorpus, FIJY1042.txt)

In example 5 above, the word *one* is a part of a relative clause structure *the one who*. This structure is widely used, and it occurred in the data frequently. There were also other instances of the relative clause structure, which were *the one that* and *the one which*.

6. I don't believe that **one** or the other controls the world. (Swedish subcorpus, FIAB1068.txt)

Example 6 includes an example of the structure *one – the other*. In general, this structure appears in the material several times but in most of the examples the structure is considered to function as a determiner. However, in the example above the structure is considered to represent a pronoun because the word *one* or *the other* do not have a main word in the sentence. Since they do not determine any other word, they are considered being independent, and in this case, they are considered being pronouns.

3.3.2 Determiners

This paragraph presents some of the instances where the word *one* functions as a determiner in the material.

7. In other words, I do not think it is within our power to decide whether to depend on society or not – we always end up depending on it in **one** way or the other. (Swedish subcorpus, FIAB1064.txt)

In example 7, the word *one* is a part of the structure *one – the other*, but in this case *one* functions as a determiner because it precedes the word *way*. Thus, the head of the noun phrase is *way,* and *one* functions as its determiner. In the material, there were also other instances of a similar structures but instead of the word *other* in the structure the word *another* was used. Moreover, the structure *on the one hand*, like in example 8, where *one* functions as a determiner as well was used quite frequently in the material.

8. On the **one** hand, the people considering themselves 'native Swedes' may feel that only assimilation is acceptable. (Swedish subcorpus, SWUL5009.txt)

9. **One** such program could be The Bold And The Beautiful. (Finnish subcorpus, FIJY1007.txt)

Example 9 represents an example where the word *one* occurs as a determiner in the material. Nevertheless, the word *one* does not represent a numeral nor is it part of a structure in this case. The function of the word *one* is a determiner and in Finnish it could be translated into the word "eräs".

3.3.3 Numerals

Even though numerals belong to the same category as determiners generally, in this study they are considered being just numerals which differ from determiners. The reason why numerals are one category on their own is because there were so many cases of numerals in the data. This paragraph gives a few examples where the word *one* functions as a numeral in the context.

10. As we live in a material world, we want to have everything including, of course, cars, more than **one** house, vacations etc. (Spanish subcorpus, SPM03029.txt)

In example 10, the word *one* functions as a numeral because, it is clear that in this context the speaker means more than one and not e.g., more than two.

11. It must be known and spoken all over the world and not just on **one** or two continents. (Finnish subcorpus, FIJO1002.txt)

In the example above, the word *one* functions as a numeral because it is being compared to number two in the structure *one or two*.

12. This book was written for about **one** hundred years ago and the story takes place in New England at the end of the eighteenth century. (Swedish subcorpus, SWUL8032.txt)

In the example 12, the word *one* is a part of a numeral which is *one hundred* and it expresses time in this case.

3.3.4 The *One of*-structure

The structure *one of* is considered to be one category in the present study because it occurs relatively often in the material, and it is widely used in general in English. These cases of the word *one* could also be categorised as numerals, there are examples of numerals regarding this structure in the OED, but because the structure has an equivalent *en av* -structure in Swedish and because it has a high occurrence in the material, it is considered as a category on its own.

13. The Internet is expanding all the time and it could become **one** of the most important media. (Swedish subcorpus, FIAB6014.txt)

14. Nowadays **one** of the great debates in the Spanish society is about the period of military training. (Spanish subcorpus, SPM07002.txt)

Both examples 13 and 14 above include an example of the structure *one of* and represent the same function. This structure occurred relatively often in the material, and it had many instances in all the subcorpora that was analysed.

3.3.5 The genitival expression *One's*

The genitive expression *one's* is considered being a category on its own because the expression is widely utilized. Since the categories should be quite specific, it is natural to categorise *one's* as one category on its own and because there were several examples of the structure in the data.

15. The feeling of knowing the answer to a question when the competitors do not, is really flattering to **one's** self-esteem! (Swedish subcorpus, FIAB1012.txt)

16. It's complicated and often confusing to try and make up **one's** mind about what basis the president should be elected on. (Finnish subcorpus, FIJY1023.txt)

These two examples above represent the instances where the genitive *one's* is utilized in the material. Since the genitive *one's* functions as a generic genitive, it is the only function it has and therefore, all the instances of it are similar with each other.

3.3.6 Idiomatic Expressions

The category of idiomatic or phrasal expressions includes instances where the word *one* is used idiomatically, and the structure that can be categorised as an idiomatic expression is an established structure. As will be noted, they are often listed separately in dictionaries. The expressions in question are covered through the examples that were found in the material.

17. Spanish, for **one**, is very important in Latin America but little used anywhere else, except for Spain. (Finnish subcorpus, FIJO1003.txt)

In example 17, the word *one* is part of the expression *for one*, which is an established pattern and has a specific meaning, namely that of 'used for emphasizing that someone is thinking or behaving in a particular, even if other people are not' (*MacMillan Dictionary*, s.v. *for one*). This particular pattern was only found in the Finnish subcorpus and there were only four instances where it occurred.

18. For example when Mr Worthing is declaring his love to Wendolyn she asks him to follow the correct steps **one** by one. (Spanish subcorpus, SPM06006.txt)

One by one is an established pattern and structure, which is considered being an idiomatic expression due to its particular meaning and function. This pattern occurred a few times in all the three subcorpora.

19. The North American indians believe that we, human beings, are **one** with Nature. (Swedish subcorpus, SWUL2007.txt)

Example 19 includes an example of an idiomatic use of the word *one*. The example of the word *one* above is an instance of the structure *to be one with*, which is used idiomatically. The pattern to be one with was only found once in the Swedish subcorpus but there were a few examples where the word *one* was used idiomatically without being part of the pattern *to be one with*. For instance, the example 20 below includes another instance of the idiomatic use of the word *one* which was found in the data.

20. But Amy is not one to dream about unachievable things; she has her feet firmly on the ground. (Finnish subcorpus, FITU3008.txt)

The word *one* is frequently used in several different idiomatic expressions (there are various examples of different idiomatic expressions in dictionaries like e.g., OED) but due to the small size of the ICLE subcorpora, the usage spectrum of the word *one* is significantly smaller and therefore, it highlights only the main uses of the word *one*. For this reason, it is also interesting to see whether there are great differences between all the three subcorpora.

3.3.7 Errors

The error category includes all the instances where the word *one* is used incorrectly or there was a spelling mistake which related to the word *one* in the data. Below, there are a few instances which were categorised as errors in the data.

21. In one place, the system and the ideology which punish faults again life, freedom or human dignity can not fall in the same vices; as freedom, life or dignity are basic human rights, all the people are worthy of that, not only those **one** with a good behaviour. (Spanish subcorpus, SPM04056.txt)

Example 21 was categorised as an error because the word *one,* which is bolded in the example, should be *ones* or another option would be to leave it out altogether in order for the sentence to be grammatically correct.

22. Freedom is **one** the pillars of democracy and continues to inspire heated debate. (Finnish subcorpus, FIHE1023.txt)

In example 22, there is an error because the word *one* should be a part of the structure *one of*. However, the preposition *of* has been left out and therefore, the instance is considered an error in the data.

3.3.8 Irrelevant instances in the data

The irrelevant category includes all the instances where the word *one* is considered irrelevant regarding the present study and its research questions. These irrelevant instances of the word *one* are e.g., examples where *one* is a part of another word, it belongs to a compound, or it is a part of a numeral.

23. It wasn't a very happy home-coming since there was no **one** there at the airport to meet her. (Finnish subcorpus, FITU1004.txt)

The example 23 includes an instance which has been recognized incorrectly as an instance of the word *one* by AntConc software when in fact it is a part of the indefinite pronoun *no one*. Since the example does not represent an instance of how the word *one* is utilized in the material, it is considered as irrelevant when it comes to the present study.

24. Many of those who moved to these towns bought **one**-way tickets and went by train. (Swedish subcorpus, SWUL7045.txt)

In the example 24 above, the word *one* is a part of a compound which is *one-way*. Therefore, this instance of the word *one* is irrelevant as regards the present study.

In addition to the examples of the irrelevant instances mentioned above, there were also other cases where the word *one* was considered being irrelevant. These examples were basically errors like e.g., the word *someone* had been spelled incorrectly with a space between *some* and *one* and therefore, the AntConc software had incorrectly counted these instances as cases of the word *one*. Similar errors to that in example 23 had happened with indefinite pronouns like *anyone* and *everyone*.

4 ANALYSIS

This chapter presents the results of the corpus analysis. Since the three subcorpora of the ICLE corpus, namely Finnish, Swedish, and Spanish, are different in size when compared to each other, regarding the total amount of words and the instances of the word *one* that were found in the data, the comparisons between the three subcorpora are made in percentages. In addition, the normalized frequencies of the word *one* in different categories per 100,000 words in the three subcorpora are counted and utilized in the comparisons in the data analysis.

The results of the different categories in the three subcorpora are presented in the tables individually in order to compare the numbers and percentages between all the three subcorpora. There is a table for every category in order to illustrate the numbers and percentages which are discussed in more detail in the paragraphs for each category.

4.1 Pronouns

The pronoun category includes all the instances where the word *one* occurs as a pronoun in the material except for the pronoun subject because it is analysed separately due to its great occurrence in the data. Since the pronoun subject, namely the generic subject, is analysed separately from other pronoun instances of the word *one*, the pronoun subject is covered in its own section (4.2) and Table 2 represents only the instances of *one* as a pronoun subject, whereas Table 1 presents all other instances of the word *one* as a pronoun in general.

Corpus & words in total	**Instances of *one* as a pronoun**	**N/100,000**	**Percentage of *one* as a pronoun compared to other instances of *one* in the subcorpora**
Finnish: 170,976	113	**66.5**	**15.8%**
Swedish: 280,263	178	**63.6**	**15.2%**
Spanish: 198,131	242	**122.2**	**36.2%**

Table 1. Instances of the word one as a pronoun per 100,000 words and the percentage of one as a pronoun compared to other instances of the word *one* in the subcorpora.

Table 1 above provides information about the word *one* as a pronoun in general in the data. As can be seen from the Table 1, the Finnish and Swedish subcorpora are almost equal when it comes to the number of instances per 100,000 words and the percentages. Nevertheless, Spanish subcorpus has more than twice as many instances of the word *one* as a pronoun compared to both Finnish and Swedish subcorpora as regards the percentages. It is relevant to point out that the percentage of 36.2% in the Spanish subcorpus is more than twice as big compared to both Finnish and Swedish subcorpora with 15.8% and 15.2%. However, when the three subcorpora are compared to each other with the instances of the word *one* per 100,000 words the difference between the numbers is not that big as it is in percentages. The instances per 100,000 words in the Spanish subcorpus is 122.2, the Finnish subcorpus is 66.5 and the Swedish subcorpus is 63.6. Thus, the Spanish subcorpus' number is not more than twice as big compared to Finnish and Swedish subcorpora. This difference between the percentages and the number of instances per 100,000 words is due to the differences in corpus sizes.

The Spanish subcorpus stands out clearly when compared to both Finnish and Swedish subcorpora. Nonetheless, the Finnish and Swedish subcorpora stand out form

Table 1 due to their evenness both in percentages and number of instances per 100,000 words.

4.2 Pronoun subjects

This paragraph presents all the instances where the word *one* occurs as a pronoun subject in the material. Table 2 below demonstrates all the instances of the word *one* as a pronoun subject both in percentages and occurrences per 100,000 words in all three subcorpora.

Corpus & words in total	**Instances of *one* as a pronoun subject**	**N/100,000**	**Percentage of *one* as a pronoun subject compared to other instances of *one* in the subcorpora**
Finnish: 170,976	164	**96.5**	**23%**
Swedish: 280,263	274	**97.8**	**23.4%**
Spanish: 198,131	43	**21.7**	**6.4%**

Table 2. Instances of the word *one* as a pronoun subject per 100,000 words and the percentage of *one* as a pronoun subject compared to other instances of the word *one* in the subcorpora.

As for the word *one* as a pronoun subject, Table 2 above demonstrates that there is a clear difference between the Spanish subcorpus and the Finnish and Swedish subcorpora. More specifically, the Spanish subcorpus includes only a small number of instances of the pronoun subject whereas the Finnish and Swedish subcorpora's instances are almost a quarter of all the instances of the word *one*. The difference between the three subcorpora is significant because the pronoun subject percentage in the Spanish subcorpus is 6.4% whereas the Finnish subcorpus' percentage is 23% and the Swedish subcorpus' is 23.4%.

As regards the pronoun subject category, the Finnish subcorpus and the Swedish subcorpus are basically equal in numbers and percentages which means that L1 speakers of Finnish and Swedish use the pronoun subject *one* similarly in L2 English. What is significant regarding the pronoun subject category and its results is that L1 speakers of Spanish do not utilize much the generic subject in L2 English.

The difference between the pronoun category and the pronoun subject category when it comes only to the Spanish subcorpus is significant. Table 1 shows that 36.2% of the instances of the word *one* in Spanish subcorpus represent other types of use as a pronoun but only 6.4% of the instances of the word *one* in the Spanish subcorpus are pronoun subjects. The difference is also remarkable as for the Spanish subcorpus' instances compared to Finnish and Swedish subcorpora's instances. There are clearly most instances of the word *one* as a pronoun in the Spanish subcorpus but when it has to do with the pronoun subject, the Spanish subcorpus has clearly the least instances of the word *one* as a pronoun subject. There are almost four times more instances of one as a pronoun subject in the Finnish and Swedish subcorpora compared to the Spanish subcorpus whereas the Spanish subcorpus has twice as many instances of *one* as a pronoun in general. The Spanish subcorpus differs greatly from the Finnish and Swedish subcorpora and in addition, Finnish and Swedish subcorpora are basically equal again.

4.3 Determiners

This paragraph presents the numbers of instances and percentages of the word *one* as a determiner in the material. Table 3 below represents the precise numbers for each subcorpus.

Corpus & words in total	**Instances of *one* as a determiner**	**N/100,000**	**Percentage of *one* as a determiner compared to other instances of *one* in the subcorpora**
Finnish: 170,976	152	**89.4**	**21.3%**
Swedish: 280,263	312	**111.4**	**26.6%**
Spanish: 198,131	100	**50.5**	**15%**

Table 3. Instances of the word *one* as a determiner per 100,000 words and the percentage of *one* as a determiner compared to other instances of the word *one* in the subcorpora.

The Spanish subcorpus stands out once again in Table 3 when compared to the Finnish and Swedish subcorpora. However, as for the word *one* as a determiner, the difference is not that distinct as it is in Tables 1 and 2. Nonetheless, the difference is visible because 15% of the Spanish subcorpus' instances of the word *one* represent *one* as a determiner whereas the Finnish subcorpus' percentage is 21.3% and Swedish is 26.6%. Nevertheless, it is clear from Table 3 that the word *one* was used as a determiner the least often in the Spanish subcorpus and the difference to the Swedish subcorpus, in which there were the most instances of *one* as a determiner, was 11.6%.

4.4 Numerals

This paragraph presents the numbers of instances and percentages of the word *one* as a numeral in the material. Table 4 below demonstrates the numbers and differences between the three subcorpora.

Corpus & words in total	**Instances of *one* as a numeral**	**N/100,000**	**Percentage of *one* as a numeral compared to other instances of *one* in the subcorpora**
Finnish: 170,976	53	**31.2**	**7.4%**
Swedish: 280,263	107	**38.2**	**9.1%**
Spanish: 198,131	58	**29.3**	**8.7%**

Table 4. Instances of the word *one* as a numeral per 100,000 words and the percentage of *one* as a numeral compared to other instances of the word *one* in the subcorpora.

The numeral category is significantly different compared to previous categories due to the evenness of numbers and percentages in all three subcorpora as can be seen in Table 4. The numbers and percentages demonstrate that the numeral category for the word *one* is not one of the biggest categories in the analysis because not even one of the three subcorpora includes over 10% instances of *one* as a numeral. The Finnish subcorpus has the lowest percentage with 7.4%, the Spanish subcorpus is next with 8.7% and the Swedish subcorpus has the most instances with 9.1%. It is noteworthy to point out that even though the Finnish subcorpus has the lowest percentage, the Spanish subcorpus has the lowest number of instances with 29.3 instances per 100,000 words whereas the Finnish subcorpus has 31.2 instances per 100,000 words. In addition, the evenness of percentages is visible in Table 4 because the difference between the lowest percentage, which is the Finnish subcorpus' 7.4%, and the highest percentage, which is the Swedish subcorpus', is 9.1%. Therefore, the difference between the lowest and highest percentage is only 1.7%.

One of the reasons why the numeral category's numbers and percentages are so close to each other in every subcorpus is probably that the use of word *one* as a numeral

is so common. Already in the early stages of language learning one of the things that a language learner learns first is the numbers and *one* as a numeral is therefore probably the first context for the word *one* that language learners acquire.

4.5 The *One of* -structure

This paragraph presents numbers and percentages of the *one of* -structure in the material. Table 5 below demonstrates how the category the *one of* -structure was presented in all the three subcorpora.

Corpus & words in total	**Instances of the *one of* -structure**	**N/100,000**	**Percentage of instances of *one of* -structures compared to other instances of *one* in the subcorpora**
Finnish: 170,976	117	**68.8**	**16.4%**
Swedish: 280,263	169	**60.4**	**14.4%**
Spanish: 198,131	167	**84.3**	**25%**

Table 5. Instances of the structure *one of* per 100,000 words and the percentage of the structure compared to other instances of the word *one* in the subcorpora.

The *one of* -structure is most utilized in the Spanish subcorpus with 25%. The percentage is distinctively bigger than the percentages in the Finnish and Swedish subcorpora. The percentage of *one of* -structure in Finnish subcorpus is 16.4% and in the Swedish subcorpus it is 14.4%. Thus, the Spanish subcorpus' percentage is almost 10% bigger than it is in the other two subcorpora.

The *one of* -structure is a basic structure or a pattern which is frequently used in English. The structure is not considered being a difficult structure but rather due to its great occurrence in English language it is probably one of the most used structures and

therefore, language learners learn it early on in their learning process. Moreover, there are similar structures in Swedish, *en av*, and in Spanish, *uno de*. However, these equivalent structures are not used the same way as the English *one of* -structure. In Finnish, the word *yksi* is the equivalent to the English *one of* -structure but without a preposition.

4.6 The genitival expression *One's*

This paragraph covers the category *one's* genitive in the material. All three subcorpora and their numbers and percentages can be seen in Table 6 below.

Corpus & words in total	**Instances of *one's* genitive**	**N/100,000**	**Percentage of genitive *one's* compared to other instances of *one* in the subcorpora**
Finnish: 170,976	67	**39.4**	**9.4%**
Swedish: 280,263	60	**21.4**	**5.1%**
Spanish: 198,131	18	**9.1**	**2.7%**

Table 6. Instances of genitive *one's* per 100,000 words and the percentage of the genitival expression compared to other instances of the word *one* in the subcorpora.

In the category *one's* the Finnish subcorpus has the biggest number of instances, which is 39.4 per 100,000 words, and the percentage is 9.4% in the material. After the Finnish subcorpus comes the Swedish subcorpus with 21.4 instances per 100,000 words and with 5.1%. The corresponding numbers in the Spanish subcorpus are relatively small compared to those in the Finnish and Swedish subcorpora because the instances of *one's* in the Spanish subcorpus is 9.1 instances per 100,000 words and the percentage is only 2.7%.

Even though none of the three subcorpora has more than 10% of *one's* in their data, there is still a great difference between the three subcorpora. The Swedish subcorpus' percentage is almost twice as big as Spanish subcorpus' and compared to the Finnish subcorpus, it is almost twice as small.

4.7 Idiomatic Expressions

In this paragraph, the numbers and percentages of *one* in idiomatic expressions in the three subcorpora are presented. Examples of instances that are counted as idiomatic expressions are illustrated with examples in paragraph *3.3.6 Idiomatic Expression*. Table 7 below demonstrates the precise numbers and percentages of idiomatic expressions in the material.

Corpus & words in total	**Instances of idiomatic expressions**	**N/100,000**	**Percentage of idiomatic expressions compared to other instances of *one* in the subcorpora**
Finnish: 170,976	7	**4.1**	**1%**
Swedish: 280,263	8	**2.9**	**0.7%**
Spanish: 198,131	7	**3.5**	**1%**

Table 7. Instances of idiomatic expressions per 100,000 words and the percentage of the idiomatic expressions compared to other instances of the word *one* in the subcorpora.

Idiomatic expressions are considered being one category because these expressions do not fit in any other category even though there are not that many instances of idiomatic expressions in the material. Basically, all the three subcorpora are even when it comes to numbers and percentages. There is not much difference between the Finnish and Spanish subcorpora's 1% compared to Swedish subcorpus' 0.7%. Therefore, one of the

conclusions that can be drawn from Table 7 and the numbers it illustrates is that L1 speakers of Finnish, Swedish, and Spanish use almost the same number of idiomatic expressions in their L2 English.

4.8 Errors

This paragraph presents all the errors that were found in the material in numbers and percentages. Examples of such instances that are counted as errors in the present study are demonstrated and explained in the paragraph *3.3.7 Errors.*

Corpus & words in total	**Instances of errors**	**N/100,000**	**Percentage of errors compared to other instances of *one* in the subcorpora**
Finnish: 170,976	3	**1.8**	**0.4%**
Swedish: 280,263	7	**2.5**	**0.6%**
Spanish: 198,131	12	**6.1**	**1.8%**

Table 8. Instances of errors per 100,000 words and the percentage of the errors compared to other instances of the word *one* in the subcorpora.

Table 8 above demonstrates both in numbers and also in percentages how many instances of errors there are in the material. Based on the analysis, the Spanish subcorpus seems to have most errors with 1.8% compared to the Swedish subcorpus with 0.6% and the Finnish subcorpus with 0.4%. Therefore, it can be concluded that even though there are some errors regarding the word *one* and its use in the data, the errors do not have a significant role in the data due to the small number and percentages that the errors represent in the categories and thus, in the analysis. Nevertheless, it is necessary to have the category for errors so that every instance of the word *one* can be categorised into some category instead of simply ignoring some of the instances of the word *one.*

However, it is noteworthy to point out that the errors that were found in the material had to do only with the instances and uses of the word *one*. Thus, it is probable that there are more errors in the data regarding other things than the word *one*, but they are not considered being relevant in the present study

4.9 Irrelevant hits

In this paragraph, the instances of the word *one* as irrelevant regarding the present study and its analysis are presented. Table 9 illustrates the numbers and percentages of *one* as irrelevant compared to other categories. Examples of irrelevant instances of the word *one* in the material are demonstrated in the paragraph *3.3.8 Irrelevant instances in the data*.

Corpus & words in total	**Instances of irrelevant instances of *one***	**N/100,000**	**Percentage of irrelevant instances compared to other instances of *one* in the subcorpora**
Finnish: 170,976	34	**20**	**4.8%**
Swedish: 280,263	50	**17.9**	**4.3%**
Spanish: 198,131	16	**8.1**	**2.4%**

Table 9. Instances of irrelevant instances of *one* per 100,000 words and the percentage of the irrelevant instances compared to other instances of the word *one* in the subcorpora.

Table 9 demonstrates all the cases of the word *one* that were considered being irrelevant in the data. The Finnish subcorpus has the most irrelevant cases of the word *one* with 4.8%. Next comes Swedish subcorpus with 4.3% and the smallest number and percentage of the irrelevant cases of the word *one* has the Spanish subcorpus with 2.4%. It is noteworthy to point out that the Finnish subcorpus has twice as many instances of the word *one* as irrelevant than the Spanish subcorpus.

The difference is remarkable especially when Table 8 is also analysed. Table 8 shows that the Spanish subcorpus has over four times more instances of the word *one* as error than the Finnish subcorpus but the situation is the opposite regarding Table 9 and the irrelevant category for the word *one*. Thus, it is evident that there are differences between the three subcorpora as regards the different categories for the word *one* and their occurrence in the analysed material. These differences and the possible reasons behind them are discussed in more detail in Chapter 5.

5 DISCUSSION

This chapter focuses on the differences of using the word *one* in different contexts by L1 speakers of Finnish, Swedish, and Spanish in L2 English. These differences between the Finnish, Swedish and Spanish subcorpora are discussed in more detail and possible reasons for these differences are provided in this chapter. Based on the analysis of the material, it becomes clear that there are differences, some rather significant as well, between the three subcorpora regarding the different usages of the word *one* and its occurrence in different categories in the data. As noted already in the Introduction, the current study aims to provide answers for the following research questions:

1) How are the uses of the word *one* different when it is used by L2 English speakers whose L1 is either Finnish, Swedish or Spanish?
2) How does L1 interfere with L2 English when it comes to the usage of the word *one*?

To begin with, there were significant differences in some of the categories between the Finnish, Swedish, and Spanish subcorpora as regards the instances of the word *one*. For example, the results of the pronoun category and the pronoun subject category revealed that the Finnish and the Swedish subcorpora had almost precisely the same percentages in both categories. The instances of the word *one* in Finnish subcorpus as a pronoun was 15.8% and as a pronoun subject it was 23% whereas in the Swedish subcorpus the number of *one* as a pronoun was 15.2% and as a pronoun subject it was 23.4%. Nevertheless, the Spanish subcorpus stands out clearly in both categories because in the pronoun category the Spanish subcorpus has more than twice as many instances of *one* as a pronoun compared to Finnish and Swedish subcorpora with 36.2% but as for the pronoun subject

category, the Spanish subcorpus includes almost four times less instances of *one* as a subject pronoun with 6.4%. All the instances of the word *one* are demonstrated in Table 10 below.

	ICLE-Finnish	**ICLE-Swedish**	**ICLE-Spanish**
Pronouns	15.8%	15.2%	36.2%
Pronoun subjects	23.0%	23.4%	6.4%
Determiners	21.3%	26.6%	15.0%
Numerals	7.4%	9.1%	8.7%
The *One of* -structure	16.4%	14.4%	25.0%
The genitival expression *One's*	9.4%	5.1%	2.7%
Idiomatic expressions	1.0%	0.7%	1.0%
Errors	0.4%	0.6%	1.8%
Irrelevant hits	4.8%	4.3%	2.4%
Total:	100%	100%	100%

Table 10. All the instances of the word *one* in the three subcorpora.

It is clearly the case that both Finnish and Swedish people are more familiar with the subject pronoun *one* in English than the Spanish people. Krashen (1981, 67) has noted that L1 influence on L2 may be a sign of low acquisition. Since the word *one* occurred a lot, especially when compared to other subcorpora, as a pronoun in the Spanish subcorpus, it could be argued that the Spanish speakers' L1 affected their L2 English as regards the use of *one* as a pronoun in general. In addition, Ellis (2000, 51) has noted that the overuse of some structure in L2 might indicate L1 transfer to L2. Since the L1 speakers of Spanish used *one* as a pronoun a lot, it could also be a sign of overuse of *one* as a pronoun which could be caused by their L1 Spanish. Respectively, the low percentage of *one* as a pronoun subject in the Spanish subcorpus, only 6.4% compared to 23% and 23.4% in Finnish and Swedish subcorpora, may indicate, as Ellis (2000, 51) has noted, that L1 transfer may lead to L2 learners avoiding some structure in L2 because it does not exist in their L1.

As Spanish does not have a similar structure as English does with the generic pronoun subject *one*, this difference could be due to avoidance in L2 English but nevertheless, it could also be caused by low acquisition of L2. In addition, the Finnish language does not have a similar structure either as the pronoun subject *one* in English, but Swedish does because the equivalent structure is in Swedish *man* -structure which is an active structure with a generic subject. For instance, the sentence *Man borde inte äta i bilen* could be translated into English for example with a sentence *One should not eat in the car*. Nevertheless, since L1 speakers of Finnish used *one* as a subject pronoun quite a lot, in fact nearly as much as L1 speakers of Swedish who have a similar structure/word in their L1, it suggests that Finnish people have a good knowledge of English language and that they are competent in L2 English. Moreover, Axelsson and Hahn (2001) suggest in their study that L2 learners of a certain language might tend to overuse a structure, like the progressive in Axelsson and Hahn's (2001) study, when their L1 does not recognise a similar structure. Nevertheless, this conclusion also means that the speakers have a good knowledge of the L2 since they know how and when to use a structure that does not exist in their L1.

When it comes to the determiner category for the word *one,* there was more variation among the three subcorpora. The Swedish subcorpus had the biggest percentage of instances with 26.6% and there was a difference of more than 5% to the Finnish subcorpus which had 21.3%. Nevertheless, the difference between the Swedish and the Spanish subcorpus was more than 10% because the Spanish subcorpus had 15% of instances of *one* as a determiner. As can be seen in the section 3.3.2 (Determiners), the instances where the word *one* functions as a determiner includes cases where *one* functions simply as a determiner and is not part of any larger or more complex structure,

but the instances also include structures/patterns like *on the one hand* and *one way or the other* where *one* is a part of a structure, but it still functions as a determiner in the context. These structures/patterns where *one* is considered to be a determiner are however more advanced structures and using them requires a good knowledge on English language. In other words, people who use these kinds of structures/patterns have a high acquisition of English language and how it functions.

L1 speakers of Swedish had the biggest percentage of *one* as a determiner which can be interpreted in a way that they are the most used to using these kinds of structures/patterns in their L2 English. L1 speakers of Finnish use *one* as a determiner less than L1 speakers of Swedish but still they use it more than L1 speakers of Spanish in L2 English. Therefore, it seems that the L1 speakers of Spanish use *one* as a determiner the least and the difference to L1 speakers of Swedish is quite significant. These differences between L1 speakers of Swedish, Finnish and Spanish may be due to the acquisition levels of English and how competent people are regarding their L2 English.

The numeral category was quite even regarding the percentages among all three subcorpora because the Swedish subcorpus' percentage was 9.1%, the Spanish subcorpus' was 8.7% and the Finnish subcorpus' was 7.4%. Since the difference between the Swedish and Finnish subcorpora was only 1.7%, it can be concluded that there were not significant differences in the percentages between the three subcorpora. Therefore, these results suggest that L1 speakers of Finnish, Swedish, and Spanish use *one* as a numeral in similar ways and to the same extent in their L2 English. A probable reason for this result is that numerals are used in similar ways in all languages worldwide. Therefore, in this case it could be concluded that the use of numerals is an indication of positive transfer as Ellis (2000, 51) has noted, because the L1 and L2 are similar when it comes

to the usage of numerals and thus, the L1 helps in learning the principles of how numerals are also used in the L2.

The *one of* -structure category implies that L1 speakers of Spanish use the structure *one of* significantly more than L1 speakers of Finnish and Swedish. 25% of instances of the word *one* were instances of the structure *one of* in the Spanish subcorpus whereas the Finnish subcorpus had 16.4% of the instances and the Swedish subcorpus had 14.4%. Thus, the results indicate that L1 speakers of Spanish use the structure *one of* considerably more whereas L1 speakers of Finnish and Swedish use the same structure less and more similarly in English. One possible reason for this result is that the structure *one of* is widely used and it is a basic structure which does not require a great knowledge on English language in order that it can be used and understood. As Ellis (2000, 51) noted, the overuse of some structure might be an indication of low acquisition because those who do not have a comprehensive knowledge on e.g., English language may tend to use considerably more some specific structure that they are familiar with instead of using several different structures that would require a comprehensive understanding of the language and how it functions.

The results of the category *one's* demonstrate that the Finnish subcorpus has the most instances with 9.4% and it is almost two times more than the Swedish subcorpus' percentage which is 5.1%. Moreover, the Swedish subcorpus' percentage 5.1% is almost two times bigger than the Spanish subcorpus' percentage which is 2.7%. All in all, the difference between the percentages of the Finnish and Spanish subcorpora is a great one because the Finnish percentage 9.4% is three and a half times bigger than the Spanish percentage 2.7%. Thus, the L1 speakers of Finnish use the genitive *one's* significantly more than L1 speakers of Spanish. In addition, the difference between the L1 speakers of

Swedish also differs remarkably from both L1 speakers of Finnish, and also L1 speakers of Spanish.

One possible reason for these results might be the fact that Finnish language does not have prepositions, but instead Finnish language uses case suffixes to express the same things which other languages indicate with prepositions. Correspondingly, both Swedish and Spanish languages utilize prepositions. Therefore, it might be a natural result that L1 speakers of Finnish would tend to avoid structures which require using prepositions and instead they would prefer using structures like *one's* which do not include prepositions and would be thus easier to use. Respectively, it might be more familiar and possibly also easier for L1 speakers of Swedish and Spanish to use structures which include prepositions and, thus they would tend to avoid or just simply use the structure *one's* less in English because a similar structure is not that frequently used in their L1.

The category of idiomatic expressions is one of the smallest categories in the present study regarding the numbers and percentages of instances of the word *one* as a part of an idiomatic expression in the material. Basically, all the three subcorpora have the same number of idiomatic expressions in their material because the Finnish and Spanish subcorpora's instances of the word *one* is 1% in their data. The Swedish subcorpus' percentage is 0.7% which is not that much either. The difference between the three subcorpora is only 0.3% which basically means that there is no difference between the Finnish, Swedish, and Spanish subcorpora regarding the instances of idiomatic expressions in the material. The use of idiomatic expressions is a sign of high acquisition and using them requires a great knowledge on English language because in order to use

them one has to understand what their meaning actually is and in which context they can be used.

Even though the percentages were not big in any subcorpus, the Swedish subcorpus still stands out a little with 0.7% especially because the number of words was in the Swedish subcorpus considerably bigger when compared to the Finnish and Spanish subcorpora. Therefore, 0.7% probably is equivalent to the Finnish and Spanish subcorpora's 1%. Once again, it seems that high acquisition of English language explains the relatively frequent use of idiomatic expressions. Nonetheless, as regards the use of idiomatic expressions with *one* in the material, there are not significant differences between the three subcorpora. Therefore, it can be concluded that not one of the groups used more idiomatic expressions than the other groups. As already noted, Ellis (2000, 51) argues that avoidance might be an indication of L1 transfer. Since idiomatic expressions cannot usually be translated straight from one language to another, they are not that frequently used by L2 speakers of English. Since idiomatic expressions were not used frequently in the material, it suggests that none of the L2 English speakers are competent enough with using English idiomatic expressions frequently.

The error category is also one of the smallest categories in the present study as for the percentages of its instances in the material. However, regarding the instances of the word *one* as an error there are differences between the subcorpora in the material. The most errors were found in the Spanish subcorpus with 1.8% of all the instances of the word *one* in the material. The Spanish subcorpus' percentage 1.8% differs greatly from the Swedish subcorpus' percentage which is 0.6% and from the Finnish subcorpus' percentage which is 0.4%. Once again, the Finnish and Swedish subcorpora are almost even in the percentages and the difference to the Spanish subcorpus is significant.

When it comes to the error category, it especially reveals that L1 speakers of Spanish do not have as good knowledge of English as L1 speakers of Swedish and Finnish do. Thus, the result suggests that L1 speakers of Spanish have lower acquisition of L2 English compared to Finnish and Swedish L2 speakers of English. Nonetheless, it is also worth mentioning that both L1 speakers of Finnish and Swedish seem to have a good knowledge on English language because there is not that great a difference between the Finnish and Swedish subcorpora and their results as regards the error category. Westergaard's (2003) study revealed that unlearning the V2 word order was quite challenging, and it happened gradually. Even though there was a lot L1 transfer to L2 in the beginning, progress still happened during the observation period and the results suggest that SVO word order is most likely learned eventually but it just takes time. The number of errors might indicate how long and how much the L2 has been used and thus the less there are errors the longer and more the L2 in question has been used. Moreover, Erarslan and Hol (2014) noted in their study that the errors that were made in L2 were traceable to the participants' L1. Thus, it is logical to suggest that errors are a sign of L1 transfer.

As for the numbers of irrelevant hits, the Finnish and Swedish subcorpora are close to each other's percentages because the Finnish subcorpus' percentage is 4.8% and the Swedish subcorpus' is 4.3% regarding the irrelevant instances of the word *one* in the material. The Spanish subcorpus' percentage is 2.4% which is two times smaller compared to the Finnish subcorpus' percentage. This is a surprising result because in the error category Spanish subcorpus had clearly the most instances but as regards the irrelevant cases of the word *one* the Spanish subcorpus has two times less instances compared to the Finnish and Swedish subcorpora.

One reason for this result might be the fact that AntConc software counted all the instances of the word *one* even if it was a part of a compound word like e.g., *one-way*. Nonetheless, there were also instances of compound words which included the word *one* but there should have been a line in the middle of the compound word. However, in some cases the line had been left out and in that case the instance of the word *one* was counted as an error and not as irrelevant. Therefore, some of the examples of the word *one* where it was supposed to be a part of a compound were considered as errors even though they would have been considered as irrelevant cases if they had been spelled correctly. Thus, the number of errors in the Spanish subcorpus includes examples which would have been counted as irrelevant cases if they had not been misspelled. Therefore, in reality there would have been more irrelevant cases, like compound words, and less errors in the Spanish subcorpus. Nonetheless, it is not that straightforward to draw the line between instances of errors and irrelevant instances but the categorisation and the grounds for it had to be done in some way. If the grounds for different categories had been different, probably the results would have been to some degree different as well.

To sum up, the answer for the first research question is that L1 speakers of Finnish and Swedish use the word *one* in L2 English rather similarly and to almost same extent whereas the L1 speakers of Spanish stand out clearly in the study. As for the second research question about how L1 affects L2, the results suggest that the overuse of some structure or pattern is an indicator of L1 transfer as is also avoiding using some structure or pattern in L2. In addition, errors might suggest that L1 has interfered with L2 which has led to errors in L2. All in all, the results presented in this chapter indicate that the Spanish subcorpus stands out in the study and the Finnish and Swedish subcorpora are relatively similar in many categories even though not in every category. These differences

between the subcorpora are likely due to the knowledge of English and how competent the L1 speakers of Finnish, Swedish, and Spanish are in using L2 English. The results suggest that L1 speakers of Spanish have lower acquisition of L2 English compared to L1 speakers of Finnish and Swedish who seem to have a quite good knowledge on L2 English based on the present study's results. Nevertheless, as Housen (2002) noted in their study of learning verb forms and functions, individual variation exists even when L1 and other factors in language background are taken into consideration.

When it comes to the possible reasons for the present study's results, there could be a number of different possible reasons. In addition to the possible reasons already mentioned, other reasons might be the fact that TV-shows are dubbed in Spain which means that Spanish people do not hear English among other languages and therefore, they are not exposed to language input. In Finland and in Sweden TV-shows have subtitles which means that Finnish and Swedish people hear authentic English and thus, they are influenced by it whether they want to or not. Another fact that might explain the similar results between the Finnish and Swedish subcorpora is that Finland and Sweden have rather similar educational systems. Moreover, Finnish and Swedish people are well educated in general. These factors might help to understand the results of the present study, but it is still not that straightforward to say which factors are the crucial ones when it comes to explaining the results of the study.

6 CONCLUSIONS

This thesis has attempted to explain how L1 affects L2 and in this study the focus has been on the English word *one*. This corpus study has analysed the use of the word *one* in L2 English by L1 speakers of Finnish, Swedish, and Spanish. In this thesis, the corpus-driven approach has been utilized and the material consists of three subcorpora of the ICLE, in other words the International Corpus of Learner English. In order to analyse the material, different categories were created for the word *one* so that the comparisons between the Finnish, Swedish, and Spanish subcorpora could be done easily. All the instances of the word *one* were analysed and categorised. Moreover, the number of instances and the percentages of the word *one* in all the categories were counted, and the tables in Chapter 4 represent the occurrence of the word *one* in different categories in all the three subcorpora. As expected, some of the categories were quite large and respectively, some of the categories had only a few instances and therefore, they were relatively small. Nevertheless, the size of the categories reveals in which contexts the word *one* is used the most and in which contexts it is not that frequently utilized.

As already noted earlier in the Chapters 4 and 5, the Spanish subcorpus stands out in the results. Nonetheless, the Finnish and Swedish subcorpora were surprisingly similar although not in every category. Nevertheless, it becomes clear from the results that based on this study's material, the L1 speakers of Spanish are not that competent in using L2 English and understanding how the language functions. Respectively, L1 speakers of Finnish and Swedish seem to understand quite well how English language functions and thus, they seem to be competent in L2 English. Nonetheless, there are some categories where there were basically no differences between the three subcorpora like

the category for idiomatic expression and the category for numerals. This result might be explained due to the fact that numerals function basically the same way in any language and that idiomatic expressions are not that frequently used in general. However, it might also be the case that none of the studied groups were competent enough to use idiomatic expressions in L2 English.

As regards the question about how L1 possibly affects L2, the present study's results suggest that L1 does affect L2 and based on the analysis the effects can be seen e.g., in errors and in structures that both L1 and L2 have. Moreover, as for the structures that L1 and L2 do not share, the cause might be either avoiding usage of the structure in question or overusing it. For instance, L1 speakers of Spanish used significantly *one* as a pronoun in general but they clearly avoided using *one* as a subject pronoun.

A follow-up study could be conducted on a similar topic, but the study's L2 English material could be compared to a data set that would include material which was written by L1 English speakers. Thus, the differences between L1 speakers of English and L2 speakers of English could be compared with each other.

REFERENCES

Corpus

ICLE = *International Corpus of Learner English*. Version 2.

Works Cited

Axelsson, Margareta Westergren, and Angela Hahn. 2001. *The use of the progressive in Swedish and German advanced learner English: A corpus-based study.* ICAME journal 25 (2001): 5-30.

Baker, Paul, et al. 2006. *A Glossary of Corpus Linguistics*. Edinburgh University Press.

Biber, Douglas., et al. 1998. *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge University Press.

Brown, Cynthia A. 1998. *The role of the L1 grammar in the L2 acquisition of segmental structure*. Second Language Research 14.2 (1998): 136-193.

Collins Dictionary. n.d. Accessed February 22, 2022. https://www.collinsdictionary.com/dictionary/english/one.

Cook, Vivian. 2016. *Second Language Learning and Language Teaching*. Fifth Edition. Taylor and Francis.

Ellis, Rod. 2000. *Second Language Acquisition*. 4. impr., Oxford University Press.

Erarslan, Ali, and Devrim Hol. 2014. *Language interference on English: Transfer on the vocabulary, tense and preposition use of freshmen Turkish EFL learners.* ELTA Journal 2.2 (2014): 4-22.

Granger, Sylviane, Joseph Hung, and Stephanie Petch-Tyson, eds. 2002. *Computer learner corpora, second language acquisition, and foreign language teaching*. Vol. 6. John Benjamins Publishing.

Granger, Sylviane, et al., eds. 2009. *International Corpus of Learner English*. Vol. 2. Louvain-la-Neuve: Presses universitaires de Louvain.

Hawkins, Roger. 2019. *How Second Languages Are Learned. An Introduction.* Cambridge: Cambridge University Press.

Hornby, A. S. 1975. *Guide to Patterns and Usage in English*. 2. ed., Oxford University Press.

Housen, Alex. 2002. *A corpus-based study of the L2-acquisition of the English verb system.* In Granger et al. 2002.

Hunston, Susan, and Gill Francis. 2000. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Benjamins.

Jordens, Peter, and Josine A. Lalleman. 1996. *Investigating Second Language Acquisition*. Reprint 2010, Mouton de Gruyter, https://doi.org/10.1515/9783110812831.

Krashen, Stephen D. 1981. *Second Language Acquisition and Second Language Learning*. Pergamon Press.

Königs, Frank G. in Fäcke, Christiane, eds. 2014. *Manual of Language Acquisition*. Berlin, [Germany: De Gruyter, Web.]

Lado, Robert. 1957. *Linguistics across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor: University of Michigan Press.

Levey, Sandra. 2019. *Introduction to Language Developmen*t. Second edition., Plural Publishing, Inc.

MacMillan Dictionary. n.d. Accessed February 22, 2022. https://www.macmillandictionary.com/dictionary/british/one_1.

McEnery, Tony, and Andrew, Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice.* Cambridge University Press.

Moon, Rosamund. 1998. *Fixed Expressions and Idioms in English: a Corpus-Based Approach*. Clarendon Press.

Ringbom, Håkan. 2007. *Cross-Linguistic Similarity in Foreign Language Learning*. Multilingual Matters.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford University Press.

Schwartz, Bonnie D., and Rex A. Sprouse. 1996. *L2 cognitive states and the full transfer/full access model*. Second language research. 12, 40-72.

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. J. Benjamins.

Westergaard, Marit. 2003. *Unlearning V2: Transfer, markedness, and the importance of input cues in the acquisition of word order in English by Norwegian children*. EUROSLA yearbook 3.1 (2003): 77-101.