MASTER'S THESIS

Article use in the writing of young Norwegian EFL learners: a corpus-based study

Sofie Larsen

Candidate number: 124

Master´s in Education with English Didactics
Department of Language, Literature, Mathematics and Interpreting
Supervisor: Kristian Andersen Rusten
Submission Date: May 14th, 2019

Sammendrag

Formålet med denne masteroppgaven er å undersøke hvordan unge norske elever bruker artikler (*a, an, the, zero article)* i engelsk skriftlig. Oppgaven er en korpus-basert studie som bruker korpuset *Corpus of Young Learner Language* (CORYL). Korpuset er utviklet ved Universitetet i Bergen og består av elevtekster hentet fra Nasjonale prøver i Engelsk fra 2005 og 2011. Elevtekstene er hentet fra elever på 7.- og 10.trinn. Oppgaven tar for seg tre ulike forskningsspørsmål: 1) Hvordan bruker unge norske elever artikler når de skriver engelsk som et fremmedspråk? 2) Hvilke feil gjør elevene angående bruk av artikler, og hvor mange feil gjør de sammenlignet med tilfellene der de bruker artiklene rett? 3) Hvordan utvikler bruken av artikler seg i skriving fra 7.- til 10.klasse?

Studien belyser et relevant teorigrunnlag, blant annet om det grammatiske fenomenet artikler, nasjonale prøver og korpuslingvistiske metoder. Teorien inkluderer en grundig gjennomgang av korrekt bruk av de engelske artiklene, samt en sammenligning av artikkelbruk på norsk og engelsk. Det vil også bli presentert mulige utfordringer med artikkelbruk for elever som lærer engelsk som et fremmedspråk.

Dette er en kvantitativ studie, der datainnsamlingen består av kvantitative data hentet fra CORYL korpuset. Denne innsamlingen er basert på korpuslingvistiske metoder. Elementer fra analysetypen *error analysis* blir brukt for å undersøke hvilke typer feil elevene gjør angående artikkelbruk i engelsk skriftlig; med fokus på overforbruk, underforbruk og feil bruk av artiklene. En tverrsnittstudie er gjennomført for å undersøke utviklingen i bruken av artikler fra 7.klasse til 10.klasse. Korpuset sine egne søkefunksjoner muliggjør å hente ut data om feil bruk av artiklene (a|an|the|zero), samt korrekt bruk av artiklene (a|an|the).

Funnene i studien viser at forskjellen mellom hvor mange feil elevene i 7.klasse versus 10.klasse gjør angående bruken av artikler generelt og bruken av bestemt artikkel ikke er statistisk signifikant. Funnene som omhandler ubestemt artikkel derimot, viste seg å være statistisk signifikant. Feil bruk av artikkel ble også sett i sammenheng med sjangertype. Funnene viste at det kun var statistisk signifikant forskjell mellom sjangeren leserinnlegg/essay, og leserinnlegg/personlig brev.

Acknowledgements

I would like to express my gratitude to my supervisor, Kristian Andersen Rusten, for his useful remarks and comments throughout the process of working with this master thesis. I would also like to thank him for bringing a lot of humour and good spirit into the sessions of supervision. This has made all the difference.

Also, I would like to thank Kari Telstad Sundet for giving me an introduction in how to use the CORYL corpus, as well as being available for further questions. This has been of great help.

Lastly, I would like to express my gratitude towards my partner, friends and family for supporting me and cheering me on. This has been a great motivation for me, and I will be forever grateful.

Sofie Larsen,

Bergen, May 2019

List of contents

List of tables

List of figures

List of abbreviations

EFL: English as a Foreign Language

SLA: Second Language Acquisition

CORYL: Corpus of Young Learner Language

L1: First language

L2: Second language

EA: Error Analysis

CEFR: Common European Framework of Reference

CA: Contrastive Analysis

LCR: Learner Corpus Research

UDIR: The Norwegian Directorate for Education and Training

Chapter 1: Introduction

1.1 Background and aim

This thesis sets out to investigate the use of articles in the writing of young Norwegian learners of English as a Foreign Language (EFL) through a corpus-based study. Articles are frequently used words in English, and many EFL learners can be expected to have difficulties with correct use of articles in speech and in writing due to a number of differences in the article systems of English and Norwegian. The purpose of doing this research, then, is to acquire a better understanding of how young Norwegian EFL learners use articles in their writing, by investigating underuse, overuse and incorrect use of articles, evidence of which can be found through an error analysis. In order to present a complete picture of article use among the Norwegian learners of English under investigation, errors in terms of underuse, overuse and incorrect use will be contrasted with the learners correct use of articles. By knowing how a substantial number of EFL learners in fact use articles and what errors they make, awareness can be raised about the challenges the learners face regarding this and the insights derived from this may potentially be applied to the teaching of English.

This thesis uses corpus linguistic methods. In order to acquire knowledge about learner language and learners' use of different grammatical features, it is – of course – necessary to obtain texts or oral speech produced by learners. Corpora and corpus linguistics are excellent tools for such research, as corpus linguistics "[…] offers a unique perspective because of its use of quantitative analyses, which allow researchers to investigate patterns of language use that are otherwise impossible to ascertain" (Duff, Bieber & Conrad, 2001, p. 332). Furthermore, corpus linguistics is a relatively modern research field that is currently growing, and a wealth of research has been carried out on learner language using corpus methods. One noticeable feature of previous corpus-based studies on learner language, however, is that they tend to focus on older or more proficient language learners – typically undergraduate University students. Consequently, there is a lack of focus on the developing language of the young language learner in previous research. Additionally, there is as of yet little research on Norwegian EFL learners within this field. Moreover, while there is previous research on the use of articles within the field of corpus linguistics, it does not present itself as a main topic. My aim in focusing on young Norwegian EFL learners and their written language is consequently to acquire knowledge about a previously neglected group of learners in similar research. This thesis will therefore attempt to fill these lacunae in this research field, at least partially.

1.2 Investigating the use of articles

Articles belong to the word class *determiner*, which in English are words or phrases "[…] that come at the beginning of a noun phrase and signal whether the information is new or familiar, or which tell us something about quantity" (Parrott, 2010, p. 25). There are two types of articles; definite and indefinite. The definite article *the* is used "with all types of noun, except proper nouns, in the singular as well as the plural" (Bækken, 2006, p. 103). *The* is also used to indicate that the listener or reader is familiar "[…] with the thing referred to" (Bækken, 2006, p. 102). The indefinite articles *a* and *an* are used to indicate that the reader or listener is being introduced to something new or unknown, and *a* and *an* is only used in the singular; even though they are "[…] typically used with countable nouns" (Bækken, 2006, p. 110). Additionally, the article is left out when one is referring to something in general, with both plural and uncountable nouns. The omitted, or *zero*, article can be exemplified as follows:

(1.1)   I like cars (*plural)*

(1.2)   She drinks tea (*uncountable)*

Articles are frequently used words in English and Master (2002) claims that this is "[…] making continuous conscious rule application difficult" for learners "over an extended stretch of discourse" (p. 332). This feature is also stressed in textbooks for English language teachers. The use of articles is decided by many different rules. Some of these rules contradict each other, and often cause learners to make mistakes. In Parrot's (2010) *Grammar for English Language Teachers,* it is stated that making mistakes regarding article use "[…] rarely leads to serious problems of communication" (p. 32). However, Parrot stresses, consistent article use errors will cause the reader or listener to have a more difficult time understanding (2010, p. 32). Bækken (2006) claims that using articles correctly is difficult for Norwegians, and that " […] this particular field of grammar is the source of numerous mistakes" (p. 99). Furthermore, interference from Norwegian is, according to Bækken (2006, p. 99), the reason why article errors occur in the linguistic output of this group of learners. The typical mistakes among Norwegians are "overuse" of the definite article, as well as often leaving out the indefinite article (Bækken, 2006, p. 99). Furthermore, Master (2002) underlines that even "[…] the most advanced non-native speaker of English as a second language […]" makes mistakes concerning the use of articles, even though he or she might be extremely proficient in other elements of the language (p. 1).

This thesis will focus on the definite article, the indefinite article, and the zero article. The reason for including all three types of article is that the corpus in question is rather small as will be explained in the immediately following section 1.3. By focusing on all the articles the aim is to acquire enough information to present a fuller description than what would be possible if the focus had been only on the indefinite article, to exemplify. Additionally, it is interesting to have a "larger" scope going into a research project like this, as features might appear that would have been overlooked had the scope of the research been narrower.

1.3 Corpus linguistics

As noted in section 1.1, this MA thesis is a corpus-based study. According to Sylviane Granger (2002) corpus linguistics can be "… defined as linguistic methodology which is founded on the use of electronic collections of naturally occurring texts, viz. corpora" (p. 1). Corpora are made up of texts and they vary in size, from several millions of words to those which consist of e.g. 100 000 words. Using corpora of a substantial size is of course beneficial as the data basis would then include more texts that most likely represent a more diverse sample. A smaller corpus may be more limited, but it can nevertheless be of great use, depending on what one is investigating. The corpus that will be used in this thesis is the CORpus of Young Learner Language (CORYL). CORYL consists of EFL learner texts collected from the National tests in English from grades 7, 10 and 11, with texts largely from 2005 and some from 2011. It consists of 129 421 words, which obviously makes it a rather small corpus [After conducting the data collection tokens have been added to the corpus. As of 04.05.19 it consists of 191 568 tokens – S.L].

Corpus linguistic has many different aims and purposes and is suitable for very many types of linguistic investigation. For example, corpus linguistics can be used to investigate both native and learner language in different genres, it can be used to investigate the native language in newspapers, and it can be used to investigate language produced during earlier historical stages. This means that corpus linguistic is not a methodology specifically for EFL, but a methodology that can be relevant in studies of all types of languages. This thesis, however, uses corpus linguistic methods to describe a certain grammatical phenomenon in learner language. Such descriptions can be useful for many purposes in terms of research on how foreign languages are acquired and the results acquired through studies on corpus linguistics can also be applied when trying to improve how foreign languages are taught (Granger, 2002, p. 1).

Corpus studies usually aim to track the frequency of a grammatical phenomenon. Granger (2002) explains how "[f]requency is an aspect of language of which we have very little intuitive awareness but one that plays a major part in many linguistic applications which require a knowledge not only of what is possible in language but what is likely to occur" (p. 2). Before corpus linguistics became a renowned research method, "[…] empirical study of grammar had to rely primarily upon qualitative analysis" (McEnery & Wilson, 2001, p. 110). This meant that the research could give "[…] detailed descriptions of grammar but was largely unable to go beyond subjective judgements of frequency or rarity" (McEnery & Wilson, 2001, p. 110). Therefore, as corpus studies have received a more renowned place in the field of grammar research, it has provided another dimension by including detailed quantitative data.

Granger (2012) points to the fact that learner corpus research has historically proven to be "[…] stronger on description than interpretation" (p. 21). Furthermore, many previous studies have "[…] focused on varieties of interlanguage that were badly in need of description, namely the upper intermediate and advanced stages of proficiency" (Granger, 2012, p.21). However, when searching for literature that investigates young learner language, specifically the language of young Norwegian learners of English as a foreign language, the results are few. This points to the fact that the language of this group of learners is also 'badly in need of description'.

1.4 Research questions

As stated in section 1.1, the aim of this thesis is to investigate how young Norwegian EFL learners use articles in their writing. Specifically, this is a descriptive study which seeks to analyse and discuss how learners in 7th and 10th grade use articles. The thesis will investigate the learners' overuse, underuse, wrong use of articles, and these will be contrasted with correct use of articles. The research will investigate the 7th and 10th graders separately and based on this a cross-sectional study will be conducted. Utilizing a corpus allows for a tracking of frequency concerning the different uses of articles, in this case being either overuse, underuse or wrong use, and correct use. Previous research establishes that there are challenges for learners regarding correct article use, and Bækken (2002) explains that this is an area of problem for many Norwegian learners of English learners as well. This led to the following research questions:

1) How do young Norwegian EFL learners use articles in their writing?

2) What errors do the learners make concerning article use, and how many errors do they make compared to the instances where the articles are used correctly?

3) How does the use of articles in writing develop from grade 7 to grade 10?

These questions will be answered through a corpus study that will use statistical tools to present the findings. Through research question 1, the aim is to present a description of article use among the selected group of learners presented in the CORYL corpus, namely 7th grade learners and 10th grade learners. This is done through an error analysis of articles in the CORYL corpus along with manual searches for correct use of articles. Research question 2 aims to provide an overview of the error types that the learners make and to compare these findings with the instances where the learners use the articles correctly. Lastly, research question 3 will as presented previously be investigated through a cross-sectional study. The reason for investigating the development in use of articles in writing from grade 7 to grade 10 is partially due to the underlying hypothesis and notion that the learners in grade 10 will make fewer errors than the learners in 7th grade. By looking at the curriculum for the English subject in the Norwegian school system, it is evident that the competence aims in the curriculum differ from grade 7 to grade 10. Article use – or other grammatical features – are not mentioned specifically in these competence aims. However, the following aims describing production of text are collected from the aims after year 7 and after year 10.

(1.3) "Use basic patterns for spelling, wording, sentence – and text building in the production of text" (The Norwegian Directorate for Education and Training, 2013a).

(1.4) "Use central patterns for spelling, wording, sentence – and text building in the production of text" (The Norwegian Directorate for Education and Training, 2013b).

These two aims are rather similar, apart from one word. The learning aim for the 7th grade includes *basic patterns*, while the aim for the 10th grade includes *central patterns*. This difference in wording indicates that it is an aim that the 10th grade learners are to be more proficient in producing English text than the 7th graders. This is regarded to be the natural development, of course, as the learners become older and have become more experienced in the subject. Thus, it can be expected that the year 10 learners will be more proficient in the article use than the year 7 learners, and this is what research question 3 aims to investigate.

1.5 Structure

The thesis is structured as follows: Chapter 2 presents the theoretical framework used in the thesis. This chapter describes and presents relevant literature as well as the theoretical approach the paper will take. Chapter 3 demonstrates the method and the data collection process. Following this, the findings are presented in chapter 4 together with an analysis. Chapter 5 concludes the paper.

Chapter 2: Theory

This chapter sketches a theoretical background for the research carried out in this MA thesis. Section 2.1 presents a description of the English article system, which focuses on the indefinite articles *a* and *an*, the definite article *the* and the zero article. Furthermore, this section includes a comparison of the article systems in Norwegian and English. Subsequently, section 2.2 presents how articles are acquired and possible reasons for erroneous and correct use of articles. Section 2.3 discusses the role of transfer in language learning. Next, section 2.4 briefly introduces error analysis, before section 2.5 presents an introduction of the CEFR-scale from the Common European Framework of Reference. Lastly, a section on the Norwegian National tests in English will be given, as the texts in the corpus used for this research are derived from these tests.

2.1 The English article system

As noted in section 1.2, Master (2002) describes the English article system as a "notoriously complex aspect of English grammar" (p. 1). This notion is also expressed by Ionin et al (2008), who describe the acquisition of English articles as a "[…] notoriously difficult process for L2 learners" (p. 555). Many studies show that leaving out and/or using the English articles incorrectly is the case among many learners of English as a L2 (Ionin et al, 2008, p. 555). These studies also show that these types of errors "[…] appear to be particularly prevalent among L2-learners whose native languages (L1s) lack articles" (Ionin et al, 2008, p. 555). However, since both English and Norwegian have articles, the latter statement is not directly relevant for the learners on which the present research is based. Nevertheless, as section 2.1.5 will make clear, there is a rather noticeable difference between the use of articles in Norwegian and English. This difference will be important to take into consideration when analysing and discussing findings in chapter 4.

Master (2002) sheds light on what he believes to cause difficulty with correct use of articles among non-native speakers of English through three principle facts regarding the article system. The first fact, that articles "[…] are among the most frequently occurring function words in the language" (Master, 2002, p. 2), was highlighted in the introduction of this thesis. With articles being so frequently used when speaking or writing learners must make continuous decisions as to which of the articles to apply – or not to apply. This can lead to challenges for the learners

and it results in errors being made. Furthermore, with the second principle fact Master (2002, p. 2) underlines how function words usually are unstressed, which can cause the learners to have difficulties with distinguishing them. This is of course more applicable to the spoken use of English, but nevertheless, one can imagine that difficulties with using and understanding articles orally may affect their use in written work as well. Lastly, Master`s (2002) third principle fact, blames the article system itself for causing difficulties for the learners. Master (2002) explains how the "article system stacks multiple functions onto a single morpheme, a considerable burden for the learner, who generally looks for a one-form-one-function correspondence in navigating the labyrinth of any human language until the advanced stages of acquisition" (2002, p. 2). In other words, the use of articles is challenging for non-native speakers who have not reached the level of proficiency where they master this use, due to the complexity of the article system.

2.1.1 Determiners

The English article system consists of three types of articles: the definite, the indefinite and zero article. Articles are function words and can also be referred to as determiners. Determiners are used when specification is needed. Hasselgård, Lysvåg & Johanson (2012) explains how "the specification is provided by the choice of a singular or a plural form and by the use of determiners" (p. 118). Function words carry little meaning, but they are important for the structure of a clause or a sentence. To exemplify this, consider examples (2.1) and (2.2):

(2.1)   My sister is **a** teacher.
(2.2)   *My sister is teacher.

The sentence in example (2.1) is correct, with the indefinite article *a* preceding the noun. However, the sentence in (2.2) is still understandable but there is a structural error as the indefinite article is missing. Using the wrong article or leaving the article out in this example can cause the reader or listener to have difficulty understanding exactly what the person is trying to convey, end even if (2.2) is comprehensible, its structure is not grammatical in Standard English.

2.1.2 The definite article

The definite article in English is *the.* It is a very high-frequency word: according to the Oxford English Corpus (OEC) (2018), which consists of over 2 billion words, the most common word in the corpus is the definite article *the.* Additionally, Master (2002, p. 2) refers to Sinclair (1991), who presents a list of the five most frequently occurring words in English. This list places *the* in first place, while *a* receives a fifth place. Looking at the list from the OEC it is interesting to note that *the* was also in first place here, while *a* is in sixth place.

According to Bækken (2006) "[t]he definite article is used with all types of nouns, except proper nouns, in the singular, as well as the plural" (p. 103). When using the definite article one is referring to something that is familiar or known to the listener or reader, expressed as a noun. Furthermore, the definite article has two types of use; the generic and the specific use. The specific use of the definite article is the most common use of this article (Bækken, 2006, p. 103). When using the definite article to refer to something specific it is assumed that what is being referred to has been introduced to the reader previously in the text. An example of the specific use of the definite article follows below:

(2.3)   I received **a** letter yesterday.
(2.4)   **The** letter came from England.

It can be observed that when the letter is introduced as new information (2.3), the indefinite article is used. Once this referent is established in the discourse, however, it may be referred to by a noun phrase marked for definiteness by means of the definite article. Example (2.4) shows the specific use of the indefinite article, as it refers to a specific letter. On the other hand, the generic use of the definite article – would do the opposite: the generic use of the definite article refers to a class or whole species (Bækken, 2006, p. 101). This can be exemplified as in (2.5) below:

(2.5)   **The** calculator is a great tool.

Here, reference is made not to a specific calculator, but rather to this type or class of machine. However, it is important to note that generic reference with the definite article must be used with some caution. Bækken (2006, p. 101) explains how in cases when singular countable nouns are referred to, as in the example above, the definite article may in some cases change the

reference to specific rather than generic. In relation to this, Bækken (2006, p. 101) underlines how important the context is to interpret the text correctly. An example of this is given in (2.6), where the definite article is used to reference a specific calculator, opposed to the generic reference in the example (2.5) above:

(2.6)   She dropped **the** calculator on the floor.

2.1.3 The indefinite article

The indefinite article has two variations: *a* and *an*. Of these two, *a* is the most frequently used of them, as pointed out in section 2.1.1. Whether to use *a* or *an* is decided by several factors. As opposed to the definite article, which refers to something that is known to the listener or reader, the indefinite article refers to something that is unknown. One is to use *a* when the article precedes a noun that begins with a consonant sound. *An* is to be used when the article precedes a noun that begins with a vowel sound. (2.7) and (2.8) below exemplify this:

(2.7)   I ran into **a** guy at the store.
(2.8)   There is **an** orange house down the street.

At first glance these rules for the use of the indefinite article might seem comprehensible. However, it is imaginable that learners might find it difficult to separate their experience of vowels and consonants being taught as letters, and not necessarily sounds. An example of this important distinction can be the following:

(2.9)   I saw **an** ugly car today.
(2.10) He is attending **a** University Fair next week.

The word following the indefinite article in example (2.9), *ugly*, begins with a vowel sound and is pronounced /ˈʌgli/. The word following the indefinite article in example (2.10) on the other hand, begins with a consonant sound, despite being the same letter as in example (2.9). *University* is pronounced /juːniˈvɜːsəti/. Both words begin with the letter <*u*>, but due to the different sounds caused by the adjoining letters forming the word, the indefinite article will differ in use.

The indefinite article is used only in the singular and it often occurs with countable nouns (Bækken, 2002, p. 110). Like the definite article, the indefinite article also has two types of use; the generic and the specific. The specific and generic uses of the indefinite article are based on a similar foundation as for the specific and generic uses of the definite article. Example (2.1) in section 2.1.1 shows the specific use of the indefinite article. The generic use of the indefinite article is used when reference is made to a whole group or class of something. This can be exemplified as follows:

(2.11) One should give **a** child plenty of encouragement.

(Example from Bækken, 2012, p. 101).

In contrast to the limitations the generic use of the definite article faces together with singular countable nouns, the generic use of the indefinite article "[…] seems to be possible with all singular countables […]" (Bækken, 2002, p. 101).

2.1.4 Zero article

There are cases where none of the articles above is used; the article is simply left out. This omitted article is referred to as the *zero article*, and the result is a bare noun phrase. According to Parrot (2010), the article is left out "[…] before plural and uncountable nouns when we are referring to something in general" (p. 28). This can be exemplified as follows:

(2.12) I eat oatmeal for breakfast.

Some countable nouns, which usually take the definite article, "[…] occur without article in English in special uses" (Bækken, 2002, p. 116). Examples of this are words describing public institutions, such as *school* or *prison.* Bækken (2002) explains how these nouns occur without an article when one is referring to the "[…] purpose of the building or place, or the activity connected with it" instead of "[…] the building or place itself" (p. 116). This is exemplified below:

(2.13) I am going to school.
(2.14) The school is under renovation.

Example (2.13) shows how the noun *school* occur without article, as this word is in this context refers to the purpose of the school, not the building itself. However, (2.14) exemplifies the use of the noun *school* referring to the building itself. This causes the noun to take on the definite article.

2.1.5 The difference between the article systems of Norwegian and English

As stated previously in this thesis, article use can be problematic for Norwegian learners. A prominent reason for the problems learners may face is the difference in the use of articles in English and Norwegian. Moen & Pedersen (2003) explain how in Norwegian "[t]he definite article is not a separate word as in English, but a post-positioned inflectional ending of the noun" (p. 54). These endings are decided by gender; masculine, feminine or neuter. These endings differ from the singular to the plural. The following examples (2.15 – 2.17) demonstrate the endings in the singular:

(2.15) Masculine: *-en/-n*: bil**en** (the car), skole**n** (the school)

(2.16) Feminine: *-a*: jent**a** (the girl), tavl**a** (the blackboard)

(2.17) Neuter: *-et/-t*: bord**et** (the table), eple**t** (the apple)

<div align="right">Examples inspired by Moen & Pedersen (2003, p. 55)</div>

In the plural there are two forms of the definite article, as opposed to three forms in the singular (Moen & Pedersen, 2003, p. 55). The definite article in the plural is the same for the masculine and feminine (*-ene*), while a distinct form (*-a*) is used for the neuter:

(2.18) Masculine: bil**ene** (the cars)

(2.19) Feminine: jent**ene** (the girls)

(2.20) Neuter: bord**a** (the tables)

<div align="right">Examples inspired by Moen & Pedersen (2003, p. 55)</div>

The indefinite article in Norwegian is more similar to English than the definite article. Like in English, the indefinite article in Norwegian is also a separate word preceding the noun. There are three forms of the indefinite article; *en, ei, et*. The appropriate use of these articles is decided based on the gender features of the noun phrase head.

(2.21) Before a masculine word: **en**, en bil (a car)

(2.22) Before a feminine word: **ei**, ei tavle (a blackboard)

(2.23) Before a neuter word: **et**, et eple (an apple)

<div align="right">Examples inspired by Moen & Pedersen (2003, p. 55)</div>

The information above can also be presented in a table such as the one below:

Table 2.1: Articles in English and Norwegian

	Singular		Plural	
	indefinite	definite	indefinite	definite
Masculine	**en** bil (**a** car)	bil**en** (**the** car)	biler (cars)	bil**ene** (**the** cars)
Feminine	**ei** jente (**a** girl)	jent**a** (**the** girl)	jenter (girls)	jent**ene** (**the** girls)
Neuter	**et** eple (**an** apple)	epl**et** (**the** apple)	epler (apples)	epl**ene** (**the** apples)

There is no doubt that there are noticeable differences in the article systems of Norwegian and English. For young learners of English it may be challenging to get a firm grip on this grammatical feature due to these differences. In the process of acquiring the necessary skills and knowledge to comprehend grammatical features and rules, transfer from the L1 to the L2 may occur. As stated in section 1.2, Bækken (2002) claims that transfer from Norwegian seems to be the reason for many of the errors that Norwegian learners of English make. Transfer from the L1, in this case Norwegian, "[…] usually refers to the incorporation of features of the L1 into the knowledge systems of the L2 which the learner is trying to build" (Ellis, 1996, p. 28). As transfer can occur in the language learning process it is important for language instructors to be aware of this, and to have a conscious and thorough understanding of the difficulties learners may face due to difference between L1 and L2. This feature is also something that has to be considered when analysing the findings in chapter 4.

2.1.6 Transfer from Norwegian

The concept of transfer was briefly mentioned in the previous section. It is deserving of a fuller explanation since it has high relevance for this research. Transfer is viewed to be a possible reason for errors in article use and it is therefore important to explore this. Views in terms of language transfer have changed through the years. From the behaviourist perspective, it was assumed "[…] that the 'habits' of the L1 would be carried over into the L2" (Ellis, 1996, p. 29).

VanPatten & Williams (2015, p. 20) describe the behaviouristic view on second language learning, noting that it was believed to be a process of acquiring a new set of behaviours, and that the first language could obstruct this process. The behaviourist perspective also maintained that transfer was partly the reason if the acquisition of the L2 failed (VanPatten & Williams, 2015, p. 20). It was said that transfer occurred "[…] when learners relied on the L1 used in attempting to produce the L2" (VanPatten & Williams, 2015, p. 20).

It is natural to discuss the different roles transfer can take on. Both VanPatten & Williams (2015, p. 20) and Ellis (1996, p. 29) discuss the fact that transfer can be both positive and negative. Whether it is positive or negative depends "[…] on the distance between the L1 and L2" (VanPatten & Williams, 2015, p. 20). When the L1 is similar enough to the L2 that the learners can simply apply old habits into a new context, the transfer is regarded to be positive (VanPatten & Williams, 2015, p. 20). Negative transfer on the other hand, is the opposite. When the L1 is so different from the L2 that the learners *cannot* apply old habits into the new context, the transfer is referred to as negative, which leads to difficulties for the learners and causes errors.

As with any other feature of SLA research, transfer has also been a debated topic. As mentioned in the previous paragraph transfer of L1 features to L2 can lead to errors. Some of the criticism of the concept of transfer in the behaviouristic view has been directed towards just this. In attempts to predict "[…] when interference would take place linguists carried out elaborate contrastive analyses of the native and target languages" (Ellis, 1996, p. 29). The contrastive analysis theory claimed that errors made in the L2 were caused by the difference in the L1 of the learner and the L2 (Ellis, 1996, p. 698). However, these claims were not empirically tested until the late 1960's, through error analysis studies (Ellis, 1996, p. 29). The error analysis studies conducted at that time questioned the claims posed by the contrastive analysis theory and showed that in cases where the L1 and L2 were different, transfer often did not occur (Ellis, 1996, p. 29). Additionally, many of the errors made by the learners "[…] appeared to reflect intralingual processes (i.e. they were the result of processed based on the learners' existing knowledge of the L2) rather than interference" (Ellis, 1996, p. 29).

Despite the criticism and questioning of the role of transfer in the late 1960s, it has regained some of its position in the SLA research field. It has moved from being very error-oriented to including focus on facilitation, overuse and avoidance (Ellis, 1996, p. 29). In other words,

transfer from an L1 to an L2 does not necessarily solely focus on errors, but also on whether some habits from the L1 causes overuse of a certain feature in the L2, or the opposite, namely avoidance of using a feature due to transfer. Errors in terms of overuse and avoidance (i.e. underuse) are highly relevant for the specific research carried out in the present MA thesis, and it is important to consider whether such errors could be caused by transfer from the L1. Lastly, there is now a general acknowledgment that transfer "[…] works in complex ways and that it constitutes only one of several processes involved in L2 acquisition" (Ellis, 1996, p. 29).

2.2 Article acquisition and possible reasons for errors and correctness

Through a corpus-based study which this research is based upon, no information about the learners' acquisition of articles is disclosed. Neither is any information dealing with how the learners have been taught this grammatical feature. Nevertheless, the acquisition of articles and possible reasons regarding errors in the use of articles is still very much relevant for this thesis. The third research question presented in section 1.4 poses the question "How does the use of articles in writing develop from grade 7 to grade 10?". This question aims to investigate possible differences in article use among different learners in 7th grade and 10th grade through a cross-sectional study. Considering this, it is useful to include a section here on how articles are acquired, and possible differences between learners of different levels of proficiency. Furthermore, being aware of the possible reasons for errors and correctness in terms of the use of articles, is useful as it may simplify the process of discovering patterns. However, previous research on article acquisition in different proficiency levels of a language – for example English – often focus on levels of proficiency that differ much more than just 7th to 10th grade. Also, such research often focuses on the difference in acquisition of articles across other languages, for example article acquisition in Spanish versus English. A great deal of this research also deals with languages that are more typologically different from Norwegian and English, such as Chinese. Nevertheless, some of this research is included in this thesis as it still may be applicable to this thesis, at least partially.

Peter Master, as referred to previously in this thesis, has written several articles regarding article use and acquisition. In one of his articles Master (1997) discusses the English article system in regard to acquisition, function and pedagogy. Without going too deep into the pedagogy of articles, as this is not directly relevant for this thesis, Master (1997) describes the teaching of the English article system as "[…] a somewhat controversial proposition" (p. 215). There is

division as to how articles should be taught, and how much time should be devoted to this (Master, 1997, p. 215). Nevertheless, Master (1997) points to the fact that evidence suggests that formal instruction has a positive effect on the acquisition of articles, and that this instruction is necessary (p. 216). It is also interesting to be aware of another notion presented by Master (1997), namely that "[…] at the lowest proficiency levels, it is probably not worthwhile directing any sustained attention to rules of article language" (p. 226). However, due to the lack of information regarding the learners' level of proficiency in the CORYL corpus, and without knowing precisely what is meant by 'lowest proficiency level', this notion should be carefully considered. Nevertheless, it is an interesting notion that indicates a possible difference in the teaching and priority of articles in writing with respect to different proficiency levels. It remains to see if this is something that will be reflected in the results presented in chapter 4.

As explained in section 2.1.5, Norwegian has an article system, but there are differences between this system in Norwegian and English. Master (1997, p. 216) refers to his own master thesis where he investigated the sequences of article acquisition. This study suggested that "articles appear to be acquired differently depending on whether they occur in the learner's first language" (Master, 1987, cited in Master, 1997, p. 216). Therefore, speakers of languages like Norwegian or Spanish, which are languages with an article system, will acquire articles in English in different sequences than speakers of Japanese, which does not have an article system. This is a feature that is found in other research papers as well. One example is Ganzhao Sun's (2016) article on 'The acquisition of English articles by Second Language Learners: The Sequence, Differences, and Difficulties". Sun (2016) refers to a range of researchers who state that native speakers of English acquire the proper use of English articles "[…] unconsciously and at an early stage" (p. 1). Furthermore, Sun (2016) also points to the fact that "English-as-a-second language learners, especially those whose first languages (L1) do not have the equivalent article systems, tend to have difficulty acquiring articles" (p. 1). Sun (2016, p. 5) refers to Master and his thesis from 1987, and states that the results from his own research are similar to what Master presented in 1987. However, some of the findings Sun (2016) presents in his article – does in fact "[…] challenge commonsense belief that "+Art" ELLs [i.e. speakers of a language with an article system – SL]" acquired English articles faster than "-Art" ELLs…" [i.e. speakers of a language without an article system – SL]" (p. 1). These findings included the fact that the +Art ELL'S had most difficulties with the definite article *the,* and that both the +Art and -Art learners of English as a second language found the zero article to be the article which was most difficult to acquire (Sun, 2016, p. 1). These findings then

challenge the commonsense belief that learners of English with a first language containing an article system does not necessarily acquire articles more easily than learners of English with a first language not containing an article system.

Master (1997) notes that in the process of acquiring English, language learners generally find "[…] grammatical items that are comprised of multiple elements" difficult to acquire (p. 220). It therefore seems safe to consider articles as one of these grammatical items. In order to arrive at "[…] the correct choice of article", one has to consider countability, number, and definiteness (Master, 1997, p. 220). This notion will be considered and discussed in relation to the findings in chapter 4. The 'multiple elements' which Master (1997, p. 220) refers to, must be viewed in the context which the errors occur. By doing this, a contribution may be made to knowledge of what elements cause the errors which the learners in the CORYL corpus make.

As for the specific features of acquisition and mastering of article use, Master (1997) suggests that speakers of languages with an article system "… overuse *the* right from the start and use Ø [zero article – SL] correspondingly less than their [-ART] counterparts [i.e. speakers of a language without an article system – SL]" (p. 218). As for the acquisition of *a*, it "[…] seems to occur almost independently of *the*" (Master, 1997, p. 218). Chaudron and Parker (1988) cited in Master (1997) "found that *a* began to emerge only after *the* was largely mastered and suggested that it may interfere with earlier correct applications of the system, even for learners with articles in their L1s" (p. 218). Sun (2016, p. 2) describes Master's English article acquisition order among learners whose L1 has an article system to be *the > null > a.* According to Master (1997), a possible reason for why *the* is learned first is that "[…] it is not limited by countability or number" (p. 225). As for the order of article acquisition, the one provided by Master (Sun, 2016, p. 2) should not necessarily be viewed as a definite order. In the study by Sun (2016) the order of article acquisition differed from Master's order. Sun (2016, p. 5) presents findings of a different order of article acquisition among several different levels of the learners, with the common denominator being the fact that the majority of the groups had most difficulties with acquiring the zero article. These notions are useful to be aware of when going into the analysis of the findings from the CORYL corpus, as they might prove to be relevant for the findings presented in chapter 4.

To conclude this section, it is worthwhile to briefly visit a possible feature that may affect the use of articles, both correct and incorrect. Despite the differences of the learners who have

contributed to the CORYL corpus, they have one thing in common; the fact that they all had the same tasks to choose between, at least within the separate grades. Therefore, one can imagine that the text type or task design potentially has an effect on the learners use of grammatical features, and more specifically, their use of articles. It is difficult to find research on this topic specifically directed towards article use. Nevertheless, there are some indications of correlations between task type and writing performance that are relevant to mention here. Kuiken & Vedder (2008:49) cited in Salimi, Dadaspour and Asdollahfam (2011) point out that "in the literature on both L1 and L2 writing, it has been suggested that some task types result in lower test scores than others, but the relationship between task type or task complexity and writing performance is by no means clear" (p. 1390). Based on this notion, the article 'The Effect of Task Complexity on EFL learners written performance' by Salimi et al (2011) presents the finding that "complex task led to the production of more complex language in terms of syntactic mode" (p. 1397). It is possibly stretching it somewhat, but this tendency might be possible to investigate – or at least consider – in this MA thesis when analysing the errors the learners make concerning article use. This can be attempted to investigate by considering the text type or the task design. Then, a pattern may be revealed as to what may cause the errors or how the task types trigger a specific type of error.

2.3 Error analysis

Section 2.1.6 briefly introduced Error Analysis in the context of transfer, and reference was made to how error analysis casts doubt on the Contrastive Analysis theory. Contrastive Analysis (CA) was the most prominent type of analysis within this field, before it was replaced by Error Analysis (EA) (Ellis, 1996, p. 47). The difference between these two is that "[…] an underlying assumption of CA was that errors occurred primarily as a result of interference […]" (Ellis, 1996, p. 47). As mentioned in section 2.3, interference describes the process where the learner`s L1 interferes with the L2 in a way that can lead to errors. CA only focused on the learner's L1 and the target language, while EA "[…] provided a methodology for investigating learner language" (Ellis, 1996, p. 48). This methodology includes five steps, suggested by Corder (1974). The steps are as follows;

1) Collection of a sample of learner language
2) Identification of errors
3) Description of errors
4) Explanation of errors

5) Evaluation of errors

<div align="right">(Corder, 1974, in Ellis, 1996, p. 48)</div>

Step 1, collection of a sample of learner language, describes how and what data is to be collected. Ellis (1996, p. 50) underlines how natural samples are preferred in such research. The second step in EA suggested by Corder (1974) is identification of errors. This means that one must have decided "[…] what constitutes an 'error' and established a procedure for recognizing one" (Ellis, 1996, p. 50). Moving on, the third step is description of errors. Describing the errors made by the learners "…involves a comparison of the learner's idiosyncratic utterances with a reconstruction of those utterances in the target language" (Ellis, 1996, p. 54). This process does not set out to discover the sources of the errors (Ellis, 1996, p. 54). Therefore, this third stage of EA involves systematically creating these descriptions. Ellis (1996, p. 54) states that the simplest way to do this is to utilize the linguistic categories and to break them down into more specific categories. An example of this could be the way the research is set up for this specific thesis. The main category is article errors, while the more specific categories are underuse, overuse and wrong use of the different categories. The fourth step in EA is suggested to be explanation of errors. If one can describe and identify the errors, explaining them is the next step (Ellis, 1996, p. 57). The process of explaining the errors deals with finding the source of the error, in other words, "[…] accounting for why it was made" (Ellis, 1996, p. 57). Ellis (1996, p. 48) explains how Step 5 has been left out in many previous studies, as this has been handled separately.

Following this brief introduction of the stages in error analysis, it seems natural to discuss the limitations and critiques of error analysis. A prominent limitation of EA is how it only focuses on errors, which in turn does not provide a complete picture of how learner language in fact is (Ellis, 1996, p. 68). Nevertheless, EA does provide a partial picture, and can be a "[…] preliminary source of information at an initial stage of investigation" (Hammarberg, 1973, p. 34, in Ellis, 1996, p. 68). A more through discussion of the limitations will be presented in chapter 3.

2.4 Common European Framework of Reference

According to the metadata provided through the CORYL corpus the corpus texts "… are coded for errors and are also linked to the Common European Framework (CEFR) levels" (CORYL,

n.d). The metadata further describes that "this means that students can search for instances of the errors produced at different CEFR levels, for the various age groups" (CORYL, n.d). The Common European Framework of Reference for Language is often referred to as CEFR, and it "[…] provides a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe" (Council of Europe, 2001, p. 1). This framework includes what is necessary to communicate through a language and how to do it efficiently (Council of Europe, 2001, p. 1). Furthermore, the framework "[…] also defines levels of proficiency which allow learners` progress to be measured at each stage of learning and on a life-long basis" (Council of Europe, 2001, p. 1).

The defined levels of proficiency describe a range of language skills that have been set into tables based on different categories. These categories include listening, reading, spoken interaction, spoken production and writing (Council of Europe, 2001, p. 26). Some of the language skills are understanding, expressing, use and production of language and communicating (Council of Europe, 2001, p. 24). These appear in a graded manner, from basic to very proficient. The CEFR-grades consist of six grades: A1, A2, B1, B2, C1, C2. A1–A2 describe the basic user of a language, B1–B2 describe the independent user of a language, while C1–C2 describe the proficient user of a language (Council of Europe, 2001, p. 24).

A small number of texts in the CORYL corpus have been assigned a CEFR grade. The CEFR grades do not specifically describe the learners' level of proficiency regarding article use. These two points lead to some implications concerning the use of the CEFR grades in this research. Since only a minority of the texts are assigned a CEFR grade, it is not possible to use this feature as a part of the analysis of the findings as a total. Also, since the CEFR grade does not disclose anything specifically about article use, it is challenging to connect this directly to the learners' use of articles. Despite this, the following paragraph discuss article errors and the levels in the CEFR scale more specifically as it may prove to be useful for the analysis in chapter 4.

The framework does not describe a specific language. Therefore it is useful with resources like English Profile. The booklet *Introducing the CEFR for English* by English Profile include information about the CEFR levels, and its intention is to present "a 'profile' of English language learners in terms of the six proficiency bands of the CEFR–A1 to C2" (2011, p. 2). This profile is based on "[…] most – but not all – of the structural features that the corpus-informed research has shown to be significant for each level from A2-C2" (English Profile,

2011, p. 12). This profile goes into detail about "[…] error types that significantly improve between adjacent levels" (English Profile, 2011, p. 25). For this thesis, the interesting error types are article errors, or determiner errors. Errors regarding form of determiner is an error type that, according to the English Profile (2011), "improve[s] significantly between A2 to B1 levels", specifically "[w]hen the articles 'a' and 'an' are confused" (p. 25). This error continues to improve from B1 to B2 and from C1 to C2 levels as well. Furthermore, on the English Profile website, is it possible to search the English Grammar Profile for categories of grammatical skills, as well as which CEFR level the different features belong to. Below is a compiled list of the found grammar profile regarding determiners, and hereunder specifically article use, according to CEFR grade level.

(2.24) A1: can use articles 'the', 'a' and 'an' before nouns.

(2.25) A1: can use 'a' and 'an' before adjectives in a noun phrase.

(2.26) A1: can use 'a' + 'very' + adjectives in basic noun phrases.

(2.27) A1: can use 'the' in prepositional phrases relating to time.

(2.28) A1: can use no article before a limited range of nouns in some fixed expressions with prepositions.

(2.29) A1: can use no article before a limited range of singular, plural and uncountable nouns when referring to things in general.

(2.30) A2: can use 'the' + adjectives in a noun phrase, to specify.

(2.31) A2: can form a noun phrase with 'the' + superlative adjectives + noun.

(2.32) B1: can use 'another' with singular nouns.

(2.33) B1: can use no article before an increasing range of nouns in some fixed expressions with prepositions.

(2.34) B1: can use 'another' to talk about something additional.

(2.35) B1: can use 'another' to talk about something different.

(2.36) B1: can use 'the other' with a singular noun to refer to the second or the opposite of two things, and with a plural noun to refer to one of a set of things.

(2.37) B2: can use the comparative phrases 'more', 'less', 'worse', 'better… the more', 'less'. 'worse…' to talk about one thing that is affected by another.

(Cambridge University Press, English Profile, 2015)

As mentioned previously, due to limited information concerning the learners CEFR levels, these will not be a prevalent part of the analysis. However, the specific examples of article use in the different CEFR levels from the English Profile are useful in that they give specific examples of what is expected of the different CEFR levels of proficiency within article use. These examples also underline the fact that there is an expectation of a clear progression in proficiency. Also, it might be interesting to be aware of these levels when discussing specific findings from the CORYL corpus. Nevertheless, it is still important to keep in mind that the CORYL corpus does not contain CEFR level information on all of the learners and their texts. Therefore, caution must be showed when discussing the CEFR levels in relation to the findings in chapter 4.

2.5 National tests in English

The corpus used in this research consists of English texts written by Norwegian pupils. These texts are collected from the National tests in English in 2005 and 2011, among learners of the age of 12-13 and 15-16 years. According to The Norwegian Directorate for Education and Training (2018), the purpose of national tests is to acquire knowledge about the pupils' basic skills in English, math and reading. Moreover, this knowledge should be the foundation of formative assessment and development of quality in the different levels of the school system (The Norwegian Directorate for Education and Training, 2018). The tests are now electronic but were previously carried out on paper. The tests consist of pictures, questions and different tasks. The pupils must answer both open-ended tasks and multiple-choice tasks. Furthermore, the test utilizes different text types, such as letters, descriptions, reviews, stories or tables. The text types that occur in the tests related to the CORYL-corpus are description, story and letter for the 12/13-year-old group, and essay, letter to the editor and personal letter for the 15/16-year-old group. Lastly, when the learners have taken the National tests in English, pupils are placed in different mastery levels based on their test scores. These scores are not included in the CORYL corpus.

The national tests were first held in 2004. In the years following, evaluations were done, and the tests reappeared in a changed format. These changes involved that the tests were to be carried out in 5th, 8th and 9th grade, and not 4th, 7th and 10th grade as previously. Therefore, the tasks (see appendix A and B) which the texts in the corpus are based on, belongs to the period before the new changes were put into practice. Whether this could possibly affect the results of this research is difficult to comment on, as there does not exist a corpus that the CORYL corpus

and the findings from it can be compared to. Moreover, the test from 8th grade is retaken in 9th grade, so that the results can be compared. The tasks that make up the test varies greatly in respect to level of difficulty, as the test aims to test pupils in all levels of proficiency and understanding (The Norwegian Directorate for Education and Training, 2017). The National test in English aims to test three areas from the competence aims in the English curriculum. These areas include general reading comprehension, vocabulary and grammar (The Norwegian Directorate for Education and Training, 2017). For this thesis, how the tests aim to test grammar is especially interesting. The Norwegian Directorate for Education and Training (2017) explains how the pupils must understand sentence structure, word forms, and function words, to comprehend grammar. As articles are recognized as function words, it is expected that the pupils should master or at least show some understanding of article use in the tests. However, it is important to note that none of the tasks given in the tests relevant for the texts in the corpus aims to test the grammatical feature that is articles specifically.

Furthermore, it seems natural to include some comments about how the National tests are perceived by school administration, teachers and pupils. A report about evaluation of national tests as a system from 2013, presents the evaluation of several features of the tests from different viewpoints. This rapport, written by Seland, Vibe & Hovdhaugen (2013), states that "overall, national tests are mainly implemented in line with the intentions set out in guidelines and instructions from the government" (2013, p. 11). Additionally, opinions are divided among school owners, teachers and pupils about the results of the tests and how they are used. Many school owners believe that the results are useful in relation to development, while the "[…] teachers are more reserved in their praise for national tests" (Seland et al, 2013, p. 13). Furthermore, the report states that the pupils' "[…] wish to get individual feedback on their performance" (Seland et al, 2013, p. 12). As the tests are designed to provide information for the government and principals about school performance, this contradicts the pupils` wish regarding the possibility of individual feedback (Seland et al, 2013, p. 12). It is then important to discuss and to be aware of how this testing affects the learners, and possibly most importantly, what their incentives are for taking these tests. Unfortunately, such information is not available through this research project, nor the aim of the research either, but still important to bear in mind when using texts produced by learners, which originates from these national tests.

2.6 Summary

This chapter presents the theoretical background for the research carried out in this MA thesis. Section 2.1 presents the English article system and the different types of articles are presented and exemplified. This section also deals with the concept of transfer in second language learning. Section 2.2 discuss article acquisition and possible reasons for errors and correctness and looks to previous research within this field. Section 2.3 briefly introduce Error Analysis and its position in language research. Section 2.4 presents the Common European Framework of Reference and its relevance to this MA thesis. Lastly, section 2.5 presents information concerning the Norwegian National tests in English and how the CORYL corpus has utilized texts from these tests in the compilation of the corpus.

Chapter 3: Method

The data collection in this thesis is based on corpus linguistic methods. Widely understood such methods encompass everything from corpus-creation – by collection linguistic data, computerizing it, and annotating the corpus – to analysing corpus data following extracting from an existing corpus (Meyer, 2002; McEnery & Wilson, 2001). The present research uses an already compiled and annotated corpus, namely the CORYL corpus. For this specific research, then, the corpus linguistic methods used include data extraction based on the annotation of the CORYL corpus and the corpus analysis tools available. This method chapter will be an in-depth presentation of the method of data collection used for this thesis, and considerable detail is included as concerns query creation as well as data handling and interpretation. Additionally, some features of Error Analysis (EA) are also drawn upon in this thesis when investigating the errors made by the learners in the CORYL corpus concerning article use. Consequently, this chapter will also discuss two of the most prominent features of EA, namely identification of errors and description of errors (Corder, 1974, cited in Ellis, 1996, p. 48).

The different elements of the method employed in this research are presented in the order which they were conducted. This chapter is structured as follows: Section 3.1 presents the reasons for choosing to work with the CORYL corpus in addition to an introduction of the corpus. Section 3.2 describes the data extraction process as well as the annotation and the analysis tools found in the corpus. This section also goes into detail on how the queries for the research were created. Section 3.3 deals with data handling and section 3.4 presents the statistical tools used in this MA thesis. Lastly, section 3.5 concludes this chapter by discussing a number of implications and other remarks of the use of the specific methods presented throughout this chapter.

3.1 Corpus of Young Learner Language

In the process of selecting the learner corpus, several things must be considered, such as which group of learners one wishes to investigate or the language one wishes to research. Granger (2012, p. 14) explains how it can be problematic to access different corpora, and the importance of finding a corpus suitable for the research in question. For this research, when deciding to research learner language based on a corpus, some criteria were set. Firstly, it was important to find a corpus that contained texts written by Norwegian EFL learners. Secondly, it was essential

that these learners were young, preferably in the range of 5th–10th grade. The reason for this was, as mentioned in the introduction, that very few corpus studies focus on this group of learners. Additionally, this is the age group that was most interesting for the researcher to investigate, as it is directly connected to the researcher's education. The Corpus of Young Learner Language (CORYL) was chosen as it meets the criteria specified above.

The CORYL corpus is an electronic corpus developed at the University of Bergen and it is made up of learner texts written by Norwegian EFL learners in 7th, 10th and 11th grade. The texts are collected from the National Tests in English conducted in 2005 and 2011. The texts can be placed into six different text types, namely: 'story', 'description', 'letter', 'personal letter', 'letter to the editor' and 'essay'. The length and quality of the texts vary, and some of the texts have been scored according to the Common European Framework of Reference scale presented in section 2.5. The texts written by the group of 7th grade learners are a mix of stories, descriptions and letters. The texts written by the group of 10th grade learners are either personal letters, letters to the editor or essays. The learners who have produced these texts are a mix of females and males, the ratio between the genders not known. The tasks which these texts are based on can be found in the appendix (A, B).

The CORYL corpus is a rather small corpus, only consisting of 129 421 words [As of 04.05.19 it consists of 191 568 tokens – S.L]. For example, in comparison, 'The Norwegian Newspaper Corpus' consists of 1 509 076 098 words and punctuation marks (Clarino, 2019). Smaller corpora mean smaller data sets, which again can lead to findings that may not be representative for larger groups. However, Granger (2012) stresses an important notion concerning the size of a corpus, namely that "[…] the optimal size of a learner corpus depends on the targeted linguistic phenomenon" (p. 9). Articles are exemplified as one of these grammatical phenomena that are so frequent that it is possible to investigate the use of them based on a small corpus (Granger, 2012, p. 9). Additionally, the goal of this research is not to provide a generalizable result, but to investigate the use of articles among learner texts compiled in this specific corpus. Based on these considerations, the CORYL corpus is eminently suitable for the study presented in this MA thesis.

3.2 Data extraction

3.2.1 Annotation and corpus analysis tools

The CORYL corpus is an annotated corpus. According to McEnery & Wilson (2001), annotation is "[…] the practice of adding explicit additional information to machine-readable text" (p. 197). There are different types of annotation used in corpora, depending on the purpose of the specific corpus, such as part-of-speech annotation or lemmatisation. Part-of-speech tagging is "[…] a type of annotation or tagging whereby grammatical categories are assigned to words (or some cases morphemes or phrases), usually via an automatic tagger although human post-editing may take place as a final stage" (Baker, Hardie, McEnery, 2006, p. 128). McEnery & Wilson (2001) explains how "[c]ertain kinds of linguistic annotation, which involve the attachment of special codes to words to indicate particular features, are frequently known as 'tagging' rather than 'annotation', and the codes which are assigned are known as 'tags'" (p. 46). A different type of tagging is error tagging, which differs from for example part-of-speech tagging in that it focuses on different types of errors rather than the large grammatical categories. The CORYL corpus is only tagged for errors and therefore excludes the possibility of investigating features that require part-of-speech tagging or lemmatisation. Notably, as a consequence of this, error tagging as a type of annotation is rather limited, as it lacks the structure needed for investigating for example invisible grammatical features, such as the zero article. However, there are also positive features regarding error tagging. Error tagging is viewed to be especially useful, because "[…] error tags are inserted into the text files and are hence presented in the full context of the text, alongside non-erroneous forms" (Granger, 2012, p. 15). This allows for researchers to view the errors in the context in which it occurs, and it might enable the users of the corpus to detect recurring features associated with the errors.

Error tagging is based on error coding. Most systems code errors according to "[…] error type (number, gender, tense, etc.), word category (noun, verb, etc.), and/or error domain (spelling, grammar, lexis, etc.) (Granger, 2012, p. 15). The CORYL corpus is only tagged for errors, which means that there is no syntactic or part-of-speech tagging. The CORYL corpus has 24 different error tags. These tags cover error domain through tags such as SP (spelling), WFU (wrong function word), and WW (wrong lexical word). Furthermore, the tags cover error type through tags such as VC (concord error) and WFO (wrong form of word). Word category is also covered with tags such as ART (any clear article error) and IT (*it/there* errors). For this research however, the error tag ART is in focus. If searched for in the CORYL corpus, the ART tag will present "any clear article error" (CORYL, nd). Speaking more broadly, this includes

wrong choice of article, as well as underuse and overuse of articles. Wrong choice of article includes cases e.g. where the learners have used the definite article *the* where the correct article is the indefinite article *a*, and so forth. Underuse errors constitute cases where the learners leave out an article where there should in fact be one. Lastly, overuse of article is when the learners use articles where the article should be left out (i.e. where the zero article should be used).

Information regarding instances where the learners use articles correctly is also a part of this data collection. As the CORYL corpus is not lemmatised, but only tagged for errors, it imposes certain restrictions as to what types of data which are possible to extract. According to McEnery & Wilson (2001), lemmatisation "[…] involves the reduction of the words in a corpus to their respective lexemes – the head word form that one would look up if one were looking for the word in a dictionary" (p. 53). Lemmatisation enables the user of a lemmatised corpus to "[…] extract and examine all the variants of a particular lexeme without having to input all the possible variants, and to produce frequency and distribution information for the lexeme" (McEnery & Wilson, 2001, p. 53). For this specific thesis, had the CORYL corpus been lemmatised it would have been possible to extract all the instances where the learners used the zero article correctly. However, due to the lack of lemmatisation in the CORYL corpus, the process of investigating the correct use of the zero article is an extremely time consuming and complex task. Doing this would entail going through every text manually to identify cases of correct use of the zero article. Despite the small size of the CORYL corpus this would be an extremely demanding manual process which is outside the scope of the present MA thesis, primarily due to time constraints. Therefore, this thesis will not enumerate correct uses of the zero article, and this category is therefore omitted from the data collection process. This is admittedly unfortunate as correct use of the zero article is obviously a relevant dimension of article use among the learners under investigation. This is, however, an unfortunate but unavoidable consequence of the corpus used as a source material for this study. Thankfully, it is still very much possible to investigate the learners correct use of the articles *a*, *an* and *the*. The approach to doing this will be described in section 3.2.2.

In addition to the annotation of the CORYL corpus, the corpus analysis tools which the corpus is equipped with are also necessary for this research. The corpus analysis tools provided by the CORYL corpus are the query, concordance, collocations, distribution and word list function. The tools most important for this specific research are the query page function, which will be described in detail in section 3.2.2, as well as the concordance and word list. The word list function presents the search items in a descending order, presenting the most frequently

occurring word of the search in number and percentage. A concordance is "[…] a comprehensive listing of a given item in a corpus (most often a word or a phrase), also showing its immediate context" (McEnery & Wilson, 2001, p. 197). This enables users of the corpus to view in which contexts errors occur, and moreover, to track possible patterns. The following figure (3.1) shows how the concordancing feature appears in the CORYL corpus when searching for article errors in 7th grade group of learners:
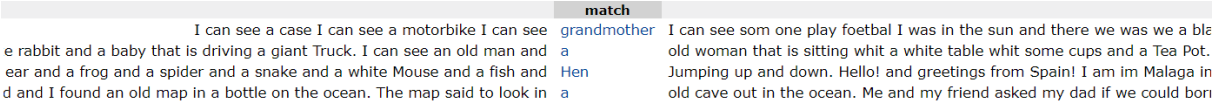
	match	
I can see a case I can see a motorbike I can see	grandmother	I can see som one play foetbal I was in the sun and there we was we a bla
e rabbit and a baby that is driving a giant Truck. I can see an old man and	a	old woman that is sitting whit a white table whit some cups and a Tea Pot.
ear and a frog and a spider and a snake and a white Mouse and a fish and	Hen	Jumping up and down. Hello! and greetings from Spain! I am im Malaga in
d and I found an old map in a bottle on the ocean. The map said to look in	a	old cave out in the ocean. Me and my friend asked my dad if we could bor

Figure 3.1: Concordance of article errors among the 7th grade learners

The search item, where the article error occurs, appears in the middle as a row, surrounded by the context on each side. Also, if one clicks on the *match* (e.g. on 'grandmother' in the first row of the concordance in figure 3.1) one can view additional information about the item in question, as well as view it in the original and full text in which it occurs. This information is provided by the annotators of the corpus. It is also possible to view the error-coded and corrected version of the text. These features are beneficial, as it can be challenging to determine the exact type of errors in some cases. Moreover, it is very useful to be able to visit the original text, and to see the context in which the error occurs. Figure 3.2 shows how the context appears. The available context is which error type the error is tagged with, the text type which the text belongs to, the anonymous ID-number of the learner who has written the text, the country of origin, age and gender of the learner, as well as the CEFR grade, and year of production, of the text. ID, gender, and the CEFR grade were sometimes not included in this context.

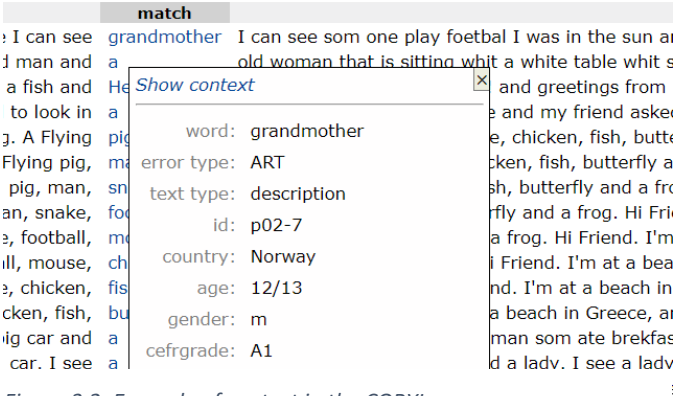|	match			
? I can see	grandmother	I can see som one play foetbal I was in the sun ai		
d man and	a	old woman that is sitting whit a white table whit s		
a fish and	He	**Show context**	☒	and greetings from
to look in	a			e and my friend aske
g. A Flying	pig	word: grandmother		e, chicken, fish, butte
Flying pig,	ma	error type: ART		cken, fish, butterfly a
pig, man,	sn	text type: description		sh, butterfly and a fro
an, snake,	fo	id: p02-7		rfly and a frog. Hi Fri
e, football,	me			a frog. Hi Friend. I'm
ll, mouse,	ch	country: Norway		i Friend. I'm at a bea
e, chicken,	fis	age: 12/13		nd. I'm at a beach in
cken, fish,	bu			a beach in Greece, ai
ig car and	a	gender: m		man som ate brekfas
car. I see	a	cefrgrade: A1		d a lady. I see a lady

Figure 3.2: Example of context in the CORYL corpus

3.2.2 Creating the queries

To retrieve any data or information from the corpus, it is necessary to create and then run queries. For this purpose, CORYL presents a query page which enables researchers to design their own queries, depending on what information they are looking for. The query page allows two ways of creating a query: one can choose to use the graphical interface or to do a manual search. If using the graphical interface, one builds a query based on the search criteria offered. One simply clicks and selects the search criteria relevant for the specific research. The manual search, on the other hand, is not necessarily as user-friendly as the graphical interface. The reason for this is that it requires "some knowledge of the Corpuscle query language" (CORYL, nd). Importantly, manually creating a query can open for a larger number of possibilities than the graphical interface. For the search for the wrong use of articles, the necessary search criteria were available through the graphical interface. Therefore, this was the approach chosen for creating these queries. Below is a figure (3.3) of how this works in the CORYL corpus:
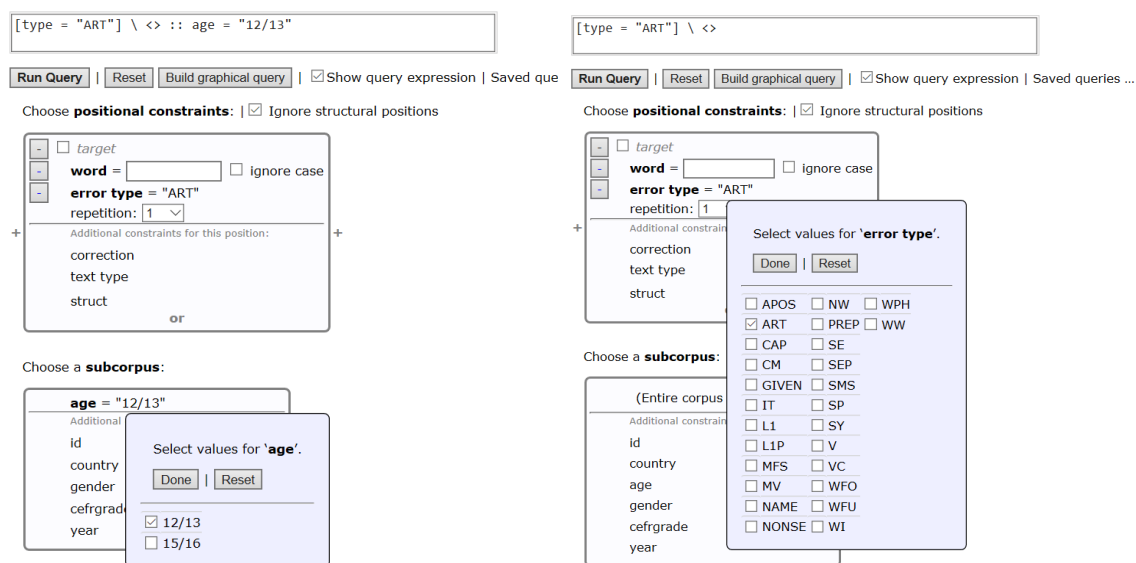
Figure 3.3: Graphical interface appearing in the CORYL corpus

Before explaining the process of creating the queries for this research it would be useful to briefly revisit the research questions posed in this thesis. As stated in section 1.4, this thesis sets out to investigate the following research questions: "How do young Norwegian EFL learners use articles in their writing?", "What errors do the learners make regarding article use, and how many errors do they make compared to the instances where the articles are used correctly?",

and "How does the use of articles in writing develop from grade 7 to grade 10?". These research questions can be broken down into three parts: article errors, correct use of articles and the two groups of learners. It is therefore necessary to create queries that will produce results that include these three parts. By using the graphical interface the corpus allows me to click on error type. Here I choose the error tag ART, which was presented in section 3.2.1. Furthermore, choosing a subcorpus (see figure 3.2) makes it possible to limit the search results. This is useful for this research as it allows for a neat and systematic process of data extraction since one can search for article errors within the two groups of learners separately. These considerations lead to the following queries:

(3.1)   [type = "ART"&!<>] \ <> :: age = "12/13"

(3.2)   [type = "ART"&!<>] \ <> :: age = "15/16"

The query in (3.1) produces results for article errors among the learners in 7th grade, while the query in (3.2) produces results for article errors among the learners in 10th grade. The query in (3.1) yielded 511 results and the query in (3.2) yielded 365 results. However, for reasons which will be explained in section 4.2, these numbers do not represent the final numbers of tokens.

As for the queries used to investigate the learners correct use of articles, the graphical interface was not sufficient. Therefore, it was necessary to create a query manually. For this query, it was necessary to exclude the findings that include articles that are tagged with an ART error. The corpus allows me to specify a search for all the occurrences of *a*, *an* and *the* which do not carry an ART tag. Furthermore, I must choose separate subcorpora to search for hits for the 7th grade learners, and then the same but separate search needed to be run on the texts written by the 10th grade learners. Having done the steps described, it is necessary to exclude the articles *a, an* and *the* that are tagged with errors, as these are included in this initial query. This is done by using another corpus tool: the word list. Here, I click on 'attribute' and choose error type. This list enables me to choose the category [no value]. This category includes all the articles that are not tagged with an error. These queries are as follows:

(3.3)   ["a|an|the"] :: age = "12/13"

(3.4)   ["a|an|the"] :: age = "15/16"

31

Another possible occurrence regarding the correct use of articles in the CORYL corpus, is articles that are used correctly but that are misspelled. In the queries in (3.3) and (3.4) the results produced only provide hits for the articles which are used correctly, and which are also spelled correctly (i.e. *a, an* and *the*). However, many of the learners misspell the articles, as demonstrated in examples (3.5)-(3.7).

(3.5)   And wee come up to **thje** house (no ID)

(3.6)   have you forgot to close **de** dhoor? (p84-7)

(3.7)   ve com not daun but **te** snake com not opp (p44-7)

While misspelled, articles such as those exemplified above cannot be considered article errors: as the examples above demonstrate, the (misspelled) articles function structurally as determiners in noun phrases. Consequently, tokens such as these should also be included in the data material under analysis here. This requires another search and therefore new queries. These queries are as follows:

(3.8)   [corr = "the|a|an" %c&!<>] \ <> :: age = "12/13"

(3.9)   [corr = "the|a|an" %c&!<>] \ <> :: age = "15/16"

The results from these queries are added together with the queries in (3.3) and (3.4), which make up the total of correctly used article among the year 7 and year 10 learners.

3.3 Data handling

Having created the queries and having run them, the data must of course be handled in a manner which facilitates presentation and interpretation. As the results from the queries do not disclose whether the article error is a case of underuse, overuse or wrong article, I had to manually go through every token and determine which of these error types the token represents. These error types then seemed to be fitting categories for systematising the findings. At this stage it was necessary to create a system into which to plot the findings and the long table format was

32

chosen. A long table is suitable because it can hold a great deal of data in a systematic manner. The variables used in the table were the ID of the texts (if possible to extract), the file (sentence) where the error occurs, the type of error (underuse, overuse, wrong article), the correct article in each case, the text type, the gender of the learner, the CEFR-grade (if possible to extract), and the year when the text was written. In cases where further comments were necessary these were added as well. The long table was created in Microsoft Excel. See figure 3.4 below for an excerpt from the long table.

ID 12/13 year[A1] FILE		Type of error	Correct article	TEXT TYPE	GENDER	CEFR-GRADE	YEAR
p02-7	I can see grandmother	Underuse of *a*	a	description	m	A1	2005
p03-7	and a old woman	Wrong article	an	description	m	A2/B1	2005

Figure 3.4: Long table excerpt

The long table excerpt in figure 3.4 is an example of the format which have been used to store and annotate the data concerning the article errors. The data collection of the instances with correct use of article among the learners is a less tedious process, as collecting this data obviously does not require manually going through the findings and sorting them into type of error, and the other categories described above. Consequently, the findings concerning the correct use of articles were not handled in the same manner. This was because the corpus has tools that simplify the process dealing with the investigation of correct use of articles. After running the queries created for this purpose, the results can be output in a table created by the corpus. These results could later be straightforwardly crosstabulated with the erroneous uses.

3.4 Statistics

In chapter 4 two types of statistics will be presented; namely descriptive and inferential statistics. Descriptive statistics "[…] are only specific to the given sample and do *not* allow the drawing of any general conclusions that would go beyond the sample" (Dörnyei & Csizér, 2012, p. 85). The difference between inferential statistic and descriptive statistic is that with inferential statistic "[…] the computer also tests whether the results observed in our sample (e.g., mean differences or correlations) are powerful enough to generalize to the whole population" (Dörnyei & Csizér, 2012, p. 85). If the tests show that the results in fact are powerful enough "[…] we can say that our results are statistically 'significant'" (Dörnyei & Csizér, 2012, p. 85). In this thesis, the descriptive statistics are presented in different forms of

data presentation such as pie charts, bar graphs and tables. The inferential statistic measure used is the *chi-square* test.

The result of collecting data material in the manner described in the previous sections is nominal – or categorical – data. Jenifer Larson-Hall (2012) explains that "[w]hen you have data with two variables, both of which are categorical, and you want to know whether these variables are related, you should use a *chi-square* test to analyse your data" (p. 265). The *chi-square* test is a null hypothesis test used to assess whether associations between the rows and columns in a contingency table are significantly different from one another. Informally stated, then, the *chi-square* test checks whether results between two or more datasets are significant or not. Examples of such data sets in relation to this MA thesis are the findings concerning article errors according to year level and text genre. Larson-Hall (2012) further explains that in order to conduct a *chi-square* one has to enter the data "[…] online in the form of a contingency table, which is a summary of counts of each category" (p.266). The *chi-square* test has been calculated by the means of the online calculator at the website www.vassarstats.net/tab2x2.html. The website calculates both Pearson's *chi-square* and Yate's continuity-corrected version. The measure used here is Pearson's *chi*-squared, uncorrected for continuity. The cut-off point value for statistical significance (i.e. the *alpha* level) adopted in this thesis is 0.05. That means that the *p*-value has to be 0.05 or lower in order for the results to be considered significant. Everything above 0.05 is consequently considered non-significant.

3.5 Implications and limitations

As with any research project, it is important to be aware of the implications and limitations of the methods one chooses to employ, as well as the strengths the choice brings with it. It is obviously important to be aware of these limitations and implications throughout this project, as this will facilitate a correct data collection and data handling. This section will discuss four issues, namely the fact that the research is done without a reference corpus, that it is a small corpus, the question of objectivity, and the use of EA.

Many previous corpus studies use a reference corpus in addition the corpus they are initially investigating. This is usually done for the opportunity of comparing for example learner and native language. Such information can be useful as it can 'place' the learner language in relation to native language. However, as this specific research project does not include a reference corpus, the opportunity to compare learner language to native language does not exist. The

results of this study can nevertheless be relevant, as the study provides a thorough description and analysis of the use of articles among a group of young Norwegian learners of English as a foreign language. This is nevertheless a limitation of the chosen method.

Furthermore, a second limitation regarding this research is – as stated in section 3.1 – that the CORYL corpus is a rather small corpus. This might affect the possibility of extracting a sufficient amount of data. A small corpus consists of a small number of texts which in turn represent only a small group of learners. The consequence of this fact is that the results produced by this research are not generalisable in the statistical sense as there is no way of knowing whether the data are representative for a larger, more diverse group of learners. A larger corpus on the other hand is more likely to present a more diverse group from whom the texts are collected and the results may therefore be more generalisable. Still, it is important to repeat Granger's (2012) statements in section 3.1, about how "[…] the optimal size of a learner corpus depends on the targeted linguistic phenomenon" (p. 9). As articles are very frequent in English, the size of the CORYL corpus was therefore not seen as such an issue that a different corpus had to be considered. Nevertheless, its size can still be viewed as a limitation in this research.

A third possible limitation concerning this research is whether it is objective or not. When dealing with corpora which have already been compiled and annotated by other researchers, the limitation concerning objectivity is that one has to accept the considerations that these annotators have done. In other words, I have had to assume that they have done a correct job in terms of tagging the corpus for article errors and that what they have left as correct use of article is in fact correct. However, I have still made my own considerations in regard to the articles tagged for errors. A small number of these errors was considered not to be errors and have therefore not been included in the data material, which is thoroughly described in section 4.2 in the following chapter. It would be a massive and extremely time-consuming task to check whether the articles that are *not* tagged for errors are in fact correct. It is therefore assumed that the corpus annotators have done this correctly. When working with a corpus that has been annotated by others this is an inescapable factor. It is then important to be aware of possible erroneous error tags and to be able to make independent judgments whether all errors are in fact errors.

Lastly, it is a limitation to heavily rely on an error analysis as a method for this research. As EA only focuses on errors, in this case errors made by young Norwegian learners of EFL, it does not present a full 'picture' of the state of which their learner language is in. As research using only EA fails to include anything about the features the learners can do correctly, it also

paints a rather negative picture of the learner language. Nevertheless, as the CORYL corpus is only tagged for errors, it became natural to base the method on error analysis. Furthermore, since there are not many corpora that deals with the language of young Norwegian learners of English, the options were limited as to which direction this research could take. However, fortunately, the possibility of investigating the learners correct use of articles does exist, as stated in section 3.2.1 – not possible to investigate the correct use of the zero article. Thus, the aim is that this contribution to the data collection might be a source of balance to the EA that is carried out. Hopefully, this will present the article use among young Norwegian learners of English in a more nuanced manner.

3.6 Summary

This thesis is based on corpus linguistic methods, such as data extraction and – specifically for error analysis – description and explanation of errors. By using these methods the aim is to investigate learner language in terms of article use, both correct and incorrect, among young Norwegian learners of English. Section 3.1 presented the CORYL corpus and the reason for choosing this specific corpus was explained. Section 3.2 discuss the data extraction and the tools used in this work, while section 3.3 presents and explains how the data was handled. Section 3.4 includes a brief presentation of the statistical features and test used which are relevant for this research. Section 3.5 discussed implications and limitations of the methods employed in this MA thesis.

Chapter 4: Findings and analysis

4.1 Introduction

As noted in section 1.4, this thesis aims to investigate the use of articles among young Norwegian learners of English as a Foreign Language. Thus, the investigation carried out here is a study of errors made in Norwegian EFL learners' use of articles, and it contrasts errors with the corresponding cases of correct use of articles. Additionally, the aim is to investigate the development in the use of articles from 7th grade to 10th grade among the learners whose texts appear in the CORYL corpus. This chapter will therefore present quantified overviews of correct and incorrect uses of the different articles, followed by analyses of these data.

The chapter is structured as follows: first, section 4.2 briefly presents and discusses exclusions and practical limitations regarding the findings in this chapter. Section 4.3 subsequently presents and analyses findings about errors and correct use of the definite article, section 4.4 presents and analyses findings concerning errors and correct use of the indefinite article, and section 4.5 presents and analyses findings dealing with errors made concerning the zero article. Section 4.6 section deals with the development in the use of articles from 7th to 10th grade and considers both the correct and incorrect use of articles. Section 4.7 summarize the findings and concludes the chapter.

4.2 Exclusions and practical limitations

Before presenting the findings, it is necessary to explain some limitations that affected the data collection, and in turn the findings. Firstly, the two queries searching for article errors produced 876 hits in CORYL in total. However, only 734 hits were part of the final data material. The reason for this is that some sentences contain examples of ART errors tagged on two words which clearly are part of the same noun phrase. An example of this is the following:

(4.1)   I can see **lazy guy** hvo is eating bananas (p152-07).

The noun phrase in question is *lazy guy* and the article error is the missing indefinite article *a* at the left edge of the phrase. The correct structure is then *a lazy guy*. The corpus marks both *lazy* and *guy* as containing ART errors, meaning that one article error is counted as two by the corpus. As this is an annotational error, only one error has been counted in cases like these.

Another issue that affected the results was that some of the errors that appear in the search results turn out not to be article errors after all. An example of this is given below:

(4.2)   but A Aslak had see a sneak (no ID)

By looking at the context where this sentence occurs it seems more likely that the *A* is a typographical error, rather than an article error. The text in which this error occurs have several phrases which include the proper noun *Aslak.* In these cases the learner successfully omits the article, as this noun does not take either the definite or the indefinite article. Therefore, examples like these were the annotators have applied an article error tag to words that are not in fact an article error have been excluded from the data material.

Furthermore, it must also be noted that in the process of collecting the data after conducting a query, it is necessary to do some interpretation. Some of the results consist of a great deal of what the CORYL corpus labels as *nonsense* and L1 words. Without any further explanation than the error tags, it is therefore necessary that interpretations are made. Also, sentences that are especially difficult to interpret and where it is unclear what the error is, due to poor sentence structure, have been excluded from the results. Therefore, had this study been replicated there is a chance that the results would vary to a minor degree due to these exclusions.

Also, note that the examples presented throughout this chapter are presented without context. Therefore, there may be sentences that appear to be correct, but which are in fact incorrect in their original contexts. The following sentence in an example of such a case:

(4.3)   '*If you take **a** bus*' (p250-10)[1]

The noun phrase *a bus* has been tagged as containing an article error by the CORYL annotators. Now, if the learner had been discussing which bus to take in order to be in time for something, as in e.g. '*If you take a bus leaving around 5 or 6 o'clock, you should make it',* the use of the indefinite article would have been correct. However, in this instance, the learner wrote '*It's about 15 min. from Bergen centre if you take a bus, maby less if you drive by your self'* (p250-10). The correct article is therefore the definite article *the*, and the corpus annotators correctly

---

[1] Example (4.3), out of context, is grammatically correct. However, in the context in which it appears, it is incorrect.

specify that the noun phrase *a bus* features an article error. When presented out of context, however, as the case is in example (4.3), the fact that an error is present may not be immediately obvious.

Furthermore, the most unfortunate limitation of this thesis and its findings is the difficulties involved in investigating the correct use of the zero article in the CORYL corpus, as briefly explained in section 3.2.1. For the analysis concerning the development in the use of articles from grade 7 to grade 10, the error analysis includes the errors made regarding overuse and underuse of the zero article. However, the correct uses of the zero article are not retrievable from the corpus. Therefore, when using these numbers combined to investigate the development in article use, the data material is missing quantifications concerning the correct use of the zero article. Since the CORYL corpus is only tagged for errors and not lemmatized or organized as a treebank, it does not have the structure necessary to investigate the correct use of the zero article through finding correct noun phrases with correct use of the zero article. In other words, the corpus does not enable searches for invisible grammatical features, such as the zero article. This is very unfortunate and an inescapable factor, and it is of course a weakness of the quantitative material presented below. Since it is only possible to retrieve quantifications on the correct use of the articles *a, an* and *the,* the data material is evidently lacking an important part. Nevertheless, as there is not an abundant selection of corpora that consist of learner language produced by Norwegian learners in English, choosing the CORYL corpus was the most optimal solution, despite the limitation of the corpus. A possible solution to this problem, however, would have been to leave incorrect uses of the zero article out from consideration. By doing this the data material would have been equal, and the focus would only be on the correct and incorrect use of the articles *a, an* and *the.* This was considered, but the incorrect use of the zero article was included nevertheless since it still illuminates an area of errors that is also deserving of investigation.

A final general comment concerning the difference between the two groups of learners that have been investigated is also warranted before the findings are presented. The texts produced by the 7th grade learners are of different text types than the texts produced by the 10th grade learners. The texts written by the 7th grade learners are letters, stories or descriptions. The texts written by the 10th grade learners, on the other hand, belong to the genres essay, letter to the editor and personal letter. The tasks given to the learners [see appendix A, B – S.L] are different, and since the different genres open up for different ways of writing, the use of articles is affected

as a consequence. An example of this is a description task (Appendix A) given to the 7th grade learners. Briefly explained, it asks the learners to list what they see in a given picture. A natural way of wording oneself in such a task would be 'I see a cat. I see a dog. I see an orange ball'. Therefore, this task – as compared to an essay or a letter to the editor – affects the number of articles used and consequently – also the number of correct and incorrect uses of articles. A letter often has a more 'natural' distribution of articles, contrary to a task that asks the learners to list all the nouns they see in a picture. Therefore, it is difficult to compare the two groups on an equal basis. However, when using an already compiled corpus, there are limitations as to what one can do to change or avoid such features as this one.

4.3 The definite article

The results of the data collection make it evident that both groups of learners make a considerable number of errors in their use of the definite article. This section aims to answer the first and parts of the second research questions posed in section 1.4, namely: "How do young Norwegian EFL learners use articles in their writing?" and "What errors do the learners make concerning article use?" Sections 4.3.1–4.3.3 present the findings concerning the learners overuse, underuse and wrong use of the definite article in the CORYL corpus. As is well known, the definite article *the* in English differs significantly from the definite article in Norwegian. Contrary to the situation in English, the definite article in Norwegian is formed via the inflectional endings *-et, -en, -a, -ene.* As a consequence of this, errors and confusion may occur for the learners when applying the definite article in their writing.

4.3.1 Overuse

As noted in section 3.2.1, overuse of an article, in this case the definite article, constitutes cases where the learner uses an article in a case where the zero article should have been used. Below are given four examples (4.4–4.7) of such errors compiled from the search for article errors in the CORYL corpus:

(4.4)   Dear John I am in **the** Miami (p11-7)

(4.5)   Elvis Presley make **the** music come to the world (p175-10)

(4.6)   she had problems with **the** breathing [2](p122-10)

(4.7)   One day Jack and I came home from **the** school (p213-7)

Example (4.4) shows a recurring error in the overuse of the definite article. The learners tend to use the definite article in front of proper nouns, such as 'Miami'. Other examples of such use are 'In the New York' (p50-7), 'you simply must see the Holmenkollen' (p105-05) and 'You most see the Galdhøpiggen' (p201-10). As discussed in section 2.1.6, which deals with transfer from Norwegian, errors in the L2 may occur due to influence from the L1. However, this specific error is not necessarily caused by transfer, as neither Miami nor New York are proper nouns that require an article in Norwegian. Galdhøpiggen and Holmenkollen do not require an article in Norwegian either, but as one can see, those two proper nouns end in *-en.* As stated in section 2.1.5, the definite article in Norwegian can take this form. It is therefore possible that the learners behind these errors believed that the *-en* ending of the two proper nouns, Galdhøpiggen and Holmenkollen, was in fact the definite article in Norwegian, rather than a part of the name of the places.[3]

(4.5) showcases another example of overuse of the definite article. In this case, the learner has written a text about Elvis Presley. Based on the corpus having tagged this article as an error, as well as the context it appears in, it is reasonable to believe that the learner in fact is referring to *music* as an uncountable noun. Therefore, the correct article is the zero article. Example (4.6) can be considered as a grammatically correct sentence when taking the definite article if the word *breathing* is acting like a noun. However, in this context *breathing* is acting like a verb. Therefore, using the definite article is overused in this sentence, and the zero article would be correct.

Additionally, as example (4.7) shows, the learner in this case has used the definite article *the* in front of the noun *school*, which is incorrect in this context. *School* acts like a mass noun referring to school as an institution where teaching happens, as opposed to referring to a specific school or school building, which would have required an article. Therefore, it is interesting to see that there were 23 instances where the learners in both 7th and 10th grade made this specific

---

[2] Example (4.6) can be construed as grammatically correct, if *breathing* is acting like a noun and not a verb. In this context, however, *breathing* acts like a verb, and use of an article is therefore ungrammatical.

[3] Note that the *-en* ending in Galdhøpiggen and Holmenkollen most definitely must have been a definite article historically.

error. This error might be due to the fact that in Norwegian *school* takes the definite article ending 'skol**en**' and the error may therefore be caused by interference from Norwegian. Still, this is merely a guess, as it is not possible to conclude on the learners' choice of article based on the information provided in the corpus. It is however still important and useful to be aware of such errors that might indicate a pattern in errors about overuse of the definite article as shown in example (4.7).

As concerns overuse of the definite article, the findings show that the 7th grade learners made a total of 47 errors of this type. This group wrote texts in three different genres, these being description, letter and story. Eight of these errors were made in texts in the genre of description, eight errors occurred in the letters, while a total of 31 errors of overuse of the definite article occurred in the stories. As for the 10th grade learners, the findings showed that they made a total of 95 errors of overuse of the definite article. The genres of the texts by this group were essay, letter to the editor and personal letter. Seven of the errors occurred in the personal letters, while 52 errors occurred in letters to the editor and 36 errors occurred in the essays.

4.3.2 Underuse

Underuse of the definite article compromises cases where the learners leave out the definite article when it is correct to include it, as also stated in section 3.2.1. In (4.8)–(4-11), four examples of such errors from the CORYL corpus are presented.

(4.8)   Next day I Woke up of a nois (p168-07)

(4.9)   Yesterday we saw Tower of London (p29-7)

(4.10)   So listet we us in to house (p58-7)

(4.11)   taking care of environment (p120-10)

An interesting fact that became apparent through the data collection was how many learners wrongly left out the definite article in the phrase 'the next day/morning', as shown by example (4.8) above. The data material shows that this happened 24 times among the 7th grade learners. Many of these sentences can be translated into Norwegian in the following manner: *Dagen etter* or *Morgenen etter*. Those translations include the suffixed Norwegian definite article *-en*.

Therefore, the omission of the English definite article *the* in these sentences might be because the Norwegian noun *dagen* does not have a prefixed article. However, as discussed in section 4.3.1., the learners seem to overuse the definite article together with *school*, which in Norwegian has the same definite article as *dagen*. A question that can be raised is then whether some of these mistakes are simply just arbitrary – that is, simple performance errors rather than indicative of inadequate acquisition. The cases in question belonged to the genre of story. Another interesting finding is that the majority of underuse errors of the definite article occur in this genre. These errors made up a total of 70 out of 89 tokens, i.e. 78.7% of the cases of underuse of the definite article among the 7th grade learners. A point for the analysis and discussion in this chapter is therefore whether genre is a factor which potentially influences or facilitates article use errors.

Example (4.9) show a case where a learner has incorrectly left out the definite article in front of the proper noun *Tower of London*. This error is most likely due to transfer from Norwegian. If one is to say this sentence in Norwegian one would say *I går så vi Tower of London*. The learner has then probably just assumed that *Tower of London* does not require a definite article. Furthermore, in Norwegian, example (4.10) and (4.11) take on the definite ending *-et.* The nouns in these examples require the definite article *the.* The learners behind these examples have therefore most likely made the same assumption as done in example (4.9), with the definite article in Norwegian being a part of the noun, and then excluded the definite article.

The findings showed that the 7th grade learners made a total of 89 errors of underuse of the definite article. 70 of these errors occurred in the texts written as stories, while 13 of the errors occurred in the letters, and 6 of the errors occurred in the texts in the description genre. The 10th grade learners made 82 errors of underuse of the definite article. 37 of these errors occurred in their essays, 32 of the errors occurred in letters to the editor, while the last 13 errors occurred in personal letters.

4.3.3 Wrong use

Wrong use of the definite article is when the learners have used the definite article incorrectly and where the indefinite article would be the correct article. Examples of such use are presented below.

(4.12) should not get **the** driving license (p05-10)

(4.13) I found the front door wide open and **the** feel of emptiness inside the house seemed to be coming out. (p150-07)

(4.14) and a man sitting in **the** cher (p155-07)

(4.15) free like **the** bird (186-10)

Example (4.12) shows a case where the learner has used the definite article incorrectly. The correct article in this sentence would be the indefinite article *a.* This example is collected from a text written as a letter to the editor and the learner is discussing who should be allowed to have a driving license. The noun, *driving license,* is a countable noun but used in the singular in this context, and as presented in section 2.1.3, the indefinite article is used together with nouns such as this. This is also the case with (4.13), where the learner is describing how ***a*** feeling of emptiness was coming from inside a house. In this context, the countable noun *feeling* is used in the singular, which indicates that the correct article is the indefinite article *a*. Example (4.14) is gathered from an answer to a description task and the learner was to describe what he or she saw in a picture. The learner is therefore referring to something that the reader is not familiar with, or which is unknown for the learner. This indicates that the correct article to employ is the indefinite article, as the indefinite article refers to something that is unknown to the reader. Therefore, when choosing to use the definite article in this context the learner has made an error. (4.15) is an example of a fixed expression often used. However, it is not about a specific bird and the reader is not familiar with this bird. Therefore, it is correct to use the indefinite article *a,* and not the definite article. It is also worth mentioning that in Norwegian this expression is *fri som fugl**en***. The error in this example may be caused by interference from Norwegian, or by confusion of the rules, as the noun takes on the definite article ending *-en* in Norwegian. However, if the learner has failed to acquire this idiomatic expression as a lexical chunk, this might also lead to the error being made.

In the group of 7th grade learners there were 29 cases where the learners have used the definite article incorrectly. 14 of these errors occurred in description, 11 of the errors occurred in stories, while the remaining 4 errors occurred in letters. Among the 10th grade learners there were 12 cases where the pupils have used the definite article incorrectly. 5 of the errors occurred in

essays, 3 of the errors occurred in letters to the editor, and the last 5 errors occurred in personal letters

4.3.4 Overview of findings

The findings regarding the different types of wrong use of the definite article can be presented as in figure 4.1. This figure shows the distribution of the uses of the definite article which are tagged as errors in the CORYL corpus.
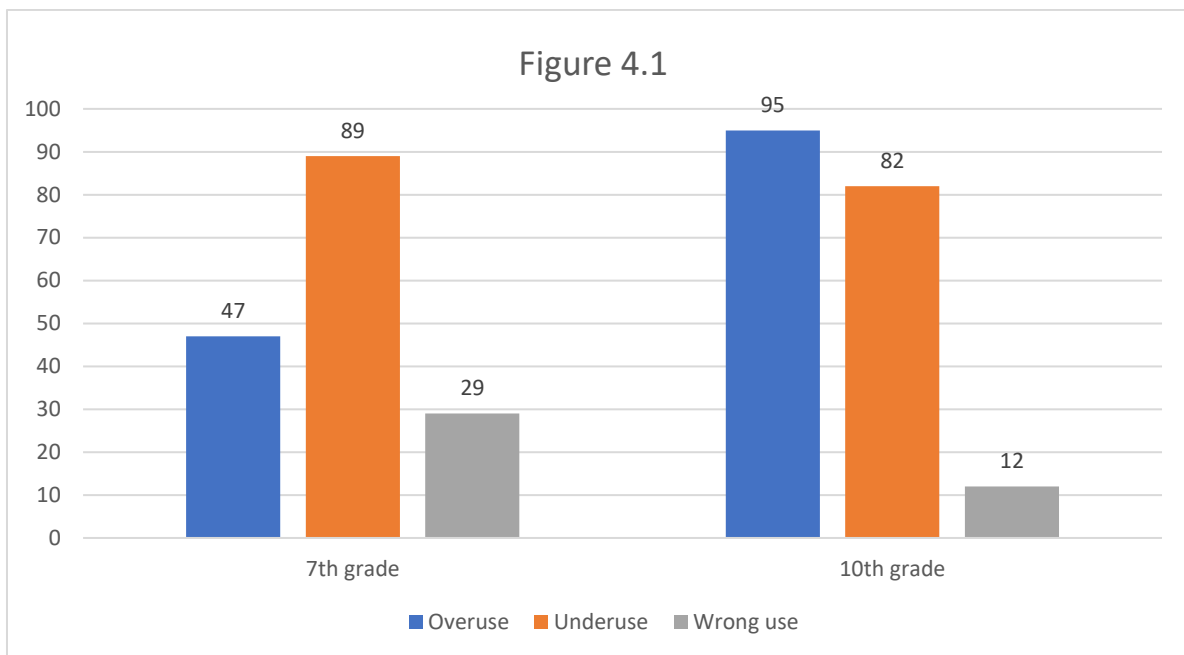

Figure 4.1

Figure 4.1: Numbers of overuse, underuse and wrong use of articles among 7th grade and 10th grade learners

The 7th grade learners made a total of 165 errors concerning the definite article, while the 10th grade learners made a total of 184 errors concerning the definite article. The most prominent observation is the difference between overuse of the definite article in 7th and 10th grade. The 10th grade learners produced 95 errors of overuse of the definite article, which is twice the number of errors produced by the 7th grade learners. There is also a rather noticeable difference between the two groups of learners concerning the wrong use of the definite article – here, the 7th grade learners can be observed to make more than four times the number of errors of the 10th grade learners. However, the number of underuse of the definite article is almost the same: the 7th grade learners produced 7 more errors than the 10th grade learners.
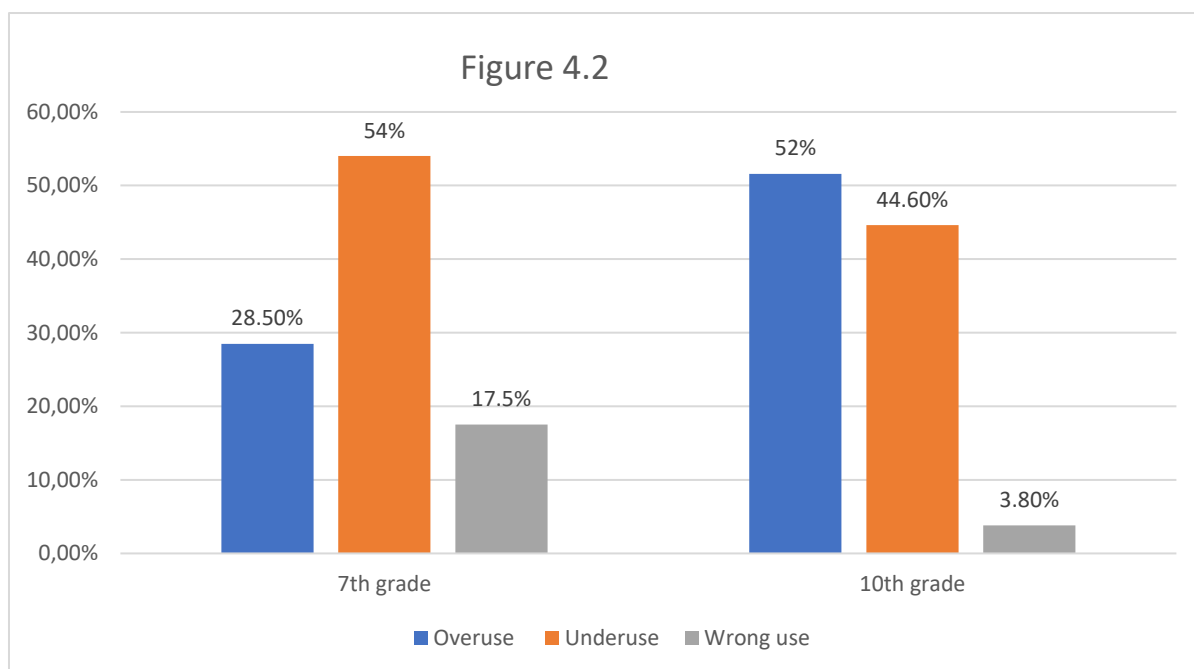
Figure 4.2

Figure 4.2: Percentage of the total number of errors made concerning the definite article within each group of learners

In this figure, the results have been presented as a percentage of the total number of errors made within each group of learners. As such, it presents a clearer picture of which errors dominate within each group. As both figure 4.1 and 4.2 show, the 10th grade learners make the most errors of overuse of the definite article, and as figure 4.2 shows, the 10th grade learners' errors concerning overuse of the definite article make up 51.6% of the errors of the definite article, opposed to the corresponding figure for the 7th grade learners errors which is 28.5 % of the errors in that group. The numbers concerning underuse of the definite article are the most similar for both groups, with a difference of only 9.4%. The discrepancy between the groups in terms of wrong use of the definite article was also rather small, with 13.7% separating the 7th graders and the 10th graders.

Since the CORYL corpus does not contain or disclose any information regarding how the learners have been instructed to apply articles or the proficiency level of the learners, it is somewhat challenging to consider why the learners make different errors. However, there is one feature in particular that is interesting to investigate in this context, which the corpus actually does disclose information about. A possible reason for the rather large discrepancy between the number of errors of overuse of the definite article among the two groups of learners might be linked to task and text type. The following figure shows the number of errors concerning overuse of the definite article in the text type in which it occurred. This error type

is highlighted since it was the error that occurred the most frequently in individual cases, but also because the findings were somewhat unexpected as the 10th grade learners made significantly more errors of this type than the 7th graders. The number of errors within each text type has been normalized by 100 as a percentage of the total number of errors within each group of learners, and text type.
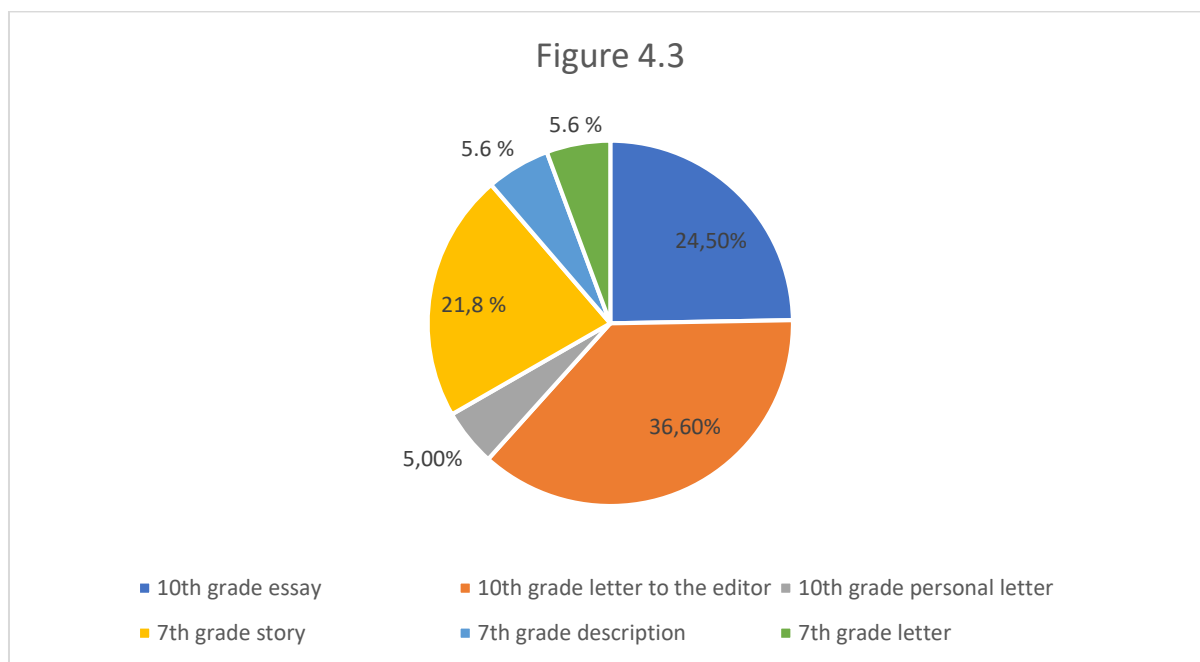

Figure 4.3: Percentage of the overuse of the definite article within each group of learners and text genre

The text type in which there occurred the most errors of overuse of the definite article was 10th grade letters to the editor, making up 36.6% of the errors. The task given to the learners was the following:

"Young people today care less about the environment than adults".

This headline appeared in your local newspaper recently. Write a response of 1 or 2 paragraphs to this editorial, providing arguments that will really convince the readers of your point of view.

(National tests in English 10th grade, 2005, Appendix B)

The learners then write about the same topic and include many of the same nouns in their texts. In line with the task nouns like *young people* and *adults* recur frequently in the texts. These nouns are plural nouns and the learners are referring to something in general in this context

when they write about young people and adults. Therefore, the correct article here is the zero article. An interesting consideration is therefore what the possible outcome could have been if the task had promoted wider use of other nouns – nouns that, for example, require the definite article. It is important to keep in mind that the design of the tasks might affect the learners use of articles, and in turn the number and type of errors they make.

The text type essay generated 24.5% of the errors among the 10th grade learners, while the 7th graders produce 21.8% of the errors regarding overuse of the definite article in the text type story. It is interesting to note that it was the 10th graders who produced the most errors, both percentagewise and in absolute numbers, concerning overuse of the definite article. The text types 7th grade letters and the 7th grade descriptions both generated 5.6% each of the errors, while the text type personal letter written by the 10th grade learners generated 5% of the errors. Due to how low these final numbers are, finding a possible reason for these errors will turn into merely speculating and guessing, since there are so few examples of the errors. Therefore, the focus lies on the text type that generated the most errors, namely the 10th graders letter to the editor task.

4.4 The indefinite article

As with section 4.3, this section also aims to answer the first and parts of the second research question posed in section 1.4. The indefinite article also turned out to be a source of error for both groups of learners in this corpus study. In English the indefinite article is *a/an,* depending on the sound following the article. However, in Norwegian, the indefinite article is either *en, ei, et,* depending on the gender of the noun following the article. The following presentation of the results regarding the different errors of the indefinite article will be divided into the two types of article; *a* and *an.*

4.4.1 Overuse

Before presenting the numerical findings it can be useful to view some examples of the errors that will be presented and analysed. Overuse of the indefinite article comprises cases where the learners have used either *a* or *an* in instances where the article should have been left out. (4.16)–(4.19) provide four examples collected from the CORYL corpus that exemplify overuse of the indefinite article.

(4.16)   We drink **a** te and spiste pizza (p156-07)

(4.17)   I can see a lady with **a** curly brown hair (p293-07)

(4.18)   He make **a** sochs ower the wole world (p201-10)

(4.19) I saw my boyfriend kissing **an** another girl (p275-10)

As stated in section 2.1.3, the indefinite article is used only in the singular, and it often occurs with countable nouns (Bækken, 2002, p.110). Example (4.17) contains an uncountable noun, 'curly brown hair'. Based on the context in which this example occurred, it was evident that the learner was not describing that the lady had *one* single curly brown strand of hair, but rather a head of curly brown hair. Therefore, this article error is categorized as overuse of the indefinite article, as the correct article is the zero article. Additionally, example (4.18) is an error similar to example (4.17). The noun, *sochs (socks)*, is a plural noun. Furthermore, the learner is here referring to *socks* in general. These features are not compatible with the use of the indefinite article, nor the definite article, and the correct article is therefore the zero article. This is also the case with example (4.16). The learner is referring to *tea* in general, and since *tea* is an uncountable noun, the correct article is the zero article. It is also important to note that the learners do not make a great number of errors regarding overuse of the indefinite article, so it would be incorrect to state that this is a prevalent problem.

(4.19) is one of those cases (as in 4.2) where one must consider whether the learner has made an article error or whether it is a typographical error since the article is so similar to the following word. However, it is more likely that (4.19) is an actual article error than the case with (4.2). The determiner *another* is actually a combined word which consists of the indefinite article *an* followed by the determiner *other*. Therefore, the learner overuses the indefinite article *an* when it is put in front of the determiner *another*. The correct use of article here is therefore the zero article. This error has most likely occurred due to transfer from Norwegian. The Norwegian way of saying *another* is ***en annen**.* In Norwegian, when referring to *another*, two words are used, and one of these words is the indefinite article in front of the following determiner. Therefore, is it reasonable to believe that the learner did not take into consideration the fact that *another* is a combined word and has therefore used the separate indefinite article *an* in addition to the determiner *another.*

There were 29 errors of overuse of the indefinite article *a* among the 7th grade learners. 18 of these errors occurred in descriptive texts, 6 errors occurred in the letters, while the last 5 errors occurred in stories. There were 2 errors of overuse of the indefinite article *an.* There was one error each in the genre description and letter. In the group of 10th grade learners, there were 13 errors of overuse of the indefinite article *a*. 8 of these errors occurred in essays, while the 5 remaining errors occurred in letters to the editor. There was 1 error of overuse of the indefinite article *an*, and it occurred in an essay.

4.4.2 Underuse

Underuse of the indefinite article involves instances where the learners have left out the indefinite article where it would have been correct to include it. Examples of such errors from the CORYL corpus are presented below.

(4.20)   They have made fier (p07-7)

(4.21)   I Can see lazy guy hvo is eating bananas (p152-07)

(4.22)   Greetings from angry girl (p285- 10)

(4.23)   And we have slalom hill (p86-10)

Example (4.22) shows a noun in the singular: *angry girl.* In the context in which this article error occurs, the learner is referring to a specific girl. Therefore, the correct article in this case is the indefinite article *a*. In Norwegian one could write *hilsen **en** sint jente* or *hilsen sint jente*. The latter example is the equivalent to what the learner has written in English. This error may therefore be caused by interference from Norwegian. In other words, the Norwegian construction may have transferred into the English rendition.

A very frequently occurring feature regarding the underuse of the indefinite article seemed to be related to the tasks the learners had been given. The 7th grade learners received a task which asked them to look at a picture and to explain what they saw, in writing. This task led to many sentences that started with *I can see…,* followed by a description of what they saw. Therefore, the learners used the indefinite article a great deal, both correctly and incorrectly. The errors concerning the indefinite article and the underuse of this mostly occurred in these sentences,

where the learners left the article out, such as in example (4.21). The learner who has written this sentence has incorrectly left out the indefinite article *a* from the noun phrase *lazy guy*. However, it is also interesting to note that the learner has correctly left out the indefinite article preceding the noun *bananas.*

As stated in section 2.1.4, the zero article is used "[…] before plural and uncountable nouns when we are referring to something in general" (Parrott, 2010, p. 28). However, in examples (4.20) and (4.23), the article has been left out even though the learners are not referring to something in general. These two examples show references to something that is unknown to the reader. Furthermore, the examples include nouns in the singular. It is therefore clear that the learners have incorrectly left out the indefinite article in these sentences.

The group of 12/13-year-old learners made a total of 172 errors of underuse of the indefinite article *a*. 14 of these errors occurred in letters, while 33 errors occurred in stories. A total of 125 of these errors occurred in descriptive texts. There were 4 errors of underuse of the indefinite article *an* in the same group of learners. 2 of the errors occurred in stories, while 2 errors occurred in descriptions. Among the 15/16-year-old learners there was 40 errors of underuse of the indefinite article *a*. 16 of these errors occurred in essays, 13 occurred in letters to the editor, while the last 13 errors occurred in personal letters. As for the indefinite article *an*, the same group of learners made 2 errors. One of which occurred in an essay, while the other error occurred in a letter to the editor.

4.4.3 Wrong use

Wrong use denotes instances where the learners used the indefinite article incorrectly, which makes the definite article the correct choice. Examples of such errors from the CORYL corpus are presented below.

(4.24)   and I found **a** front door open (p231-07)

(4.25)   If you take **a** bus (p250-10)

(4.26)   I eat my pizza wend I hear **a** sound of a snake (p92-7)

(4.27)   and we'll be **a** race known as (p193-10)

Compared to the underuse and overuse of the indefinite article, the number of wrong uses was considerably lower. (4.24)–(4.27) all show how the learners use the indefinite article when referring to something that is familiar to the reader, and about something specific. The definite article is therefore the correct article in these sentences. Furthermore, it is interesting to note that the definite article is the correct article in Norwegian as well. All of the nouns in examples (4.24)–(4.27) have the definite ending *-en* in Norwegian; *inngangsdøren, bussen, lyden, rasen*. It can therefore be questioned what might actually cause these errors, as none of these nouns take an article that precedes the noun in the sentence. Transfer from Norwegian does not seem to be the reason for the errors.

There were eight cases where the 7th grade learners used the indefinite article incorrectly, where the definite article is correct. One of the errors occurred in description texts and seven errors occurred in stories. Among the 10th grade learners there was a total of seven cases where they have used the indefinite article incorrectly, where the definite article was correct. Three of these errors were found in essays, while the remaining 4 errors occurred in letters to the editor.

4.4.4 Overview of findings

Figure 4.4 is a representation of the distribution of the errors of the indefinite article, whether overuse, underuse or wrong use, for the two age groups under investigation. Looking at the figure it is evident that the number of errors regarding the indefinite article is noticeably higher among the 7th grade learners than among the 10th grade learners. This goes for all the three different categories of uses investigated in this thesis; overuse, underuse and wrong use.
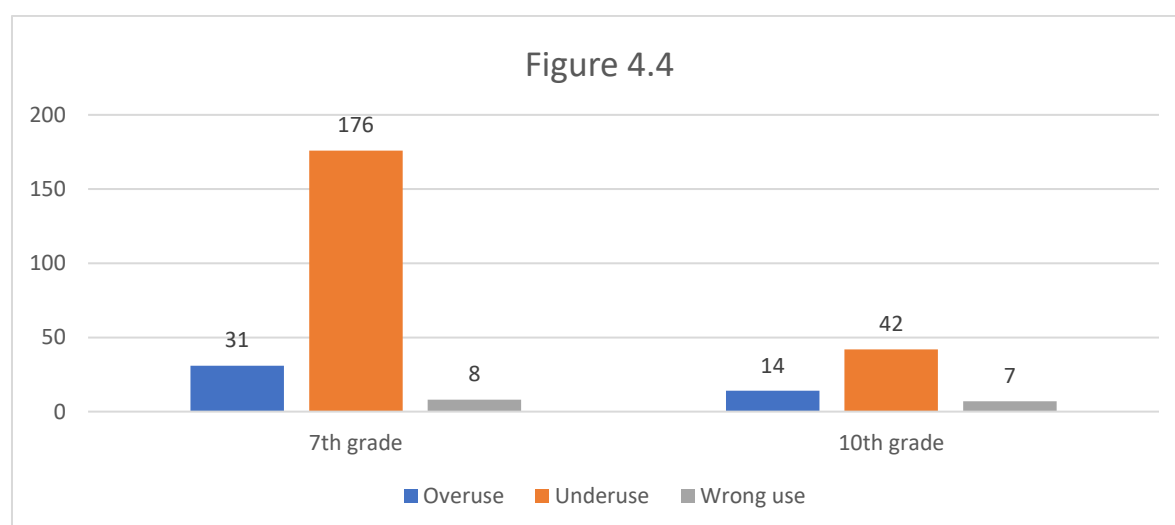

Figure 4.4: Numbers of overuse, underuse and wrong use of articles among 7th grade and 10th grade learners

The 7th grade learners made a total of 215 errors concerning the indefinite article, where a total of 176 of these errors was underuse. The 10th grade learners made a total of 63 errors concerning the indefinite article. 42 of these instances was underuse of the indefinite article. Overall, looking at the observed number of occurrences concerning errors, the 7th grade learners clearly made most errors. These findings are in line with the findings from Master (1997) in section 2.2, namely that the indefinite article *a* "(…) began to emerge only after *the* was largely mastered" (p. 218). It is therefore useful to consider Master's (1997) notion in relation to these findings, as there seem to be some sort of similarities. One can imagine that the number of errors regarding underuse of the indefinite article is lower among the 10th graders than the 7th graders, due to the 10th graders being more proficient in the use of the indefinite article. It is still interesting to consider Master's (1997, p. 218) findings, as the number of individual cases of underuse of the indefinite article was so different from the 7th graders to the 10th graders. However, when normalizing the findings, the presentation of the findings changes somewhat, which can be seen in the following figure.
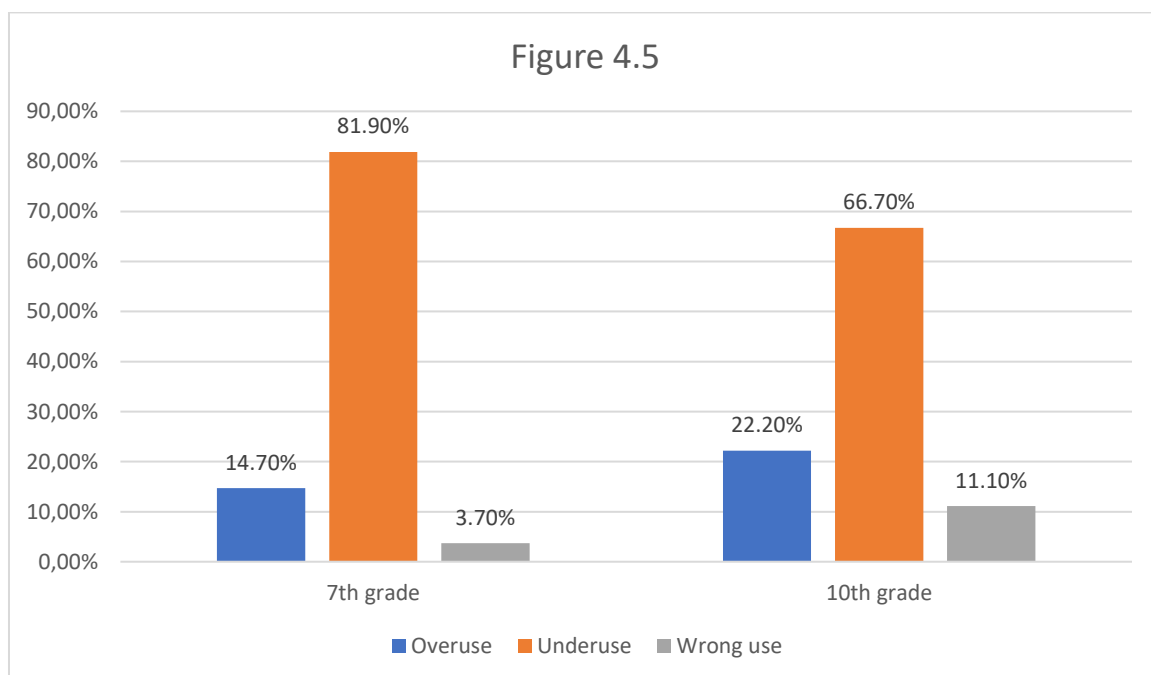

Figure 4.5: Percentage of the total number of errors made concerning the indefinite article within each group of learners

In this figure, the errors have been normalized by 100, as a percentage of the total number of errors within each group of learners. By doing this, it presents a clearer picture of which errors dominate within each group. The 10th grade learners made fewer errors in all the categories in total. However, when looking at the normalized findings the 10th grade learners made more

errors percentagewise concerning both overuse and wrong use of the indefinite article. The 7th grade learners made a much higher number of errors in terms of underuse of the indefinite article and this therefore affects the normalized findings. These findings show that 81.9% of the total number of errors were underuse errors in the year 7 group, and that that figure is 66.7% in the year 10 group. Furthermore, the 10th grade learners' errors of overuse of the indefinite article make up 22.2% of their total amount of errors of the indefinite article, while the 7th grade learners' errors of overuse make up 14.7% of their total number of errors of the indefinite article.

As noted for the erroneous use of the definite article in section 4.4.2, the reason for the rather large number of errors regarding underuse of the indefinite article among the 7th grade learners might be affected by the task given to the learners. It is therefore interesting to take the different text types into consideration to see whether the underuse of the indefinite article is more prevalent in one text type compared to another. The following figure shows the distribution of the errors involving underuse of the indefinite article according to the text type in which they occurred.

Figure 4.6

- 7th grade: Story
- 7th grade: Description
- 7th grade: Letter
- 10th grade: Essay
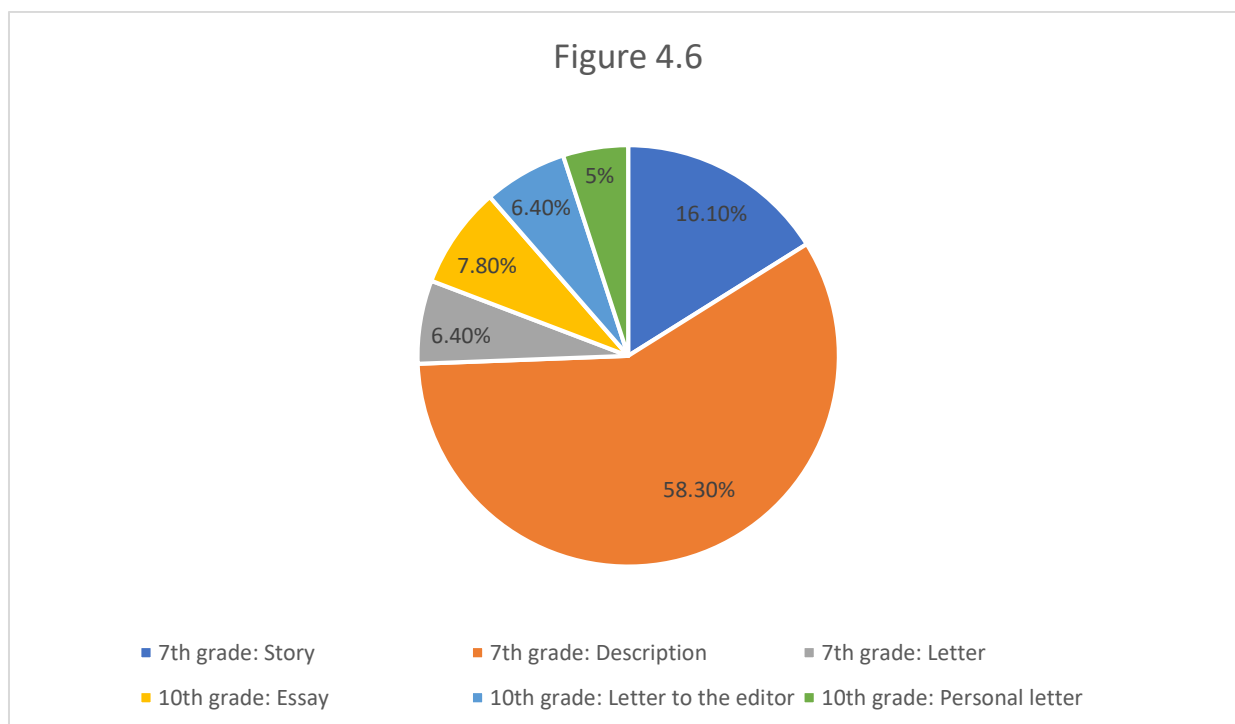- 10th grade: Letter to the editor
- 10th grade: Personal letter

Figure 4.6: Percentage of the underuse of the indefinite article within each group of learners and text genre

Section 4.4.2 briefly discussed the possibility of the task given to the 7th grade learners having affected their underuse of the indefinite article. Figure 4.6 presents the distribution of errors

regarding underuse of the indefinite article and text type. It is evident that the text type description (7th grade) contained the majority of these errors, making up 58.3% of all the errors of underuse of the indefinite article. The description task given to the learners included the following picture and the following text; "Look at the picture. What do you see?".
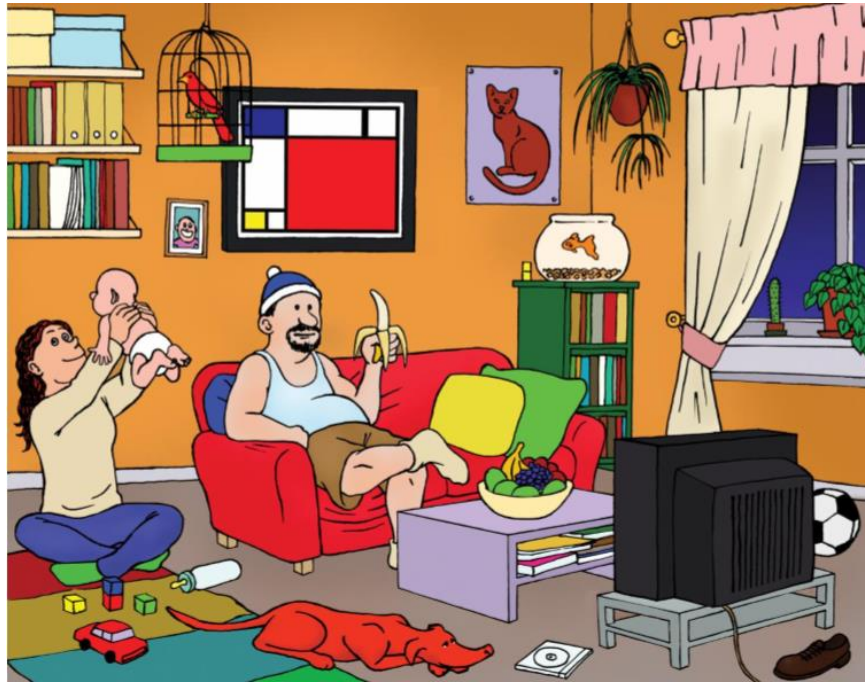

Figure 4.7: Task from National Tests in English, 7th grade (Appendix A)

The task leaves little room for other than listing the things one can see. Most of the learners therefore begin their sentences with *I can see….* There is some variation as to how the learners then proceed with describing what they see. Some of them choose to write *I can see* numerous times, while others choose to write *I can see a dog, an orange wall, a baby….,* and in this way list several things they see in one sentence. Regardless of how the learners have chosen to write, the indefinite article still must be where it belongs. This was not the case in many of these texts. With a total of 127 instances of underuse of the indefinite article within the text type description, or 58.3% of all the errors of underuse of the indefinite article, it is possibly an indication that this text type and task formulation have contributed to the large number of errors regarding underuse of the indefinite article in the 7th grade group.

4.5 The zero article

As stated in section 2.4, it is – based on the CEFR levels – expected that learners on level B1 and also A1 are able to apply the zero article correctly. Since the CORYL corpus presents errors of overuse of articles, it was therefore natural to investigate how the learners used the zero article. However – as discussed in section 4.2 – these results do not include the number of correct uses of the zero article. This section is also a part of answering the first and second research question posed in section 1.4, concerning how young Norwegian EFL learners use articles, and what type of errors they make regarding article use in their writing. The results presented in this section are a part of the previous findings discussed in section 4.3 and 4.4, simply viewed and investigated from another point of view. The results of errors of zero article can be viewed from two different angles. The first is cases where the pupils have left out the article where there should have been an article, in other words, cases where the learners may have wrongly applied the zero article. The second way to view the results is to look for cases where the pupils have used an article, but where it would be correct to leave out the article, or, in other words, to apply the zero article. Examples of such errors from the CORYL corpus are presented below.

(4.28) I heard noys coming from up stairs (p108-7)

(4.29) buy cheap products from U.S.A. (p280-10)

(4.30) we took a kurv with **a** icecrem (p58-7)

(4.31) Day on **the** school was long (p172-10)

Among the 12/13-year-old learners there were 265 cases where the article was left out where it would have been correct to include either the definite or indefinite article. Furthermore, there were 78 cases where the learners used an article, but where the zero article should have been used. As for the 15/16-year-old group, there were 124 cases where the learners used zero article where they should have included either the definite article or indefinite article. Additionally, there were 109 cases where the pupils used an article where there should have been zero article. These results can be presented as in the following diagram:
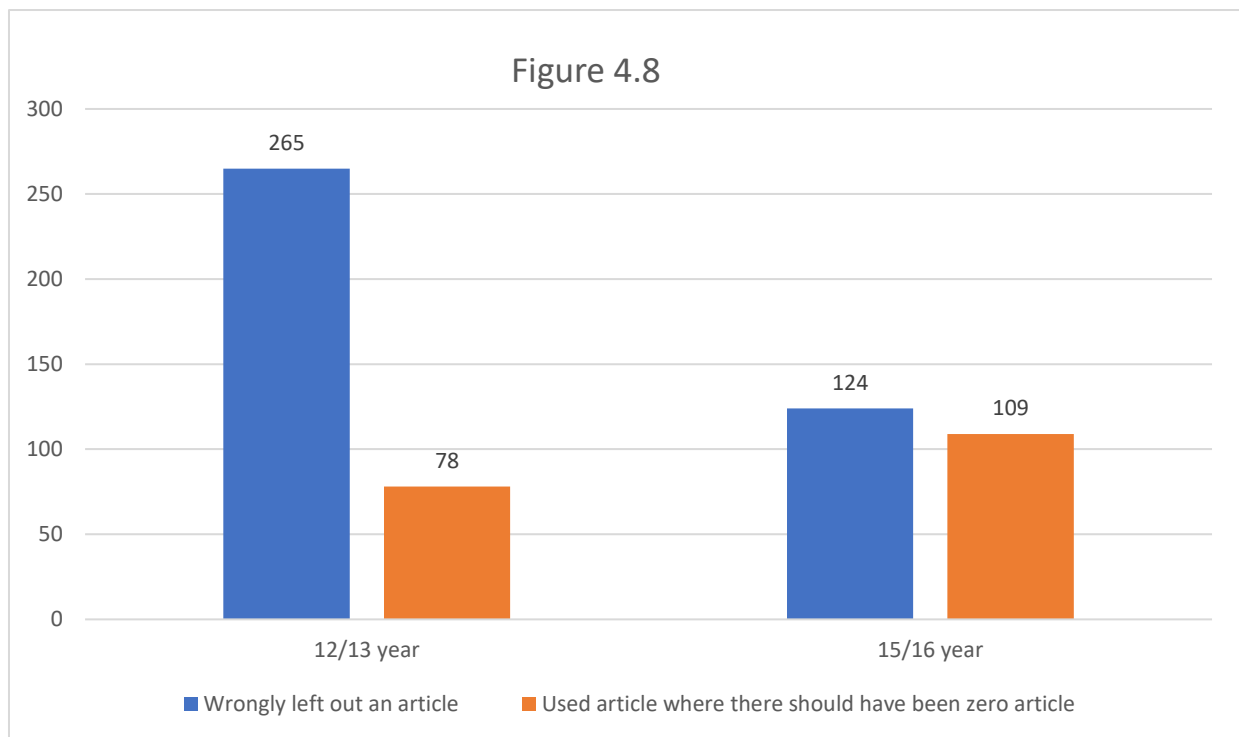
Figure 4.8

Figure 4.8: Wrong uses of the zero article

The findings dealing with the cases where the pupils have wrongly left out an article shows a rather large difference from the 7th grade to the 10th grade. With over double the amount of errors regarding the incorrect application of the zero article, it can be assumed that this is a difficult feature for these learners. However, the findings did not reveal specific patterns with these errors, and the examples provided for this section are therefore rather randomly selected. As stated in section 2.1.4, the article is left out "[…] before plural and uncountable nouns when we are referring to something in general" (Parrot, 2010, p. 28). Example (4.28) presents an instance where the learner has incorrectly left out the indefinite article, in this case, in front of the countable noun 'noise' in the singular which is referring to something that is unknown to the reader. These facts indicate that the indefinite article is the correct article to apply. Furthermore, if one translates this example into Norwegian, it would be the following: *Jeg hørte en lyd komme fra etasjen over.* Here the Norwegian indefinite article *en* is used. Transfer is not the reason for the error as the learner has chosen to completely exclude the article altogether.

Example (4.29) shows an occurrence of which the definite article has been left out in front of the proper noun *U.S.A,* as one would do when writing about it in Norwegian. This error is therefore most likely due to interference from Norwegian. Example (4.30) shows an instance of a wrongly applied indefinite article where the correct article would have been the zero article.

In this instance, the learner is talking about ice-cream, which is an uncountable noun in this context. The same is the case with (4.31); the learner's focus is on the school as an institution of learning rather than as a building. Therefore, the definite article is wrongly included instead of the zero article. Without trying to indicate the learners' CEFR levels, it is interesting to look at the level indicators presented in section 2.4. The CEFR levels indicating article use from the English Profile show that is expected both at level A1 and B1, that the use of the zero article before nouns when making general references is somewhat mastered (Cambridge University Press, 2015).

Figure 4.9 show the distribution of the wrongly left out article according to text genre and year level.
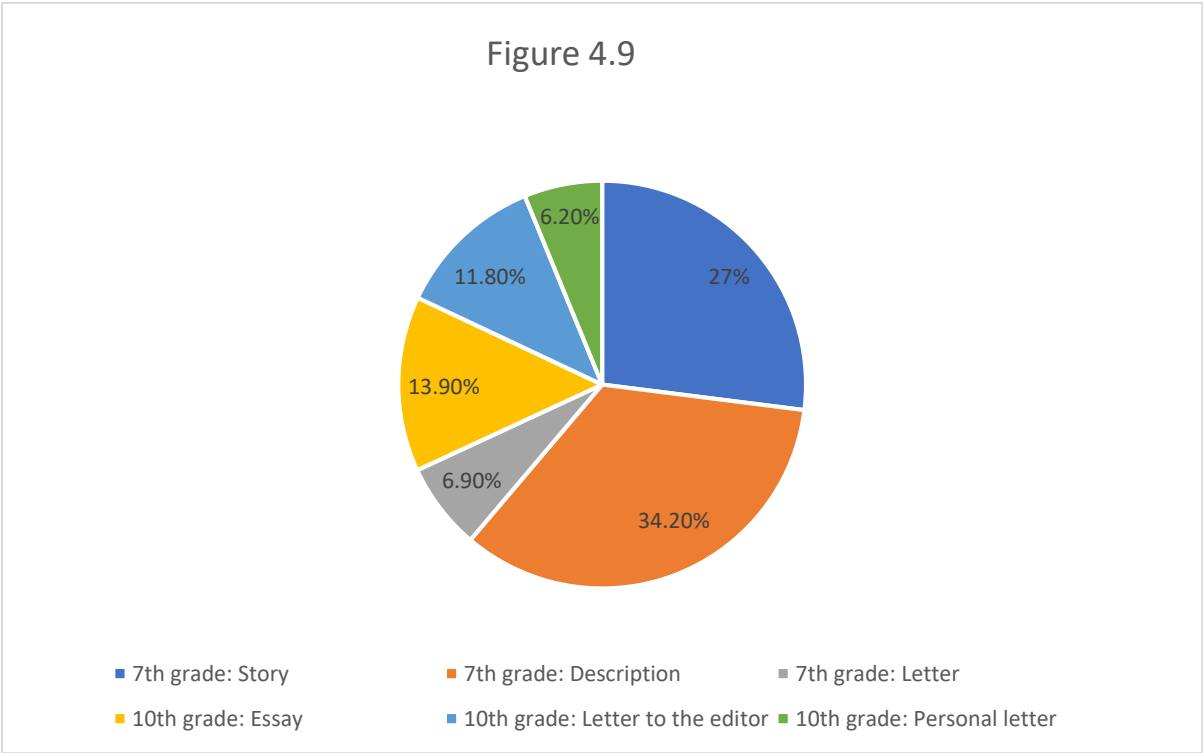
Figure 4.9: Percentage of the underuse of the zero article within each group of learners and text genre

Figure 4.9 shows that description is the text type that generates the most errors as to wrongly leaving out the indefinite or definite article, with 34.2% of the errors. The text type story generates the second most errors regarding wrongly leaving out the indefinite or definite article with 27% of the errors. It is interesting to note that both of these text types come from tasks created for the 7th grade learners. After description and story, essay is the text type that generates

the most errors among the year 10 learners, producing 13.9% of the errors. The remaining text types do not generate the same high amount of errors as the three mentioned above.

Figure 4.10 shows the distribution of overuse of the definite and indefinite article in normalized quantifications.
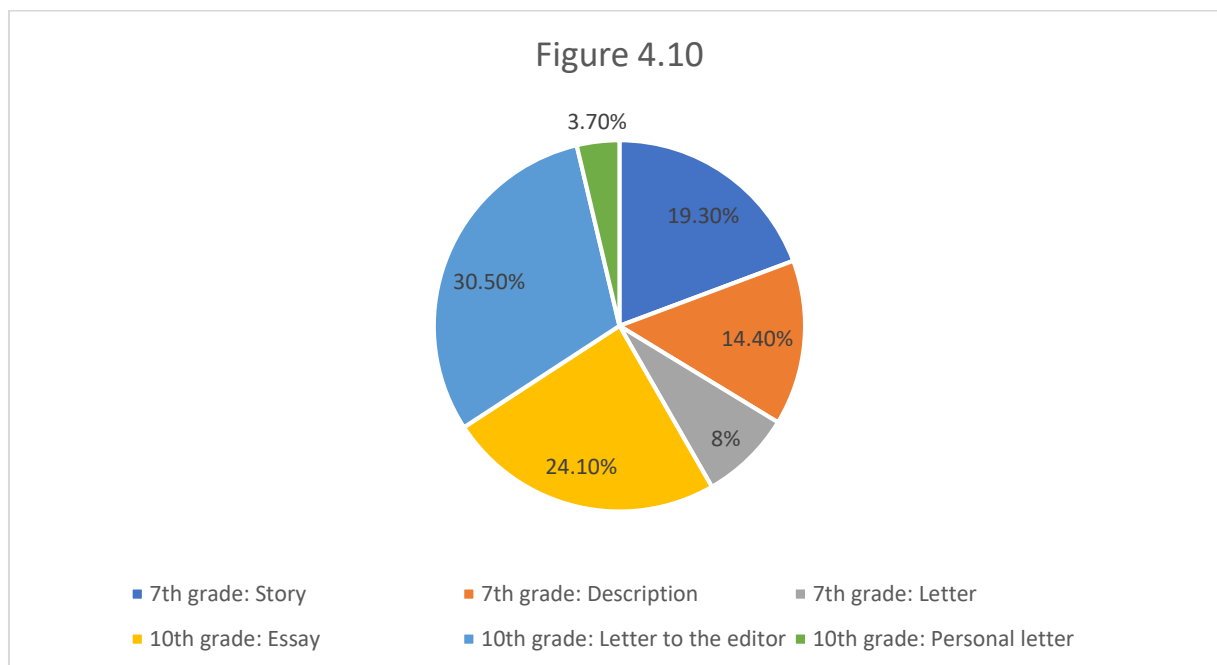

Figure 4.10: Percentage of the overuse of the definite and indefinite article within each group of learners and text genre

The most noticeable difference between figure 4.9 and 4.10 is that it is the 10th grade learners and their text types that generate the most errors in relation to failing to include the zero article, rather than the 7th grade learners. The text type letter to the editor generates 30.5% of the errors regarding overuse of the definite and indefinite article, or underuse of the zero article. Furthermore, the text type essay generates 24.1% of the errors. The 10th grade learners were given tasks within both of these text types. The text type that generates third most errors, is stories written by the 7th grade learners, with 19.3% of the errors. The task given to the learners in the text type letter to the editor was presented and discussed in section 4.3.4 and should be considered to be a possible reason for the many errors concerning underuse of the zero article.

4.6 The development in article use from 7th to 10th grade

The above sections have described and analysed how young Norwegian EFL learners use articles in their writing. Specifically, the findings and the analysis have shed light on what type

of errors these learners make in their use of articles, and possible reasons for these errors. This section therefore aims to answer the third research question: "How does the use of articles in writing develop from grade 7 to grade 10?" and the latter part of the second research question: "… and how many errors do they make compared to the instances where the articles are used correctly?" . To do so, the findings presented in sections 4.3–4.5 have been combined with the findings regarding correct use of articles. By doing this the aim is to carry out a cross-sectional study in order to track the development in the use of articles in the two groups of learners. The results are presented below in tables displaying the number of correct uses of the articles *a, an* and *the,* the number of wrong use of articles, the total sum of the correct and incorrect uses of the articles, as well as the number of incorrect article use expressed as the percentage of the total article uses. The results presented in this section therefore includes and expands on the results already presented in section 4.3, 4.4 and 4.5. Furthermore, even though the CORYL corpus is only tagged for errors, the manual search function enables searches for features that are not tagged with errors. However, as mentioned previously, it was not possible to investigate the learners correct use of the zero article. Nevertheless, it was possible to extract detailed data about the correct use of the overt articles.

When investigating the development of the use of articles from grade 7 to grade 10 it would be preferable to investigate the same group of learners over a set stretch in time through a longitudinal study. A longitudinal study is "[…] carried out over a longer period of time during which data from a group of individuals are collected at regular intervals" (ELLO, n.d). However, due to the time constraints this thesis is under and due to how the corpus is structured, this is not achievable. Therefore, a cross sectional study was most suitable. By conducting a cross-sectional study "[…] linguistic data are collected at a single point in time, usually from a relatively large group of individuals or subjects" (ELLO, n.d). For this thesis, data regarding the use of articles among learners in 7th grade and 10th grade in 2005/2011 are compiled in the CORYL corpus. These data do therefore not represent the development of article use over time, but rather the use of articles of two different groups of learners at a given point in time. There is no doubt that this fact is to some extent a limitation of the research. However, with the time available for this MA thesis, and the lack of other opportunities to obtain learner texts from the same learners over a stretch of time, there is little that can be done other than being aware of the limitations it brings.

Table 4.1 presents the number of correct uses of articles, the number of wrong uses of articles, the total, and a percentage that shows the wrong use of articles relative to the total use of articles. The data are subdivided according to year level.

Table 4.1: Correct and wrong use of all article use

	Correct use	Wrong use	Total	% wrong
7th grade	3439	438	3877	12.7 %
10th grade	2422	296	2718	12.2 %

The numbers which make up the sum of wrong use in this table are gathered from the findings from all the errors tagged with ART error from the CORYL corpus. Table 4.1 shows that the 7th grade learners' use of all the articles included 12.7% wrong uses, while the 10th grade learners' use of all the articles included 12.2% wrong uses. The difference between the two groups of learners is then 0.5%. This is a negligible difference, and as expected, the difference is not statistically significant in a *chi*-squared test of independence ($\chi^2=0.27_{(df=2)}$, p=0.60).[4] Since the 10th grade learners presumably have had almost three years more of English instruction in school it is somewhat surprising that the difference in percentage of errors regarding all article use is so small. However, as noted throughout this chapter, the uses of articles both correct and incorrect, might be affected by the text types and tasks given to the learners in the National tests in English. Table 4.2 and 4.3 below presents findings compiled from the CORYL corpus concerning the correct use of the articles *a, an* and *the*, and the incorrect uses of the definite, indefinite and zero article in relation to text type.

Table 4.2: Correct and wrong use of articles in relation to text genre 7th grade

7th grade	**Correct**	**Wrong**	**Total**	**% Wrong**
Description	1564	211	1775	11.9%
Letter	366	49	415	11.8%
Story	1509	178	1687	10.6%
Total	3439	438	3877	11.3%

---

[4] The test used is Pearson's *chi*-squared, uncorrected for continuity.

10th grade	Correct	Wrong	Total	% Wrong
Essay	1495	126	1621	7.8%
Letter to the editor	492	121	613	19.7%
Personal letter	435	49	484	10.1%
Total	2422	296	2718	10.9%

Based on the results presented in table 4.2 and 4.3 it is possible to test the findings to see if the differences between the text types are random or significant. This is done through a chi-square test. The findings concerning text type and correct and incorrect use of articles within the group of 7th grade learners do not vary greatly. In table (7th) there is one pressing issue that must be addressed, namely the findings in regard to the percentage of errors in the text type *description.* Contrary to prior assumptions – discussed in section 4.2 – there are not significantly more errors in the descriptive texts than in letters or stories. The difference between description texts and letters was not significant ($\chi^2=0_{(df=2)}$, p=1), and neither was the difference between description texts and stories ($\chi^2=1.55_{(df=2)}$, p=0.21). As for table (10th), the case was different here. The difference between essay and letter to the editor was significant ($\chi^2=64.76_{(df=2)}$, p=0.0001), and so was the difference between letter to the editor and personal letter ($\chi^2=19.09_{(df=2)}$, p=0.0001). A point for discussion is therefore what causes the text type letter to the editor and the task given to facilitate 19.7% errors of article use.

As briefly touched upon in section 4.3.4, the reason for the many occurrences of article errors within this text type might be linked to the nouns occurring in the task given to the learners. Some of these nouns were *young people, environment* and *adults.* In Norwegian these nouns would be *unge folk* (young people), *miljøet* (the environment) and *voksne* (adults). A recurring error among the article errors within the text type *letter to the editor* was that the learners left out the definite article in front of the noun *environment.* In Norwegian, this noun takes on the definite ending *-et.* A possible reason for this error is therefore transfer, as the learners might assume that the noun does not take on a preceding definite article in English, as there is no preceding article before the noun in Norwegian. Another recurring error was also how the learners wrongly included the definite article in front of the noun *adults*, which in the context they wrote about it, does not require the definite article. *Adults (voksne)* does not take on an

article in Norwegian either, when referring to them as a general group of people, as done in the examples of letters to the editors in the CORYL corpus.

It has not been possible to acquire information regarding this but based on the results in the corpus concerning article errors within the text type *letter to the editor*, it seems as though the learners have been given the opportunity to write about whether very old people or young people should be allowed to have a driving licence. Errors of article use within these texts show that the learners use the wrong indefinite article in front of the phrase *a 80 year old*, they underuse the indefinite article preceding the noun *driving license*, and they also underuse the definite article in front of nouns such as *bus, train* and *wheels.* These considerations, combined with the discussion in the previous paragraph, create a pressing question. This is whether it is the text type itself – letter to the editor – or the nouns occurring in these tasks that causes the many article errors. If the tasks in this text type include nouns that cause difficulty for the learners in terms of article use and lead them to making a number of errors, it is reasonable to assume that a different result would possibly be the outcome if the learners had been given tasks with nouns that they did not find as challenging in regards to applying articles. Therefore, it is difficult to conclude on whether it is the text type itself that facilitates so many article errors opposed to other text types, or rather the specific nouns included in the tasks. That being said, as table 4.3 shows, article errors are by far most prevalent in this genre.

4.6.1 The definite article

Research question 3 aims to investigate how the use of articles in writing develops from grade 7 to grade 10. In order to carry out such an investigation, it is necessary to compare the findings concerning correct and incorrect article use across the two groups of learners, namely the 7th and 10th grade learners who have contributed to the CORYL corpus. Table 4.4 is especially useful for visualising this development. This table provides a contrastive picture which can shed light on the tendencies as concerns incorrect use. The number of correct occurrences of the definite article, alongside the number of erroneous occurrences of the definite article are presented. The wrong uses are also included and presented as a percentage of the total. The data are subdivided according to year level.

Table 4.4: Correct and wrong use of the definite article

	Correct use	Wrong use	Total	% wrong
7th grade	1587	165	1752	10.4 %
10th grade	1608	186	1794	11.7 %

The numbers which make up the sum of wrong use in this table include all the instances of overuse of the definite article, underuse of the definite article, and all instances where the definite article is incorrectly used instead of the indefinite article. The definite article *the* is used correctly 1587 times in the 7th grade learner group. Of these, 436 of them occur in the text type description, 199 of the occur in the text type letter, and 952 of the occur in the text type story. The definite article is used correctly 1608 times in the 10th grade learner group. Of these, 1002 of them occur in essays, 372 of them occur in letters to the editor, while 234 of them occur in personal letters.

The number of correct use and wrong use of the definite article is quite similar between the two groups of learners. Percentagewise, 1.3% separates the 7th graders from the 10th graders in terms of wrong use of the definite article. The 7th graders errors regarding the definite article make up 10.4% of the total, while the 10th graders errors make up 11.7% of the total. The difference between the groups is not statistically significant ($\chi^2=0.9_{(df=2)}$, p=0.34). As noted in section 4.6, it is somewhat puzzling that the difference between the two groups of learners is so small. There are specifically two reasons for why this result is unexpected, namely the difference in instruction of English, and also what Master (1997) writes about the definite article, as mentioned in section 2.2. According to Sun (2016, p. 2), *the* is the first article in the order of article acquisition presented by Master (1997). Furthermore, Master (1997) also suggests that this is the first article that is learned, due to the fact that it is not "(…) limited by countability of number" (p. 225). Since the 10th grade learners most likely have had more instruction in English, is reasonable to expect that they would have produced fewer errors than the 7th graders, and based on the Master's (1997) article on article acquisition and pedagogy, it is also reasonable to expect that the 10th grade learners would master the use of the definite article *the* more proficiently than the 7th graders. Section 4.3.4 addressed the notion of whether the task type given to the 10th grade learners might have affected their incorrect use of the definite article, and specifically their overuse of the definite article. A possible implication is therefore whether the outcome possibly could have been significantly different had the learners received a different task. In other words, would the learners still make the same amount and type of

errors concerning article use had they revived different tasks that possibly facilitated a different use of articles?

4.6.2 The indefinite article

As with section 4.6.1, this section and the findings that will be presented are useful in terms of comparing the results between the two groups of learners in order to track the development in article use from 7th to 10th grade. Table 4.5 includes a presentation of the total number of correct uses of the indefinite article (*a* and *an*) together with the number of errors connected to this article. The data are subdivided according to year level.

Table 4.5: Correct and wrong use of the indefinite article

	Correct use	Wrong use	Total	% wrong
7th grade	1852	215	2067	11.6 %
10th grade	841	63	904	7.5 %

The numbers which make up the sum of wrong use in this table include all the instances of overuse of the indefinite article, underuse of the indefinite article, and all instances where the indefinite article is incorrectly used instead of the definite article. The indefinite article is used correctly 1852 times among the 7th grade learners. The indefinite article *an* is used correctly 22 times among the 7th grade learners. Of these, 6 of them occur in the text type description and 15 of the occur in the text type story. As for the indefinite article *a*, this is used correctly 1830 times. Of these, 1122 of them occur in description texts, 166 of them occur in letters, while the remaining 542 instances occur in stories.

The indefinite article is used correctly 841 times among the 10th grade learners. The indefinite article *an* is used correctly 56 times in this group of learners. Of these, 32 occur in essays, 12 of them occur in letters to the editor, and the remaining 234 correct uses of the indefinite article *an* occur in personal letters. The indefinite article *a* is used correctly 756 times. Of these, 459 of them occur in essays, 108 correct uses occur in letters to the editor, while 189 occur in personal letters.

As table 4.5 shows, the 7th grade learners use of the indefinite article included 11.6% wrong use. The 10th grade learners use of the indefinite article included 7.5% wrong use. There is a difference of 4.1 %, which is the largest difference between the two groups of learners and

article errors in regard to the indefinite article, and the difference between the groups is statistically significant ($\chi^2=8.74_{(df=2)}$, p=0.003). The fact that the difference is statistically significant indicate that the difference has not occurred simply by chance, rather most likely caused by a specific feature. A possible reason for the number of these errors being higher among the 7th grade learners is the task the learners were given. The 7th grade learners received a task that might have induced a more frequently use of the indefinite article, both correct and incorrect. When the results are presented together with the percentage of wrong use of the indefinite article, it is evident that the 10th grade learners are more proficient concerning this feature than the 7th grade learners.

4.7 Discussion

Section 4.3 – 4.6 aims to answer the following research questions, as presented in section 1.4:

1) How do young Norwegian EFL learners use articles in their writing?
2) What errors do the learners make concerning article use, and how many errors do they make compared to the instances where the articles are used correctly?
3) How does the use of articles in writing develop from grade 7 to grade 10?

The findings presented in this chapter have provided information concerning how the learners who have contributed to the CORYL corpus use articles in their writing and therefore answers research question (1) and partially research question (2). The learners make errors within all of the categories under discussion which include underuse, overuse and wrong use of articles. However, the number of errors within each of these categories vary and some categories generate more errors than others. These categories were overuse of the definite article, underuse of the indefinite article and underuse of articles in general. The reasons for why the article errors occur in the corpus are somewhat complex. Some of the errors are clearly caused by interference from Norwegian while other errors seem to be prompted by the way the pupils are asked to answer the different tasks. Some of the tasks seem to generate more article errors than other tasks, especially the text type letter to the editor. Therefore, genre has been considered a contributing factor as to why errors in article use occur. Also, some errors seemed to be due to typographical errors, while it was challenging to determine reasons for the occurrence of other

66

errors. Nevertheless, based on the findings presented in this chapter, it is evident that article use among young Norwegian EFL learners is a challenging grammatical feature to acquire. These findings are in line with previous research and theory concerning article use, errors and acquisition presented in chapter 2 (Bækken, 2006; Master, 1997; Sun, 2016).

The latter part of research question (2), how many errors the learners make compared to the instances where the articles are used correctly, is answered thoroughly with a cross-sectional study in section 4.6. The results show that the learners use articles correctly a large number of times more than they use them incorrectly. As concerns research question (3), the 7th grade learners make in total 12.7% errors concerning the total use of articles, while the 10th grade learners make in total 12.2% errors concerning the total use of articles. By using the *chi*-square test it was clear that these results were not statistically significant. However, the findings presented in section 4.6 suggest that there is a small overall decrease of article use errors from grade 7 to grade 10, but the changes are too insignificant to determine this development. Nevertheless, table 4.4 and 4.5 indicate that it is the incorrect use of the indefinite article that contributed most in term of the small difference there actually is between the two groups of learners. While there was a difference on 1.3 % between the 7th grade learners and the 10th grade learners concerning the wrong use of the definite article, the 10th graders making more errors than the 7th graders percentagewise, there was a difference on 4.1% between the two groups concerning wrong use of the indefinite article with the 7th graders making the most errors.

Additionally, a notion that is interesting to consider here is how the findings from this research can be of use or be applied to the teaching of English as a foreign language. Corpus linguistics has a history of contributing to creating teaching resources, such as grammar books, and also dictionaries. In the article 'Corpus-based approaches to foreign language pedagogy' (2002), Fanny Meunier discuss how both native and learner corpora can influence EFL grammar teaching "[…] distinguishing between three domains of application: curriculum design, reference tools and classroom EFL grammar teaching […]" (p. 119). Reference tools such as dictionaries and EFL grammar textbooks have both benefited from corpus research. Dictionaries – which historically have included the usual grammatical and lexical information – now include frequency and they "[…] register information in the form of language/usage notes illustrating, among other things, differences between spoken and written language" (Meunier, 2002, p.125). This additional information has been gathered by the help of corpus research, and information concerning e.g. frequency would be challenging to obtain otherwise. This MA thesis has produced findings concerning the frequency and type of article errors

among young Norwegian learners of English, as well as the frequency of the correct uses of articles.

Furthermore, learner corpus research makes it more manageable to identify features that learners struggle with, and it also considers the learners L1 which can lead to grammar teaching that is more suitable for the target learners (Meunier, 2002, p. 123). This MA thesis has identified the types of article errors which young Norwegian learners of EFL make. However, as concerns e.g. curriculum design, this specific MA thesis does most likely not carry direct relevance. Nevertheless, a corpus study like the one carried out in this MA thesis can contribute with a description of learner language, specifically concerning the specific feature of article use. As stated in section 1.3, such descriptions are needed, as there is little literature that investigates the language of young Norwegian learners of English as a foreign language.

In conclusion, this specific corpus-study is too small to have any direct impact on the creation of an English-Norwegian dictionary, to exemplify. Nevertheless, this research may function as a building block to a larger and more comprehensive study, of a perhaps larger corpus which would be more representative. Furthermore, if looking at it from an angle of a student teacher or a practicing teacher, findings from studies like this might be of use as it can raise awareness of a grammatical feature that is problematic for Norwegian learners of English as a foreign language. Being aware of possible problematic areas for learners is useful when planning how to teach a grammatical feature and also when detecting what might cause the difficulties the learners are facing. Specifically for this thesis, the findings and results can be used by teachers to design tasks that aims to give the learners practice in applying the correct article and also to create a set of guidelines as to which errors one should be aware of.

4.8 Summary

This chapter presents and discuss the findings of the research carried out in this MA thesis. Section 4.1 introduce the structure of this chapter. Section 4.2 follows with a presentation and discussion of exclusions and practical limitations concerning the findings. Section 4.3 presents and discusses the findings regarding the definite article. Section 4.4 presents and discusses the findings concerning the indefinite article, while section 4.5 focuses on the zero article. Section 4.6 deals with the findings regarding development in article use from 7th to 10th grade, and a cross-sectional study of this development is presented and discussed in this section. Section 4.7

discusses the research questions in relation to the findings, and thoughts about possible further research and how this thesis contributes to the field of corpus linguistic studies is discussed.

Chapter 5: Conclusion

The aim of this master thesis was to investigate the use of articles in the writing of young Norwegian learners of English as a foreign language. This has been done by using corpus linguistic methods and the CORpus of Young Learner Language (CORYL). The reason for investigating this specific topic was the lacking focus of language development research on young Norwegian learners of English. Previous corpus-based studies on learner language tend to focus on older or more proficient language learners – typically undergraduate University students. Therefore, acquiring information about young Norwegian EFL learners is important in order to continue to develop the curriculum, the teaching of grammatical feature such as article use, and teaching supplies such as books and tasks to promote grammatical proficiency.

The manner in which the data was collected was thoroughly described in chapter 3. Corpus linguistic methods were employed, such as data extraction based on the annotation of the CORYL corpus as well as the corpus analysis tools that were available through the CORYL corpus. Based on these methods, chapter 4 presented the findings alongside an analysis and discussion of these findings, including an error analysis, an overview of the correct use of the articles, as well as a cross-sectional study of the development in article use from 7th grade to 10th grade.

The findings and analysis in this thesis produced some expected results, and some unexpected results. The overall findings regarding correct and wrong use of all article use showed that the 7th grade learners made the most errors, but not by a great deal. By testing the findings concerning the overall use of article errors with the *chi*-square test, it was clear that the difference was negligible. Furthermore, the findings showed that the 7th graders made fewer errors concerning the use of the definite article, while they made the most errors concerning the use of the indefinite article. As for the 10th grade learners it was the opposite, as they made the most errors concerning the definite article and fewer errors with the indefinite article.

As for possible reasons for errors of article use in the writing of young Norwegian EFL learners, two main reasons were discussed in this thesis, namely transfer from Norwegian and the possibility of how different text types or task design might affect the number of article errors occurrences. As noted in section 1.2, according to Bækken (2002), transfer from Norwegian seems to be the reason for many of the errors that Norwegian learners of English make. The findings in this thesis are to a certain degree in line with this notion. Since it is not possible to

investigate what the learners thought when applying articles to their text, it is impossible to say definitely whether the errors are due to transfer or not. However, it is not unlikely that they are, which can be seen in many of the examples provided from the CORYL corpus in chapter 4. Since both Norwegian and English have an article system and since the learners are in a process of becoming proficient in English as a foreign language, it is imaginable that situations occur where the learners mix the rules of article application in Norwegian and English and errors are then made.

Another possible reason for the article errors in the CORYL corpus was considered to be due to text type and task design. Based on the findings presented in chapter 4 it was evident that the text types letter to the editor and description was the source of many article errors. The findings showed that most errors of overuse of the definite article occurred in the text type letter to the editor, while most errors of underuse of the indefinite article occurred in the text type description. The description texts were written by the 7th grade learners, and the letters to the editor were written by the 10th grade learners. These two text types also generated the most article errors concerning wrongly left out article and overuse of the definite and indefinite article, in other words, where the learners should have applied the zero article. It is interesting to note the fact that the difference between description and the other text types produced by the 7th grade learners – story and letter – was statistically insignificant. The difference between the text type letter to the editor versus essay and personal letter, which were produced by the 10th grade learners, was statistically significant. Therefore, it can be said that the text type letter to the editor facilitated significantly more article errors compared to the other text types presented in the CORYL corpus. However, it must be considered that these findings might be induced by the nouns occurring in the tasks, and therefore, the text type letter to the editor in itself is possibly not the facilitator of the many occurrences of the article errors. A decision was made to focus on the type of error that occurred the most within each type of article. Therefore, there are no figures that specifically illustrate the distribution of errors considering text type of for example underuse of the definite article or the overuse of the indefinite article underlying each type of article in chapter 4.

When conducting a corpus study like the one in this MA thesis it is interesting to consider how the findings can be of use or applied to the teaching of EFL. Learner corpus research is useful as it can present a description of learner language. These descriptions may influence teaching resources such as textbooks, dictionaries or the way in which English grammar is taught to learners of English as a foreign language. However, in order to do so, a more comprehensive

study of larger corpora is necessary. Nevertheless, the corpus-based study carried out in this MA thesis may be a building block towards a more comprehensive, and in turn, a more representative study.

This being a master thesis, there are of course restraints concerning time and space in terms of words. Therefore, there are features of this research that is not able to investigate, such as a complete developmental study of the use of articles among learners from 7th to 10th grade or to go deeper in to the CORYL corpus. These notions are then indications of what further research within this field could possibly entail. Article use is an interesting topic, and since articles are very frequently used words in the English language, learners of English as a second language are bound to face them, and possibly to face challenges with mastering this grammatical feature. This notion, and the fact that there is little research on young Norwegian learners of English as a foreign language, makes it a relevant field to do further research on. It is reasonable to believe that such research would be useful and beneficial for the teaching of article use in English, referring to both individual teachers, but also to the development of the grammatical curriculum and teaching resources.

References

Baker, P., Hardie, A., & McEnery, T. (2006). *Glossary of corpus linguistics*. Retrieved from: https://ebookcentral.proquest.com

Bækken, Bjørg. (2006). *English Grammar: An introduction for students of English as a foreign language.* Bergen: Fagbokforlaget.

Cambridge University Press. (2015) English Grammar Profile Online. Retrieved from: https://www.englishprofile.org/english-grammar-profile/egp-online

CORYL, (nd). *Coryl User Manual.* Bergen: University of Bergen.

Clarino. (2015). CMDI 1.1. Metadata. Retrieved from: http://clarino.uib.no/comedi/metadata-record?identifier=coryl

Clarino. (2019). Corpuscle: Corpus list. Retrieved from: http://clarino.uib.no/korpuskel/corpus-list?session-id=246812505434161

Council of Europe. (2001). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Retrieved from: https://rm.coe.int/1680459f97

Dörnyei, Z. & Csizér, K (2012). How to Design and Analyze Surveys in Second Language Acquisition Research. In Mackey, A., & Gass, S. (eds). *Research methods in second language acquisition: A practical guide* (Vol. 3, Guides to research methods in language and linguistics). Chichester: Wiley-Blackwell.

Duff, P. A., Biber, D., & Conrad, S. (2001). Corpus-Based Research in TESOL: Quantitative Corpus-Based Research: Much More Than Bean Counting. *TESOL quarterly*, *35*(2), 331-336.

Ellis, Rod. (1996). *The Study of Second Language Acquisition.* Oxford: Oxford University Press

ELLO. (n.d). English Language and Linguistics Online: Field Work 1: longitudinal versus cross-sectional studies. Retrieved from: http://www.ello.uos.de/field.php/TheoryModelMethod/TheoryModelMethodFieldWork1Longitudinal VersusCrossSectionalStudies

EnglishProfile. (2011). Introducing the CEFR for English. Retrieved from: https://www.englishprofile.org/images/pdf/theenglishprofilebooklet.pdf?

Granger, S. (2012). How to Use Foreign and Second Language Corpora. In Mackey, A., & Gass, S. (eds). *Research methods in second language acquisition: A practical guide* (Vol. 3, Guides to research methods in language and linguistics). Chichester: Wiley-Blackwell.

Granger, S. (2002). A bird's-eye view of learner corpus research. *Computer learner corpora, second language acquisition and foreign language teaching*, *6*, 3-33.

Hasselgreen, A., & Sundet, K. (2017). Introducing the CORYL Corpus: What it Is and How We Can Use it to Shed Light on Learner Laguage. *Bergen Language and Linguistic Studies. 7.* doi: Retrieved from: http://dx.doi.org/10.15845/bells.v7i0.1107

Hasselgård, H., Lysvåg, P,. Johanson, S. (2012). *English grammar: Theory and Use.* (2nd edition). Oslo: Universitetsforlaget

Ionin, T., Zubizarreta, M. L., & Maldonado, S. B. (2008). Sources of linguistic knowledge in the second language acquisition of English articles. *Lingua*, *118*(4), 554-576

Larson-Hall, J. (2012). How to Run Statistical Analysis. In Mackey, A., & Gass, S. (eds). *Research methods in second language acquisition: A practical guide* (Vol. 3, Guides to research methods in language and linguistics). Chichester: Wiley-Blackwell.

Master, P. (1997). The English Article System: Acquisition, Function and Pedagogy. System, 25, 215-232. Retrieved from: https://www.researchgate.net/profile/Peter_Master/publication/240920277_The_English_Article_System_Acquisition_Function_and_Pedagogy/links/59db95c30f7e9b1460fbcfa5/The-English-Article-System-Acquisition-Function-and-Pedagogy.pdf

Master, P. (2002). Information structure and English article pedagogy. System, 30, 331–348.

Retrieved from:

https://www.researchgate.net/profile/Peter_Master/publication/223357198_Information_Structure_and_English_Article_Pedagogy/links/5ab70cf90f7e9b68ef502568/Information-Structure-and-English-Article-Pedagogy.pdf

McEnery, T. & Wilson, A. (2001). *Corpus Linguistics.* (2nd edition). Edinburgh: Edinburgh University Press.

Meunier, F. (2002). Corpus-based approaches to foreign language pedagogy. In Granger, S., Hung, J., & Petch-Tyson, S. (Eds.). (2002). *Computer learner corpora, second language acquisition and foreign language teaching*.

Meyer, C. F. (2002). *English corpus linguistics: An introduction*. Cambridge University Press.

Moen, P. & Pedersen, P-B. (2003). *Norwegian Grammar.* Kristiansand: Høyskoleforlaget.

Oxford Dictionaries. (2018). What can the Oxford English Corpus tell us about the English language? Retrieved from: https://en.oxforddictionaries.com/explore/what-can-corpus-tell-us-about-language/

Parrot, M. (2010). *Grammar for English Language Teachers (*2nd edition). Cambridge: Cambridge University Press.

Salimi, A., Dadaspour, S., & Asadollahfam, H. (2011). The effect of task complexity on EFL learners' written performance. *Procedia-Social and Behavioral Sciences*, *29*, 1390-1399.

Seland, I., Vibe, N., & Hovdhaugen, E. (2013). Evaluering av nasjonale prøver som system. NIFU.

Sun, G. (2016). The acquisition of English articles by second language learners: The sequence, differences, and difficulties. SAGE Open, 6(1), 2158244016635716.

The Norwegian Directorate for Education and Training. (2018). *Kva er nasjonale prøver?* Retrieved from: https://www.udir.no/eksamen-og-prover/prover/nasjonale-prover/om-nasjonale-prover/

The Norwegian Directorate for Education and Training. (2017). *Kva måler nasjonale prøve i engelsk?* Retrieved from: https://www.udir.no/eksamen-og-prover/prover/nasjonale-prover/mestringsbeskrivelser-og-hva-provene-maler/kva-maler-nasjonale-prove-i-engelsk/

The Norwegian Directorate for Education and Training. (2013a). *Læreplan i engelsk (ENG1-03): Kompetansemål etter 7.årstrinn.* Retrieved from: https://www.udir.no/kl06/ENG1-03/Hele/Kompetansemaal/kompetansemal-etter-7.-arstrinn

The Norwegian Directorate for Education and Training. (2013b). *Læreplan i engelsk (ENG1-03): Kompetansemål etter 10.årstrinn.* Retrieved from: https://www.udir.no/kl06/ENG1-03/Hele/Kompetansemaal/kompetansemal-etter-10.-arstrinn

Thewissen, J. (2013). Capturing L2 Accuracy Developmental Patterns: Insights From an Error-Tagged EFL Learner Corpus. *The Moderns Language Journal, 97.* doi: Retrieved from: https://doi-org.galanga.hvl.no/10.1111/j.1540-47812012.01422.x

Universitetet i Bergen. (2019). CORYL. Retrieved from: http://clarino.uib.no/korpuskel/simple-query?session-id=246858507587517

VanPatten, B. & Williams, J. (2015). *Theories in Second Language Acquisition: An Introduction.* Routledge: New York.

Vassarstats. (2019). 2x2 Contingency Table. Retrieved from: http://vassarstats.net/tab2x2.html

Appendix
A: Description task 7th grade

Time: 60 minutes
Answer all three tasks. Spend half the time on task 3.

1. Look at the picture. What do you see?

_____

_____

_____

_____

_____

_____

_____

_____

WRITING TEST

Time: 90 minutes

Answer all three tasks. Choose topic a) or b) in task 3.

Spend about half the time on task 3.

1   Some English/American friends are coming to visit you. Write a <u>short</u> letter to them telling them about a place they simply must see while they are in Norway. Tell them, for example,:
WHAT.... WHERE…. and WHY.

2   **"Young people today care less about the environment than adults"**
This headline appeared in your local newspaper recently. Write a response of 1 or 2 paragraphs to this editorial, providing arguments that will really convince the readers of your point of view.

3   EITHER

    a) A teen magazine is publishing a special edition on famous people.
       Write a text, 5 – 8 paragraphs, about a person you think has made a difference to the world. Give your text a title.

       OR

    b) Write a text, 5 – 8 paragraphs, with the title:

       "The day everything went wrong"