Faculty of humanities, social sciences, and education

I kastet me down on the sofa

A dive into lexical errors and Norwegian influence in the English of pupils in Norway

Frode Fagerås Hansen & Sondre Schløpke

I

III

Abstract

This study aims to find the distribution between interlingual and intralingual influence among pupils in Norway, in addition, we want to observe to what extent the lexical errors among the pupils are influenced by Norwegian. We also observed which lexical errors influenced by Norwegian were the most common.

In order to answer our thesis and research questions, we used a quantitative research approach where we got the data from the CORYL corpus. We used three different codes to search for Norwegian influence and word errors/vocabulary mistakes. Our target group consisted of 472 pupils aged 12/13 and 212 pupils aged 15/16. The texts that are in the CORYL corpus are gathered from the National Test in English (writing) from 2004 and 2005.

This study has found that 12/13-year-old tend to create more interlingual errors than the older group aged 15/16. This is because they on a much larger scale tend to use whole Norwegian words and have Norwegian elements in their English. The older group which consists of pupils aged 15/16, do not have the same issue, as they tend to have more intralingual mistakes than interlingual mistakes.

IV

v

Sammendrag

Målet med denne studien er å finne fordelingen mellom interspråklig og intraspråklig påvirkning blant elever i Norge, i tillegg ønsker vi å observere i hvilken grad de leksikalske feilene blant elever blir påvirket av norsk. Vi observerte også hvilke leksikalske feil som er de mest vanlige for hver av gruppene som skyldtes påvirkning fra norsk.

For å kunne svare på forskningsspørsmålene brukte vi en kvantitativ forskningsmetode, der vi hentet ut data fra CORYL korpus. Her brukte vi tre forskjellige koder for å søke etter norsk innflytelse og vokabularfeil. Målgruppen vår besto av 472 elever i alderen 12/13 og 212 elever i alderen 15/16. Alle tekstene til disse elevene er fra CORYL korpus, som hentet disse fra nasjonale prøver i engelsk (skriving) fra 2004 og 2005.

I denne studien har vi også at 12/13-åringer har en tendens til å ha flere interspråklige feil enn 15/16-åringer. Dette er trolig fordi de i mye større grad har en tendens til å bruke hele norsk ord og ha norske elementer i deres engelsktekster. Den eldre gruppen som besto av 15/16-åringer hadde ikke samme problem, da de hadde en tendens til å ha flere intraspråklige feil enn interspråklige feil, slik som å velge riktig ord til riktig kontekst.

Nøkkelord: *førstespråkinnflyelse i andrespråket, intraspråklige og interspråklige feil, Engelsk læring, korpus studie, engelskutdanning, leksikale feil, språklig utvikling, CORYL korpus, engelsk undervisning og læring.*

Table of Contents

List of Tables

List of Figures

Explanation of central terms

Code: A already existing code found in the CORYL corpus, which helps navigate the corpus in the search for specific mistakes made by the pupils. An example of this can be the SP-code, which used will help find spelling errors in the corpus.

Label: A code/error category given by us to the matches from the CORYL corpus; Calque errors (CE), False friends (FF), Indeterminate (ID), L2 Vocabulary confusion (VC) and Norwegian Elements (NE). (Explanations for all of these are found in section 3.3 The procedure of the study).

CORYL corpus: A Norwegian Corpus, based in Bergen.

Match: The word/words that the CORYL corpus marked as an error when searching for a code. (Example: when searching the code L1/NOR, the Norwegian word "Skulle" shows up as a match).

Token: The total amount of words, error codes, numbers, and signs.

L1: The first language of a person (or mother tongue).

L2: A person learning a second language. (Example: A person with Norwegian as L1 learns English as an L2).

Lexical mistakes: Lexical errors are errors related to word usage. (Wrong word choice, collocation, literal translation, incompletion, word formation).

Interlingual influence: Influence from another language when using a target language. (Example: Norwegian influence when writing English).

Intralingual influence: Confusion within a language (Example: Mixing two different words that look almost identical and wrong word choice)

1 Introduction

Norwegians tend to use Norwegian elements in their English. This leads to sentences that are incorrect and can be quite confusing for native speakers. An example of this can be found in our headline, which is from our data set, where it says "I kastet me down on the sofa". This is one of many examples found where pupils tend to mix L1 and L2 Norwegian influence in English such as in this example is something we wanted to research further, because we think such a study will yield data that will be useful to teachers of English in Norway. Errors that are due to influence from another language are referred to as "interlingual errors". However, we will also study "intralingual errors", which are errors that are a result of confusion with other words in the target language.

This study takes a quantitative approach, where we want to explore:

1. Which types of lexical errors do pupils aged 12-13 and 15-16 make in Norway when they are learning English.
2. How does the frequency of these differ between the two age groups.

What we will look at is the L1 influence on L2 learning, such as lexical transfer. "Lexical transfer, stated simply, is the influence of word knowledge in one language on a person's knowledge or use of words in another language." (Jarvis & Pavlenko, 2008, p. 72).

The data we used is gathered from CORYL corpus, which has gathered texts from the National Test in 2004 and 2005 from pupils aged 12-13, 15-16 and 17-18 in English writing. The corpus consists of texts from 880 pupils and is managed by the University of Bergen.

1.1 Background

The field of vocabulary is both wide and important. It is important to be understood and to understand others. For this reason, it would be important and useful to understand how the influence of Norwegian can influence learning English. The Ministry of Education and Research (2020, p. 2) states that "Language learning refers to developing language awareness and knowledge of English as a system, and the ability to use language learning strategies.". The Ministry here talks about language learning, where language awareness is a central element. This means that when learning English, you need to be aware of for example, how

one language can affect learning another one. Mahan & Brevik (2013, p. 37) argues that lexical mistakes first and foremost come from the lack of knowledge of the word they are using. Later they mention that "Interlingual lexical mistakes are due to errors across languages. This means that when pupils and students lack knowledge of a word, they use their mother tongue as a starting point. The pupil or pupil's native language or mother tongue can therefore have a huge influence on learning a new language such as the English language.

As Mahan and Brevik mentioned, one often tends to use one's L1 as a starting point when learning an L2, but this can be deceiving when learning a new language. A lot of words across different languages might both look or sound alike but have different meanings. For this reason, awareness of how languages can differ is severely important. An example of this can be calque errors, where one translates each element of a word into English because of a lack of knowledge about the word intended to use.

Example 1: Five minutes later a big red *firecar* came, and took the snake

In the example above we can observe that the pupil has used the word "firecar" in his/her text, and what the pupil has done is translate the direct translation of the word "brannbil" from Norwegian. The pupil has translated each of the elements "brann" and "bil" into "fire" and "car" and put them into one word. Mistakes like these are probably due to a lack of vocabulary, where the pupil has not been taught the word "firetruck", and therefore improvises and creates errors like this one.

Hemchua and Scmitt (2007, p. 3) argue that the importance of vocabulary in second language (L2) writing is widely accepted, but there has been relatively little research into the lexical errors that learners produce when writing in their second language. For this reason, we want to dive into lexical errors that pupils in Norway make, to try to get a larger understanding of why they do occur.

How big is the vocabulary of English learners, and what is the difference between first and second language learners? An educated native speaker has a vocabulary of around 20 000 words, while a second language learner is lucky if they have acquired 5000 words over several years (Thornbury, 2020, p. 20). An English native speaker will grow up with the language and get input from their families and surroundings in their mother tongue.

Meanwhile, a second language learner will start learning English at school and have a few hours of lessons each week. Not only does a native speaker get a lot of years of English input and output before the second language learner even starts learning, but they also continue to get exposed to the language more than the second language learners. A classroom learner would need more than eighteen years of classroom exposure to get the same amount of vocabulary input as one year in a natural setting (Thornbury, 2020, p. 20). Therefore, we cannot expect to teach a classroom full of pupils to be even close to the same English level as a native speaker. If pupils want to get on the same level as native speakers, they must learn by their own initiative in their spare time. There are not enough hours to learn everything inside a classroom at the school to reach the same level, but there are enough lessons to become fluent in English and still be at an adequate level to communicate in English. However, this presupposes that teaching time is used effectively and purposefully.

How many words does a learner need to know to be able to communicate? Knowledge of the 2000 most frequent English words would provide familiarity with nearly nine out of ten words (Thornbury, 2020, p. 21). To be able to communicate with other people that talk English is a valuable quality to have. It opens opportunities to talk with more people across the world because there are more people that talk English than Norwegian. The world becomes more digital every year, and a lot of information is in English, and English is often used when communicating on the internet. English is also often used when traveling, and it will increase the chance to be able to communicate.

The most important words to learn are the words that are often used. There is a strong argument for equipping learners of English with a core vocabulary of 2000 high frequency words as soon as possible (Thornbury, 2020, p. 21). By learning the high frequency words, the learner will be able to understand most of the English spoken or written and would be able to understand unknown words from a contextual setting. This can motivate the learner because the learner sees the value in learning each word because it is often used. Learning the most used words will increase the chance to understand conversations and will help more than learning words that are less used on a day-to-day basis. "In fact, it has been calculated that the most frequent 100 words in English make up almost fifty percent of most texts." (Thornbury, 2020, p. 21). Learning the most used words will help with making sentences and are

important in the basics of grammar. The most used words in English are often the small grammar and function words, for example "the", "to", and "that" (Thornbury, 2020, p. 21).

To make a coherent sentence a learner needs to also learn content words, which are words that can be found in the dictionary and give a sentence meaning. It is important to not only focus on these small function words but also learn content words that bring meaning. As stated in the core elements of the curriculum "The teaching shall give the pupils the opportunity to express themselves and interact in authentic and practical situations." (The Ministry of Education and Research, 2020, p. 2). The educational departments are therefore supposed to let the pupils express themselves and get them ready to be able to communicate in practical situations. Therefore, it is important to first let the pupils learn the high frequency words and function words, and then try to expand on their content words. If the pupils can increase their vocabulary, they can be clearer and more specific in their language and therefore be better understood by others. Research has shown lexical errors cause more problems of interpretation and a higher degree of irritation to the native speaker (Johansson, 1978, p. 71).

1.2 Motivation for the project

Pupils that have Norwegian as a first language and are learning English as a second language is going to have both advantages and disadvantages. According to Thornbury (2002, p. 18) when someone learns a new language, they must learn a new conceptual system, and construct a new vocabulary network. New learners must learn a separate system from what they currently know, which can be difficult for some learners. During our practice periods, we have seen pupils mix in some Norwegian elements when writing and speaking English. We decided that we wanted to do research within this field and find out how big of an influence the Norwegian language has been on the pupils that are learning English as a second language. We also thought about what we did as pupils, and we used to make the same mistakes that the pupils in the CORYL corpus did. For example, when we were unsure about how a word was written in English, we just made it look more English by adding letters such as "ck". We wanted to go deeper in the research and find out more about the influence and connection between Norwegian and English.

Part of the motivation for this study is that it can be used by others. Teachers can use this to learn more about the influence the Norwegian language has when learning English, and what

parts of Norwegian influence the pupils struggle the most with. We hope that this thesis can be useful for English teachers in Norway, by helping them to understand the most common mistakes that pupils in Norway make.

We also believe that teachers can use this study to compare their own English class with the results from this study. We have gathered data from a quite large and well-worked-through corpus, therefore one can say our data is quite generic since it comes from 680 texts from all around the country. The information found in this thesis can be helpful to schools and teachers in the sense that they can help to tell which errors are most frequent among the different age groups.

Thornbury (2002, p. 19) states that when German-speaking learners learn the word "table", they are likely to create a link to their first language equivalent, which is the German word "tisch". When a pupil is learning a second language, they will create links between words that mean the same. In the same way a Norwegian pupil would learn that the word "table" is the Norwegian word "bord". The pupils will make these links between languages, which will make it easier for them to remember, but could also cause confusion. In Norwegian we use the word "høy" for the English words "high", "tall" and "hay". Having three equivalents for one word, can be confusing since some might struggle with understanding which of the equivalents to use in the correct setting. For example, one common mistake Norwegians make is to ask, "how high are you" instead of "how tall are you". There are lots of cases like this, and we found it interesting to research and study it to see what kind of errors we make as a learner.

Mahan & Brevik (2013) states that just because teachers of English know about common word errors does not mean that they know how to help pupils with avoiding these mistakes. Mahan & Brevik (2013) later states that it can be useful for teachers to know how to categorize these word errors and avoid them. In this study, we have looked through a lot of different word errors and labelled them. Teachers can for example use a similar category system as in our study to categorize the errors their pupils have made. When looking at English texts that pupils have written, it will be useful to categorize word errors when giving feedback to a pupil (Mahan & Brevik, 2013). If a teacher categorizes the word errors that the pupils are doing, they can not only have an easier time giving feedback to each pupil, but they

can also make teaching schemes that can help with the errors that most of the class does. It may be relevant to assess whether the word errors that the pupils make are due to them simplifying the language, they use informal or oral English or that they are not aware of certain differences between the L1 and L2 (Mahan & Brevik, 2013). It is important for the teacher to let the pupils know which errors that are most common in their writing and let them know how they can fix them. If a pupil is not aware of certain differences between the first and second language, they need to know how they can stop making these mistakes and not let their English become worse because of Norwegian influence. On the other hand, if pupils make errors by being informal where they use the verb "to get" a lot, the teacher can help them with vocabulary, but it is not a mistake that is urgent to fix.

It may also be important to know if the pupils end up choosing the wrong word because of influence from Norwegian which is called interlingual errors, or whether is due to them choosing the wrong word within the English language which is called intralingual errors (Mahan & Brevik, 2013). Tracking and categorizing the errors into interlingual and intralingual errors will help to keep track and set up lessons where the focus is on the most used errors. If a class struggles with interlingual errors, it may help to teach English vocabulary to increase their knowledge of English words that they may not currently use. On the other hand, if a class is struggling with intralingual errors, it would help to teach about the differences between the two languages and not to use Norwegian rules or words in the English language. Mahan & Brevik (2013) states that influence from the first language when talking in the second language is a trait that is difficult to get away from. When pupils struggle with a second language, they resort to using something they know more about, which is the first language. We hope that this study can help English teachers with spotting and categorizing both interlingual and intralingual errors and get knowledge about the most common mistakes that are made by pupils aged 12-13 and 15-16.

1.3 Research questions

Thesis question: Which types of vocabulary errors do pupils in Norway aged 12-13 and 15-16 make?

Research questions:

1. How is the distribution between lexical errors caused by interlingual and intralingual influence among pupils in Norway aged 12-13 and 15-16?
2. To what extent do pupils aged 12-13 and 15-16 make lexical errors that are influenced by Norwegian?
3. Which errors are most frequent among the lexical errors caused by transfer from Norwegian?

Two things that are interesting for educators or language researchers to know are how the L1 might influence learning an L2 (in this case English), and how the L2 learner might struggle with inner L2 issues. For this reason, we wanted to study the interlingual and intralingual mistakes done by two different age groups situated in Norway. By comparing these two groups we can observe a possible difference in interlingual errors and intralingual errors, which can give an indication of language development between the ages.

From an educational stance, it would be helpful to know what mistakes are the most frequent among English learners that are caused by the influence of the Norwegian language. This information can then be applied when teaching English so that teachers can help the pupils avoid these mistakes in the future.

2 Theoretical framework

The theoretical framework is categorized into three different parts. First, we will present how vocabulary is learned and how second language learning works. Later we will show what kind of mistakes learners make and introduce the two major categories of different error types. Lastly, we will present previous studies that are in the same field as our study.

2.1 How is vocabulary learned – second language learning

One important element of learning a language is to expand one's vocabulary. This is done by learning new words and putting them to use. "Language learning refers to developing language awareness and knowledge of English as a system, and the ability to use language learning strategies." (The Ministry of Education and Research, 2020, p. 2). Learning a new language requires the pupil to learn the system of the chosen language. Learning the English language requires the pupil to also acquire vocabulary, grammar, and the phonology of the English language. It is also important that the pupil learns how to communicate. It is important to have knowledge enough to understand and be understood by someone else. Below, we will address how vocabulary is learned, and the difference between learning a first and a second language.

Thornbury (2020, p. 18) makes a point that when children are learning their first language they use their first words for labelling, which is connecting words to concepts. The word *apple* must be connected to the concept apple, it is round and can be eaten. Not every round and edible object is an apple because oranges exist. Therefore, the child must learn where to extend the concept of an apple, to not include oranges. Thornbury (2020, p. 18) further states that acquiring a vocabulary also needs categorizing skills and recognizing that the word apple can be replaced with the term fruit. The next time the child learns what a pear is, the child can then put the pear into the fruit category with a different label than the apple. The label could be based on shape, taste, or color which will help separate the different words in the same category. Thornbury (2020, p. 18) points out that network building is when the child constructs a web of words, and interconnects words with each other, words like brother and family for instance, and it serves to link all labels. This will continue to expand when the child learns new words to put them into their vocabulary, and it will continue to do so for the rest of their lives. This is how vocabulary is learned for the first language, learning a second

language is different because the system on how to learn is already there, the learner must put labels on different words and categorize them in another language.

According to Thornbury (2020, p. 18), the difference between learning a first and second language is that the learners that are learning a second language already have the conceptual systems and the network of associations that link the words with each other. When a child learns their first language, they get a lot of input from their family. The child then put different labels into concepts as mentioned above. However, when a learner learns a second language they cannot just copy and paste the system that they used for their first language. "Learning a second language involves both learning a new conceptual system and constructing a new vocabulary network – a second mental lexicon." (Thornbury, 2020, p. 18). Words can have more meanings in some languages, and some words might not exist in other languages. An example of this is the Norwegian word "døgn" which means 24 hours. It is both day and night, and English does not have a word for it. The Ministry of Education and Research (2020, p. 2-3) also states that language learning refers to identifying the connections between the learner's first and second language and understanding how the second language is structured. It is important to not mix the first and second languages with each other since it can cause confusion and the learner might not be able to distinguish the structural difference between the two languages. There is one downside to having a ready-made conceptual system with its own associated lexicon, which is that the learner is likely to short-cut the process of constructing a network of associations (Thornbury, 2020, p. 18). The L1 may disturb the acquisition of L2 vocabulary, one example of this is when the learners start associating words that look the same in both languages, but do not mean the same. These words are called false friends and can cause confusion because learners may not have the knowledge or short-cut the process of constructing a network of associations and confuse an L2 word because of L1 influence.

Weinreich (1974, p. 9-10) argues that two different language systems can relate to each other in the mind of a bilingual learner. He believes there are three conceptualizations models that can be recognized in the relation between L1 and L2 lexicons. These three are:

The first: coordinate, in which two separate form-meaning links coexist in the mind of the learner.

The second: compound, where bilinguals have a single concept associated with two different words

The third is: subordinative, in which the L1 word form mediates between the L2 concept and the L2 word form

Weinreich (1974) also argues that the relationship between these is not static, but that they depend on the learner's experience with the L2 and that different types of lexical organization can coexist in the same mind.

Munoz discusses the connection between age and second language learning. Munoz sums up the empirical studies from the 1970s and states that:

> *(1) Adults proceed through early stages of syntactic and morphological development faster than children (where time and exposure are held constant)* (Munoz, 2006, p. 2).

> *(2) Older children acquire faster than younger children (again, in early stages of syntactic and morphological development where time and exposure are held constant)* (Munoz, 2006, p. 2).

> *(3) Acquirers who begin natural exposure to second language during childhood generally achieve higher second language proficiency than those beginning as adults* (Munoz, 2006, p. 2).

Older learners have a superior learning rate, particularly in the first stages of the acquisition of morphosyntactic aspects, while younger learners are slower at first, but eventually show a higher level of ultimate attainment (Munoz, 2006, p. 2).

2.2 What kind of mistakes do learners make?

Thornbury (2020, p. 29) categorizes errors into two different major types, which is form related and meaning-related. "Form-related errors including mis-selections, misformations, and spelling and pronunciation errors." (Thornbury, 2020, p. 29). We will now explain the form and meaning-related errors and explain how it relates to our own labels that we used in the dataset.

A mis-selection is when a learner uses an existing word that is similar in sound or spelling to the correct form (Thornbury, 2020, p. 29). In our dataset, we labelled mis-selections to the L2 vocabulary confusion label. For example, when pupils confuse the words "sad", "said" and "sat" with each other we put them into the L2 vocabulary confusion label. When the pupils use a word that already exists, but they confuse it for another word because it is similar to another word, then it is L2 vocabulary confusion. This is because pupils confuse their L2 vocabulary and choose the wrong word, they know about the word, but they may not have the knowledge yet to use it correctly.

Misformations often result from misapplying word formation rules, producing non-existent words and misformations often show clear influence from the learner's mother tongue (Thornbury, 2020, p. 29). The misformations were labelled as Norwegian elements in our dataset. We noticed cases where pupils produced non-existent words when they had trouble finding the English word for it. In most cases, it was clearly a result of Norwegian influence because when they had trouble finding the English word. They resulted in using the Norwegian word for it or trying to make the Norwegian word sound more English. An example of this was when a pupil wrote "onkol" instead of "uncle". The Norwegian word for "uncle" is "onkel", and in this example we can see that the pupil tried to make the Norwegian word "onkel" sound more English by replacing the "e" with an "o", thus producing a non-existent word.

"Meaning-related errors typically occur when words that have similar or related meanings are confused, and the wrong choice is made." (Thornbury, 2020, p. 29). The meaning-related errors were mostly labelled as L2 vocabulary confusion errors. An example that is discussed further later in the study is the overuse of the word "see" instead of "look". These are two words that have related meanings, which causes confusion and thus pupils end up making the

wrong choice. When looking at our dataset, we also noticed a few instances where some of the meaning-related errors were labelled as Norwegian element errors. "Meaning-related wrong-choice may derive from the learner's L1, where the meaning of an L1 word may not exactly match its L2 equivalent." (Thornbury, 2020, p. 29). We discuss the use of the English word "good" later in the discussion section, because we see many uses of the word "good" instead of "well". We believe this is because of the Norwegian word "bra", and how it is used in Norwegian, contrary to how the words "good" and "well" is used in English.

Agustín-Llach (2017, p. 65) argues that establishing the source or main causes of lexical errors in EFL productions will allow us to conclude some pedagogical implications for vocabulary instruction. Hence, she created a list of what she considered the most frequent and important error types in EFL (which are based on Agustín-Llach, 2011; Bouvy, 2000; James, 1998; Warren, 1982), below are the four most important ones.

> *1) Borrowings, which are bare L1 insertions into the L2 syntax* (Agustín-Llach, 2017, p. 65).

> *2) Lexical adaptation of an L1 word to the L2 morphological or phonological rules* (Agustín-Llach, 2017, p. 65).

> *3) Semantic confusion originates when the learner confounds two words which are semantically related in the L2* (Agustín-Llach, 2017, p. 65).

> *4) Learners also tend to calque L1 words or expressions when they lack exact lexical knowledge of the L2 equivalents* (Agustín-Llach, 2017, p. 65).

Agustín-Llach (2017, p. 65) states that learner tend to borrow L1 word, create lexical adaptations and calque L1 words or expressions when they lack lexical knowledge of the L2 equivalents. Agustín-Llach's most frequent and important lexical errors are many in the same category as we have in our study. We also investigate L1 borrowing but have called this category Norwegian word (NOR). The lexical adaptation of an L1 word to the L2 morphology can be compared with our Norwegian element error. The difference between how we did it and what Agustín-Llach did in this category is that Agustín-Llach only looks at L1 words that are trying to look more English. While our study investigates L1 words that are

trying to look more English, and English words that have Norwegian elements in them (English words that look more Norwegian).

The semantic confusion group Agustín-Llach mentioned, is similar to what we have in our study, where the pupil has chosen the wrong word for the wrong situation. Agustín-Llach (2017, p. 65) calls these errors phonetic or formal confusions. This is because some words look and might sound the same and therefore can be confusing for the learner. Agustín-Llach says

> "*Semantic and formal confusions reveal a certain degree of word knowledge, incomplete or imperfect knowledge, though. We might wonder whether the learner knows both the target and the error word, and confuses them because of their similarity or whether they ignore the target word and use a proximal, close word they have knowledge of*". (Agustín-Llach, 2017, p. 65).

The last group of errors Agustín-Llach mentioned which is relevant to our study is the "calque errors". This category is the same as we have, and Agustín-Llach describes them, as similar to what ours are.

2.3 Previous studies

In this section, we dive into previous studies, that are comparable to our study. These studies can help give a bigger understanding of linguistic influence and L1 influence among L2 learners.

2.3.1 Mahan (2013)

Previous research that is relevant to lexical errors and Norwegian influence can be that of Mahan (2013). She investigates how lexical errors are distributed in texts that are written by Norwegian speakers of English. She does this by using a taxonomy to find and determine patterns of lexical errors. Mahan conducts a search for errors in two corpora that are based on lexical dissonance. Lexical dissonance is when words create dissonance between the word itself and its context, rather than violating the rules of the language (Mahan, 2013). It consists of 49 031 words, where 1077 of the words have lexical dissonance in some way. The material in the study is taken from two separate corpora, one is written by Norwegian students at an intermediate level of English, while the other one is written by Norwegian students at an advanced level. The intermediate corpus consists of 24 732 words that are written by 18 participants at the age of 14-15 years old. The advanced corpus consists of 24 299 words that are written by 39 participants at the age of 19-21 years old.

When it comes to lexical dissonance, Mahan and we have different views on how to categorize these. Mahan has only accepted a very formal language, while we have allowed more verbal/informal English. When looking into her Examples we disagreed with how she had categorized them, as they could have been said in verbal English. A lot of words she would classify as vocabulary mistakes/lexical dissonance we have classified as non-lexical mistakes.

Mahan classifies lexical dissonance with the use of routes, effects, and influence to explain and label the errors that are found within both corpora.

Route is how learners choose the wrong word or in what way a learner selects the wrong word (Mahan, 2013, p. 31).

Effect is about what is wrong with the choice of words or about what **effect** does the wrong word have on the message (Mahan, 2013, p. 39).

Influence is about why the learner chooses the wrong words or what triggers the lexical dissonance, interlingual or intralingual (Mahan, 2013, p. 41).

In cases where two words mean the same thing, researchers and speakers will disagree about which words are "correct" and "incorrect" in different situations, which means that "lexical dissonance" does not have a conclusive answer.

In Mahan's results, we see that most of the errors created are cores-errors. These types of mistakes indicate that the largest lexical coping for Norwegian learners of English is using non-specific vocabulary. Mahan used these examples of core errors, using "like" instead of "enjoy" and "appreciate", and it represents 4 in 10 wrong words (Mahan, 2013).

The second largest route is called synonyms. Synonyms are when a learner uses a word that is somewhat related, but it is the wrong word (Mahan, 2013). Using the word "seemingly" instead of "apparently" or "going" for "walking" are two examples that Mahan (2013) uses when describing synonyms, synonyms are the cause of 1 out of 5 wrong words.

The third largest route represents 1 in every 6 words and is called transliterations. This is when a learner translates such as "get away (få vekk)" instead of "remove" and "to make a party (lage fest)" instead of "throw a party") (Mahan, 2013). Mahan's transliterations correspond to our calque label.

Cognate is the last route that occurred more often than 1 in every 10 dissonances, cognate occurs 1 in every 6 dissonances. Cognates such as "to mean that instead of "to believe", or "to have the impression" instead of "be under the impression" are examples of learners having words that interfere with each other between two languages because they are similar (Mahan, 2013). This is one of the examples where we have different views on what an error is, we disagree with Mahan and believe that writing "to have the impression" is correct and should not be marked as an error. Mahan's cognates correspond to our false friend label.

The second least frequent route is associations, which occur in 1 in every 14 dissonances. Associations are caused by a confusion of word form, such as "defiantly" for "definitely", or "the ocean raised" instead of "the ocean rose" (Mahan, 2013). Mahan's category of "associations" corresponds with our label of vocabulary confusion.

The least frequent route is perceived equivalents, which only occur in 1 in every 33 dissonances. It is caused when the L2 speaker does not know the word form or meaning of what is intended to say at all such as "to give a withering look" which is marked as an unknown intended meaning or misunderstanding the meaning of the word such as "to evolve something" instead of "to develop" (Mahan, 2013).

The largest effect in Mahan's results is called stylistic/connotational, and it is evident in over half of the dissonances. This suggests that a lack of knowledge of wrong words is related to register, which also includes inappropriate and most often informal words (Mahan, 2013). Mahan (2013) used these examples of dissonance caused by stylistic/connotational: "crazy" for "mentally ill" and "get" for "receive/find/obtain".

Semantic is the second largest effect and occurs in 1 in every 6 dissonances. Examples such as "the division of a hospital" instead of "wards" or "wood" for a "forest" proves Mahan's (2013) point that L2 speakers are not fully aware of the senses of the words that they use.

Collocation is the last effect that occurred more often than 1 in every 10 wrong words, collocation occurred 1 in every 6 dissonances. Examples of collocations such as "choose the right direction" instead of "make" or "increase your look" instead of "enhance" illustrates that the learners lack word combination knowledge (Mahan, 2013). This is another example where we do not agree with Mahan, we believe that there is nothing wrong with writing "choose the right direction" as it is highly used in informal English.

The second smallest effect is syntactic, which occurs in 1 in every 20 dissonances. Syntactic effects show confusion between words in word families or case relations, such as "get born" instead of "be born" or "a near-dying accident" instead of "near death" (Mahan, 2013). The smallest effect occurred 1 in every 20 dissonances as well, and it is called invalid. It is caused when the learner translates a Norwegian concept that does not exist in English such as "life-lie from Norwegian livsløgn" or creates a word that does not exist such as "consum for consumption" (Mahan, 2013).

As to influence, interlingual influence is 58%, while intralingual influence is 42% in the corpora. Interlingual influence consists of both strong and weak L1 influence, while intralingual influence is unrelated to the L1, and confusion is only within the target language

itself (Mahan, 2013). The intermediate students (14-15-years-old) show evidence of L1 influence in 62% of all dissonance, while the advanced students (19-21-years-old) show evidence of L1 influence in 54% of all dissonance (Mahan, 2013). It will later be discussed how Mahan's results compare to our results and see how age affects the L1 influence.

2.3.2 Mahan and Brevik (2013)

Mahan & Brevik (2013) wrote an article that is based on Mahan (2013). The article contains the most common English word errors, and how such errors can be recognized and avoided. Some words are combined more often than others, if a learner has heard a word combination many times, they are likely to use it themselves (Mahan & Brevik, 2013). Therefore, word choice is often determined by frequency and the bond between the words grows stronger the more it is used. If a word is used in combinations or context that is non-native, the word choice might be wrong, and a word error will occur (Mahan & Brevik, 2013).

Since this article by Mahan & Brevik is based on Mahan's (2013) study, we will inform that we do not agree with all the judgements that were made in their article. Some of the examples mentioned are from their article, while we do not agree that it is an error, Mahan and Brevik wrote it down as an error. This will be discussed further down on why we do not agree with their judgement.

Mahan & Brevik (2013) states that the reason for stylistic/connotation errors was due to the writing style that the learners used, it was characterized by oral language and inappropriate for academic writing. Stylistic and connotation errors made up 62% of all the lexical errors in the study, this might be because they count the verb "to get" as an error because other words such as "become" and "enter" might fit better and make the sentence more formal. Mahan & Brevik (2013) also mentions verbs such as "think", "like" and "make" as overused. Moreover, adjectives such as "nice", "good" and "big" is mentioned as well.

Half of the semantic errors were due to influence from Norwegian, and the rest of the semantic errors were due to individual errors made by each learner, and difficult to generalize (Mahan & Brevik, 2013). Sentences such as "the clock was almost eleven" are being used instead of "the time was almost eleven", and "she saw me in the eyes" instead of "she looked me in the eyes", were errors that the learners did (Mahan & Brevik, 2013). These are English words that the learners expect to mean the same as a Norwegian word.

17

The collocation errors were influenced by L1, and approximately half of the errors were due to learners combining the English words the same way they usually do while writing Norwegian (Mahan & Brevik, 2013). Mahan & Brevik (2013) mentions that verbs such as "to take surgery" instead of "have surgery" and "to take a decision" instead of "make a decision" is being overused. It is possible to use the verbs in Norwegian but requires other verbs when writing English (Mahan & Brevik, 2013). The use of adjective-noun-combinations that are possible in Norwegian but not in English was also a collocation error that learners made. The use of "plane accident" instead of "plane crash" and "an evil circle" instead of "a vicious circle" were two of the examples that Mahan & Brevik (2013) made.

The learners in Mahan's study did not have any issues with choosing the correct word class as they did with other word errors, they did also not have the habit of producing new words (Mahan & Brevik, 2013). The syntactic and invalid errors in their study only made up 9% of the lexical errors.

English word errors come from a lack of knowledge about the used word, and the main source of word errors are interlingual influence and intralingual influence (Mahan, 2013). They further state that when learners lack knowledge about a word, they resort to using their L1 as a starting point. This is what interlingual influence is, when the word errors are due to errors across languages, it is an interlingual word error. Mahan & Brevik (2013) describes intralingual word errors as errors that occur when a learner has problems within the L2. Moreover, Mahan & Brevik (2013) states that second language users get confused by the words that they can choose from and end up choosing the wrong word. Further, they state that when a learner is comfortable with words, they end up overusing them. Moreover, the learners end up underutilizing words that they do not know. The result of overusing a word such as "to give" can make the learner write "to give an advice", instead of "to offer advice" (Mahan & Brevik, 2013). Similarly, the lack of familiarity with low-frequency words can cause Norwegian learners to use high-frequency words instead. An example that Mahan & Brevik (2013) used is "on the edge of falling" instead of "on the verge of falling". The learners would rather use words that they are familiar with and use in their everyday language, rather than using more precise words and terms that they are not familiar with.

2.3.3 Hemchua and Schmitt (2007)

Hemchua and Schmitt done did a study that looks for a lot of the same lexical mistakes among learners as we have in this study. They studied the lexical errors in the English compositions of Thai learners. Their study was compiled from short texts (300-350 words) from third-year Thai university students (most of them were between the ages of 19-20). Hemchua and Schmitt`s study was more comprehensive than our study, as they look at many more error types than we have done. Some of the mistakes they investigated can be compared to our study. For example, mistakes influenced by the learner's L1 (Calque translations, borrowing L1 and coinage), false friends, and near-synonyms.

On average, each paper contained 13.05 errors, which equates to one error per 26.46 running words (Lexical errors) (Hemcua and Schmitt, 2006, p. 14). In total, they found 261 errors in the texts. Their largest group of mistakes is the "Near synonyms mistakes", which can be compared to our vocabulary confusion mistakes. In total, the near synonyms made up 51 out of 261 (19,54%) mistakes.

They also did have a look at the calque errors (L1 translation), this was their 4 largest group of mistakes, and the calque errors made up 18 of the 261 mistakes (6,9%). They did only find three occurrences of false friends, zero occurrences of borrowing (L1 words), and zero occurrences of coinage (inventing new English words influenced by the L1). Hemchua and Schmitt (2007, p. 20) state that they calculated that 23.75% of the errors could be reasonably attributed to L1 influence. It must be stressed that identifying the underlying cause of errors is inexact and problematic (Hemchua and Schmitt, 2007, p. 20)

3 Method

This section of the thesis focuses on the methodology and procedure of the study. First, we will discuss how this study takes a quantitative approach, then we will discuss using secondary data as a research method, and last, we will explain the procedure this study has taken.

3.1 Quantitative study

This study takes a quantitative approach, where the idea is to be able to generalize the findings from this research. A quantitative study is suitable for finding information from a vast number of subjects. An example of this can be if a teacher wants to find out what activity the pupils want to do during a class. The teacher can give the pupils an opportunity to vote for A, B, or C, and then the teacher can map out which of the activities most pupils want to do during that class. This would then be a quantitative study done by the teacher. Cohen et al (2018, p. 725) state "Quantitative data analysis is a powerful research form. It's often associated with large-scale research, but can also serve smaller scale investigations, with case studies, action research, correlation research and experiments.". When researchers must justify their approach and method, it is common to point out that a quantitative approach includes a larger sample and thus provides breadth to the data material (Gleiss & Sæther, 2022, p. 196).

Public archived linguistic data and statistical code have great pedagogical value for the community of linguistics (Janda, 2013, p. 8). Data from different researchers pooled together can give a greater insight into how we deal with linguistics and what one can do to improve one's grammar. Different quantitative studies can be used and compared to give a bigger understanding of something that is happening. Garshol (2019, p. 125) states that the choice to apply a quantitative methodology to the main variable(s) aims at capturing patterns that are frequent in the data, and not only the interesting exceptions to the rules.

Our data are something called descriptive data. Cohen et al (2018, p. 727) emphasise that descriptive statistics do exactly what they say; they describe and present data. Our data has been collected by using categorical variables. This means that the matches we have collected have been sorted within the given categories. This would later allow us the opportunity to

later compare and observe if any of our data correspond with each other and the data other researchers have found.

3.2 A corpus study/Secondary data

Our study takes its information from a corpus, and therefore in addition to being a quantitative study also is a corpus study or a secondary data study. According to Heaton (2008), secondary analysis can be used to test hypotheses, generate new knowledge, and to support, challenge, and extent existing theories or findings. This is something our study aims to do, as we want to use the pre-existing Corpus to generate new knowledge about the Norwegian influence in English, in Norwegian schools.

Firstly CL (corpus linguistics) provides data for language studies by supplying huge amounts of examples about a specific structure and linguists can work out how a structure is used looking at these occurrences. In this way corpus [sic] is really good tool for gathering data about language studies and it provides huge amounts of data which would take a lot of time with traditional data collection methods. (Atar & Erdem, 2019, p. 141)

In addition to this, a corpus is a good tool to gain insight into the changes in languages between different groups. A corpus can be used to help compare the language difference between different age group, different L1 influence in a language, or different social groups. Some corpora like the CORYL corpus might also be a good tool for teachers and schools. The different corpora can be used to find the most frequent mistakes that learners do in an L2 language. In the CORYL corpus, one can search for a set age group, search for either categories or words, and then find out which errors are the most frequent ones. This can then give an indication of for example what the 12-year-olds struggle with the most, if it is the spelling of a certain word or if it is "*was"* and "*were"* mistakes.

Using the CORYL corpus gives us the opportunity to create this study, since getting approval from NSD, and coding all pupil texts would cost too much time and resources and would have forced us to create a lot smaller study with just a few texts.

However, there are some challenges that arise when using secondary data, some of the data can for example be inaccurate or outdated. Therefore, it would be important for the researcher to evaluate on beforehand the quality of the secondary data. Another challenge that using

secondary data has, is that the researcher has no control over the data collection process. This can then limit the researcher from being able to use the data in a meaningful way.

3.3 The procedure of this study

We decided to use the CORYL corpus as our source of data collection. CORYL corpus is a database that consists of texts written by pupils at the age of 12-13, 15-16, and 16-17. The 16-17 age group has around 200 texts, but we will be focusing on the 12-13 and 15-16 age groups. The texts are taken from Norwegian National Tests in English between the years 2004 and 2005. The corpus has been sustained and texts have been added to the corpus up until 2017. The CORYL corpus uses a coding system that separates errors into codes, which makes it easier to categorize and work with.

We started by looking at the different codes that are used in the CORYL corpus and discussed which errors that would be relevant to use for our data. In the end, we decided to use the code "NW" (nonexistent word), "WW" (wrong lexical word), and "NOR" (Norwegian word). With these codes, we could look for errors that pupils made when they used the wrong word and used a word that does not exist in the English language. Our aim was to focus on finding errors that Norwegian pupils made because they would mix their first language with the second language. We also wanted to see where they just wrote Norwegian words as a substitute for the English word, which is why we included the code "NOR" as well. The "NW" and "WW" codes did not have Norwegian words, only Norwegian elements within them, which is why we also had to take the "NOR" code to find all the Norwegian words.

To search in the CORYL corpus we had to use a search expression with the code that we wanted to use. To search for our code, we had to type "[type = "WW"] \ <>" and replace the "WW" with "NW". Afterward, we used "NOR" as well and used these three codes. After that, we downloaded the concordance into an Excel spreadsheet with the age and year settings shown for easier categorizing.

After downloading it to an Excel spreadsheet we opened six tabs where we had two tabs each of WW, NW, and NOR errors, one for each age group. When we downloaded it from the corpus, we had to filter by age to separate it into different tabs. After that, we were done with setting up the spreadsheet and could start labeling the different errors that the pupils made.

Before we started with the labeling of the errors, we had to agree on how many, and which labels we wanted to have. We ended up with ten different labels that would help us with categorizing the errors that the corpus picked up and labeled them ourselves. We had labels for calque error, false friend, indeterminate, vocabulary confusion, Norwegian element, spelling error, abbreviation, sound, non-lexical mistake, and Norwegian word. We started labelling together to get used to it and discuss every sentence, this way we could later work alone and have the same thought process as each other. We will now explain every relevant label that is discussed in our study and show a few examples of what the pupils wrote.

Calque error – a word or phrase in one language is incorrectly transferred to another language by a literal translation of each of the individual elements. One example of a calque error that we got was "cleaninghelp". Here the pupil just translated "rengjøring/vaske" and "hjelp" and put them together in one word, when the correct word would be "maid", or the more gender-neutral "cleaner". Another example is the sentence "I just love the kingcastle", as we can see in this sentence, the pupil wanted to write "kongeslottet" in English. "Kongeslottet" is often used as a verbal term for "det kongelige slott" which translates to English as the royal palace. This pupil just tried to transfer each individual element of "kongeslottet" and ended up with "kingcastle" which is not a word.

False friend error – a cognate word used incorrectly in the L2. An example of a false friend error that we got was the use of the word "character", which resembles the word "karakter" in Norwegian. The pupil used the English word character when talking about grades, and the Norwegian word for grade is "karakter". "I was so lucky when I heard that the police came" belongs in the false friend category as well. In this example, the pupil used the English word lucky when happy would have been better. In Norwegian the word "lykkelig" is close to the English word lucky, and "lykkelig" translates to happy. In this circumstance, the pupil could have mixed the words and used the wrong word, because using the word lucky in this context does not make any sense.

Indeterminate – the meaning of the word and the source/cause of the error is unclear. We used this label when we were unsure of what the pupil meant or tried to write.

Example 2: the hous was the a snacke this coler was orange and yellow. _azakelig_ theer lising we sombody was under the hous.

23

The example above is put into the indeterminate label and the word that the corpus marked in this sentence is "azakelig". After reviewing the sentence, we could not find a label that fits the marked word. Therefore, the word ended up in the indeterminate label because we could not find out what the pupil tried to write. Another example of a sentence that was labelled as indeterminate can be seen below.

Example 3: A sofa, A cactus, A picture, some bokcses, A board, A table *hamens*, A ball, some socks, red, A bierdhous.

The word that showed up as a match was the word "hamens", we could not find the meaning of the word, or understand it in a contextual sense, therefore we had to labelled it as indeterminate.

L2 vocabulary confusion – incorrect use of L2 words, which is not due to transfer from the L1, often involving the confusion of two words in the L2. One of the errors that was labelled as L2 vocabulary confusion can be seen below.

Example 4: On the wall it is a picture of a cat. It is an goldfish in a *bottle*. Hello Nils Right now i'm in New York. It is many things to.

Here the pupil wrote that the goldfish was in a bottle, when looking at the task the goldfish is inside a fishbowl. The word "fishbowl" is not a frequent word that gets used a lot in the English. However, the word "bottle" is used more frequently. Maybe the pupil did not know or forgot about the word "fishbowl", or just thought that "bottle" meant the same as "bowl". Another example of L2 vocabulary confusion can be seen below.

Example 5: ook my bag and get myself home crying because the picnic was *broken*... When I came home I didn't cry any more.

Here the pupil has mixed the word "broken" with the word "ruined", the picnic itself cannot be "broken", since the word is more used in the context of an item being destroyed. The word "ruined" would fit better, since a picnic can be ruined by for example rain. For this reason, this mistake has been labelled as vocabulary confusion.

Norwegian element – insertion of a Norwegian word or word component, or it looks like the Norwegian word. An example of a pupil with a Norwegian element error would be writing "clatring" instead of climbing. The word "clatring" is just an attempt to make the Norwegian word "klatring" sound like English. Another example of a Norwegian element error is when pupils write. Another example of a Norwegian element can be found in the sentence below.

Example 6: VI most *undersoeke*. Nei ikke naa jeg maa ite

The word highlighted here is the word "undersoeke", what we believe the pupil tried to write here is the Norwegian word "undersøke". What it seems like in this case, is that the pupil tried to make the word sound more English, by switching out the letter "Ø" which is used in the Norwegian alphabet, with an "oe" ending. This is also done, in for example flight tickets, when someone has a name that contains the letter "Ø", it then says "OE" in the name.

Norwegian word - the pupil uses a Norwegian word in the text. There will be English words that the pupil does not have in their vocabulary and a lot of times they choose to just write the missing word in Norwegian. An example of this happening is found below.

Example 7: I just started painting the *hytte* green.

Here we can see that the pupil is writing the whole sentence in English except for the word "hytte" because he does not have the English word cabin in their vocabulary. This was one of the most common words that the 12-13 age group struggled with when looking at the Norwegian word label. Another example of a Norwegian word error can be seen below.

Example 8: I woke *opp* in my own room

Here is another word that they struggle with. They might not have the word "up" in their vocabulary, so they end up using the Norwegian word instead. We separate Norwegian element errors and Norwegian word errors because Norwegian word is only when the learner writes a word in Norwegian and does not try to make it English. This can have an impact on communication because people from other countries will not understand the meaning of the Norwegian word in a conversation.

We categorized all the errors in the "NW", "WW" and "NOR" errors in CORYL into the categories described above. After we were done with the labelling, we made one more page in our Excel document where we summarised all the labels. We counted all the labels and separated them into the different age groups 12-13 and 15-16. This way we could see the difference between the two age groups and see which errors occurred the most often in which age group. We made two tables for the different age groups and wrote down the number of errors that every label had for that specific age group. Both tables were summarised as well, which allowed us to make a pie chart. With a pie chart, we could clearly see which mistakes that were made more often in which age group. This was useful for our findings and to look at the difference between the age groups. Afterward, we also made a bar chart where we wanted to see mistakes per 1000 words. First, we had to find out how many tokens were in the different age groups. We found this information on the CORYL website. After finding out how many tokens there were in each age group, we had to start to calculate the correct number of words per age group. We did this by taking the total number of tokens minus every error code (<SIC>, <DIV> and ETC.), signs (for example: /, "", !.), and numbers so that we were only left with only the numbers of words. This had to be done for both groups (the 12-13-year-old pupils and the 15-16-year-old pupils)

Picture 1, List of words for the pupils aged 12-13

Picture 1 above, shows the total number of tokens (Treff) in the left corner, and some of the number of error codes and signs that had to be removed to be able to find the correct amount of words in the corpus. For example, we can see from the table that there are 36 354 cases of " </SIC>" that had to be removed. Removing all of these is how we found that the group of 12-13-year-old had written 109 179. We also can observe from picture one which of the words were the most frequently used among the 12-13-year-olds.

After finding out how many words there were in the corpus, we could make the bar chart to see mistakes per 1000 words. With the information on the number of total words in the corpus, we could make these charts to easily comprehend the results that we got from our analysis.

3.4 Research quality

In this section we will discuss the validity and reliability of this study, later we will also discuss the ethics of the study.

3.4.1 Validity

Validity is a huge concern when creating a quantitative study. Creating data and research questions that are coherent with each other can be difficult. The validity of a study is all about how well the different parts of the study fit together. Gleiss & Sæther (2022, p. 204-205) mention that in a quantitative study high validity will imply that one measures what one wants to measure. Cohen et al (2018, p. 246-247) state that in much quantitative research, validity often (not always) strives to be faithful to several features, for example, Controllability, replicability, consistency, predictability, the derivation of generalizable statements of behavior, randomizations of samples, neutrality/objectivity, and observability.

Validity might be the most important part of a study, which is if our data and research questions are coherent with each other. The data we have found show the change in Norwegian influence within English grammar, between the ages of 12-13 and 15-16. The data shows how the usage of Norwegian Elements, vocabulary mistakes, calque errors, false friends, and Norwegian words between the ages and what mistakes connected to Norwegian influence are the most frequent for the different groups.

Now the next step here is to find if our thesis question matches our data and if they are relevant to each other. Our thesis question was: *Which types of vocabulary errors do pupils in Norway aged 12-13 and 15-16 make*? One way we have made sure that our research questions answer the data we have found is that we have remade the questions throughout the analysis process to make sure that the research questions can be answered by the data. One of our initial thesis questions was: "What challenges do Norwegian (L1) pupils in Norway have between the ages of 12-13 and 15-16 when it comes to English and what challenges can impose when it comes to communication". When we were starting to look at the corpus, we wanted to use we soon found that it would be harder to have a specific way to look at the mistakes that the Norwegian pupils with Norwegian as an L1 had made. We realized that there were other students with other L1 languages who had participated in the National tests, and therefore if we had continued with that thesis question our research study would be less valid. For this reason, we had to continuously keep changing our thesis and research questions for every implication we met, so that the thesis and research questions truly could be used up against the data we have collected.

Another central part of a good study is controllability. To be able to check/control one's study against someone else's study can give a good indicator that the study is valid. If the study completed had similar answers/findings as most other studies in the same field, the study would be more valid (but not necessary). However, if one's study has completely different findings than other studies within the same field, the validity and methods can be questioned. If this happens the researcher, then must try to prove why he or she has gotten the findings he or she has. Our study has a limited offer of other studies to compare to. The other studies in the same field have not necessarily used the same methods or have had the same questions to answer as our study. However, other studies can still be used to compare with and control our study.

Some of the studies our study can be compared to are Mahan and Brevik (2013), Mahan (2013), and Hasselgren (1994). Brevik and Mahan and Mahan's other study show the semantical mistakes that students in the ages 15-16 and 19-21. Hasselgren's study shows some of the same as our study, but with a group of other ages than our study. These semantical mistakes they have found can be compared with one part of our study; the vocabulary confusion mistakes. But these studies do not have the same focus on Norwegian elements as

our study does and therefore can be difficult to compare them with. But their findings and conclusion are similar to our study when it comes to Norwegian influence. The controllability within our study is to some degree lacking in the sense that there are not many studies that focus on how Norwegian can influence when learning English, especially between the ages 12-13 and 15-16.

3.4.2 Reliability

Reliability is all about the quality of the research study and if the research is trustable. According to Gleiss & Sæther (2022, p. 202), the two questions that are important to answer when it comes to reliability are.

1. How has the data been influenced by the way it has been collected?
2. Can the results be reproduced by other scientists?

The first question here is how our data is influenced by the way we have collected it. Our data came from three codes in the CORYL corpus, we then downloaded the data and transferred it into an Excel format. At this point, we have not influenced the data. But, after downloading it we then erased all er error codes where it just stood *<SIC>*. After this, we were able to give labels to the matches we had gotten and summarize all the codes in the end.

When labeling the codes, we also made sure of control checking our data to ensure it is reliable. We used some time after labeling all the matches to cross-check the data between the two age groups. We did this by jumping between the NW-codes between the two age groups and checking that the words that were the same, were coded the same. Then after we did the same thing with the WW-codes and the NOR-codes. Doing this would then ensure that the data is more reliable, and the reasoning for this is that it removes some of the possibility of unintentionally creating biased data. It must be mentioned that not all the words were coded the same between the two age groups, because in some cases to try to understand what the pupil had meant, we had to look into the context of what was said.

Another important element of research and the second question Gleiss & Sæther asked is if it is possible to replicate the research, Miceli (2019, p. 13) states that "One of the ways that scientists confirm the validity of a new discovery is by repeating the research that produced it". Our study would be highly replicable if one would just use the same codes and labels (use

the same description as in this text) when it comes to identifying the mistakes. This would then show approximately the same results as we have found with a small amount of difference. Here it is important to mention that some of the "matches" we got could look like both a spelling error, a Norwegian element, or other elements. What the pupils' have written could in some cases be hard to determine a label for. For example, in the sentence below, it is difficult to find a correct label to use.

Example 9: When a tell my freiNnd I found doog and he <u>sad</u> are then broun and I So yes.

 If the students mixed the word sad and said, since the words might sound the same it would then be labelled "vocabulary confusion" (VC) but if the pupil just forgot the letter "i" in the word said it would then be coded "spelling error" (SE). This example shows that even if someone tried to replicate our study, we believe they would mostly get the same results. However, there will be some differences in results since other researchers might choose a different label than we would in some of the cases.

Miceli (2019, p. 13) says: "Several factors can contribute to non-reproducibility or non-replicability, including previously unknown variation or effects, inadequate recordkeeping, technology limitations, potential biases, lack of training, institutional barriers, or even misconduct, in rare cases.". When it comes to our study results can change over time, the CORYL corpus has several times been updated and it has been added more texts to it, it has also had a change of people who have coded the corpus and changed the person responsible for the corpus. Different people might have different opinions on which words belong to which code, and therefore be an unknown variation.

3.5 The ethics of the study

When conducting this quantitative study there were some ethical concerns to consider. When working with analyzing data it is important to have in mind that we must follow the guidelines (labels) created, to ensure that the data is as fairly created as possible. Cresswell in Cohen et al (2018, p. 139) states that the researcher has an ethical duty to ensure that the results of the research are reported fairly, credibly, and accurately, without misrepresentations and unfair selectivity (exclusion and inclusion). We have tried our best to fit the data we found into the categories we thought fitted the best, by using labels we created, though this is for others to decide how we did.

Dealing with sensitive data is something we also had to consider. In this research, we are getting our results and data from previous students` texts (from the National Test), and it is important to keep the identity of the individual secret. One of the positives of using the CORYL corpus is that it has already made everyone anonymous for us, and there are no real names mentioned in these. All the pupil`s texts are therefore untraceable, and their identities are hidden. Even though the texts and sentences found in the corpus have been anonymized we still have to show respect and dignity to the participant's work. This is important since promoting trust and cooperation in research will both be beneficial for the researcher and the participants. If the researcher does not respect and dignity of the participants, his or her paper would look less serious and the participants would probably not want to participate in any more studies, which will then be a loss for future research.

4 Data information and information about the Corpus

This chapter focuses on three different parts of the study. The first is basic information about the CORYL corpus. The second data type and specific information were sent to us. The third was the tasks the pupils were given.

4.1 The CORYL corpus

> *Coryl* [sic] *is a young learner corpus, which consists of English texts written by Norwegian pupils. The texts which make up the corpus, were collected in the course of the National Testing of* **English** *(writing), 2004-5, at the University of Bergen, and are taken randomly from pupils in 7th, 10th and 11th grade. The texts are anonymized and have been assigned to levels and half levels on the CEFR by multiple raters.* (Clarino Centre Bergen, without year).

To clarify what the Clarino Centre Bergen said, the "National Testing of English (writing)" is the National Test. The tested was carried out by individual schools, but the material for the corpus has been collected and coded by the University of Bergen by a set group of researchers.

4.1.1 The intention behind the corpus

Hasselgren and Sundet (2017, p. 1-2) state that the aim of compiling CORYL was to allow researchers to investigate the interlanguage of the learners. Hasselgren and Sundet (2017, p. 2) further state that the CORYL corpus is not well suited to research the individual's interlanguage, but rather expose language patterns when looking at groups. "The texts are coded for all errors and are also linked to the Common European Framework (CEFR) levels. This means that students can search for instances of the errors produced at different CEFR levels, for the various age groups." (University of Bergen, 2023).

4.1.2 Useful elements of the corpus

One advantage the CORYL corpus has is that it as a secondary data source is much bigger and more thoroughly created than something a single or a couple of scientists or researchers could create alone.

> "*Using secondary data has many attractions. For example, the scale, scope, and amount of data are usually much larger and more representative than a single*

researcher could gather, and the large scale and scope of data may be more robust at a level of complexity not available to smaller-scale research." (Cohen et al 2018, p. 587).

The corpus has several features that are useful to find a huge range of information. In the corpus, one can have a look at the different codes, 24 in total. Examples of some codes that can be found in the corpus are found below.

Codes we used:

NW: Nonexistent or unintelligible word (excluding clear verb- or word-form errors) (Hasselgren & Sundet, 2017, p. 214)

WW: A wrong lexical word (i.e. not of a closed class) (Hasselgren & Sundet, 2017, p. 215)

NOR (L1): Norwegian word. (Hasselgren & Sundet, 2017, p. 214)

Examples of other codes that can be found in the CORYL corpus:

> **SMS**: SMS or slang, e.g. L8 or wanna (Hasselgren & Sundet, 2017, p. 214)
> **WFO**: Wrong form of word (excl. verb), e.g. plural or –ly ending. (Hasselgren & Sundet, 2017, p. 214)
> **IT**: It/there errors (Hasselgren & Sundet, 2017, p. 215)

When searching the corpus one can sort for age groups, gender, tasks, and nationalities (some). In the corpus, it can also be found a list of words, where it is numbered how many times a certain word has been mentioned in the corpus, which tells us the most frequently used words by the pupils. When searching in the corpus one can also find the distribution of words of each age group, gender, and more.

4.1.3 Limitations of the corpus

When we started working with the corpus, we experienced that the CORYL corpus has its limitations. We discovered that some of the information on the webpage of the corpus had not been updated. For example, the total numbers of words in the corpus were not updated to what the actual number was, and some of the written information about the corpus also seemed to be incorrect/outdated. The contact person for the corpus had gone off with her

pension, and there was no one else to contact. Also, the web pages connected to the corpus did not work. A lot of the information we have gotten has been through mail correspondence with the new contact person of the corpus. This person was able to give us quite a bit of information about the corpus, including the CEFR gradings, the tasks given to the pupils, and a bit of explanation about the corpus. For example, what a token is, how many texts there are from each age group, and how many sensors had been working on the corpus. While working with the corpus, the Clarino webpage has been updated. Among other things, the information about the corpus has been updated and the layout has been updated. However, some of the connected web links on the page are not working, and there is still some information that is outdated such as the number of tokens and contact person.

In chapter 4.1.2 about the useful elements in the corpus, we mentioned that it is possible to find for example the numbers of words for each age group, which is possible, but the number mentioned here also includes a lot of *sic*`s and other error codes. These numbers we found to be misleading, and therefore we had to remove all the error codes ourselves to calculate the correct number of words.

4.2 Data type and information

The data that is found in the chapter below is from the comparison of two age groups. The first age group is 12-13 years, and the second is 15-16 years. All the data found in the corpus used have been gathered from the National Tests in English between 2004-2005. In the National Tests, the students were tested in both reading and writing (writing has now been removed from the National Tests). Through mail correspondence with Eli Moe (24.11.22), we were informed that all the texts in the CORYL corpus were gathered from all around Norway. We were also told that the corpus consists of 472 texts from pupils in the ages 12-13 (7th. grade) and 212 texts from pupils in the ages 15-16 (10th. grade).

The age group 12-13 had written 109 179 words in total (231 words/pupil) and the age group 15-16 had written 91 065 words (429 words/pupil). In total, the parts of the corpus we have used then consist of 200 244 words. This specific data we had to calculate ourselves, since this information was not to be found in the corpus. (How we calculated this is found in the procedure of the study).

4.3 Tasks given to the pupils

This section focuses on the tasks given to the pupils, from 2004 to 2005, for both age groups. The tasks can be found in the appendix.

4.3.1 7th grade (age 12-13) from 2005

The 12-13-year-old pupils (in 7th grade) had three tasks they had to complete in the Norwegian National Test in English (writing). They were to write on the given lines on the paper and they had 60 minutes to complete the tasks. The introduction text suggested using around half of the time on the third task, probably since it is the largest part of the National Test.

In the first task, the pupils were asked to look at and describe what they could see in the picture. In the picture, there are a lot of things happening and a lot of items to describe. The pupils were given eight lines to write on.

In the second task, the pupils were to write a postcard from one of three places depicted in the task. The pupils were to write where they are, what they are doing, and what they like/dislike about the place they are visiting. They could choose to write from a campsite, a city, or a harbor.

In the third task, the pupils were asked to choose between task A) or B) and write three to four paragraphs. In task *A,* the pupils were asked to start their text with the sentence *"One day when I came home from school, I found the front door wide open.".* In task *B*, the pupils were asked to write their text using a given picture. The picture depicted some boys building a treehouse and a snake sneaking its way up the tree.

4.3.2 10th grade (age 15-16) from 2004

Like the pupils in 7th grade, the pupils in the 10th grade also had three tasks to complete. Different from the pupils in the 7th grade the pupil in the 10th grade were given more time to complete the tasks given to them. The 10th grade pupils were given 90 minutes to complete all the tasks, and they were recommended to spend half of the time on the third task (the longest task).

In the first task, the pupils were supposed to imagine that they are going away for the weekend and that they must write an e-mail to some friends to get them to join. To help the

pupils it is mentioned they can write about where, when, and what the others should bring in the example. In this task, it is not mentioned how much the pupils are supposed to write.

In the second task, the pupils were asked to write for the school paper about their stance on the school`s pending decision to introduce separate classes for boys and girls. The pupils were asked to write one or two paragraphs for the school newspaper, arguing for their view on the case, but showing that they can understand both sides of the issue.

In the third task, the pupils had to choose between task A) or B). In task A the pupils are asked to; Write a text of five to eight paragraphs with the title "The next ten years of my life". In task B the pupils were asked to; Write a text of five to eight paragraphs, finishing with the sentence: " …. That's what friends are for.".

4.3.3 10th grade (age 15-16) from 2005

Same as the 10th grade from 2004 and 7th grade from 2005, the 10th grade pupils from 2005 were also asked to answer the three tasks given. As in the 10th grade test above, the pupils were asked to complete all the tasks within 90 minutes and were also advised to spend half their time on the third task (the longest task).

In task one, some English/American friends are coming to visit. The pupils were asked to write a <u>short</u> letter to them, telling them a place they must see while visiting Norway. The text gives the pupils examples of things to write such as, what…, where…. and why… someone should see while visiting.

In the second task, the pupils are supposed to write a response (one to two paragraphs) in response to a headline that occurred in the newspaper. The headline stated that *"Young people today care less about the environment than adults".* In the task, it also stated that the pupil's response should contain arguments to convince the reader of their point of view.

In the third task, the pupils had to choose between <u>either</u> task A) or B). Task A) stated: *"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.".* Task B stated: "*Write a text, 5 – 8 paragraphs, with the title: "The day everything went wrong".*

5 Findings

In this chapter, we will present our findings. First, we will dive into the mistakes made by the 12-13-year-olds, then the 15-16-year-old. After this, there is a comparison between the two groups, before we present the most common errors among each group.

5.1 Does the influence from the Norwegian language and vocabulary mistakes differ between the ages of 12-13 and 15-16.

In this section, we present we will present the date found. Below, there are two tables presented, one for the age of 12-13 and one for the age of 15-16. These tables show two things, 1. How many errors the age groups have made, and 2. How many mistakes are produced per 1000 words the pupils have written.

The abbreviations in the tables have the following meaning:

CE stands for "Calque error"

FF stands for "False friend"

ID stand for" Indeterminate"

NE stands for "Norwegian element"

SE stands for "Spelling error"

VC stands for "Vocabulary confusion"

NOR stands for "Norwegian word"

5.1.1 Labels from the 12-13-year-old pupils

In this section of the master thesis, the focus is on all the individual labels for the ages 12-13. All the examples shown in each section are gathered from the CORYL corpus. (The words that came up as a match are put in cursive and underlined in the examples).

Labels from the 12-13-year-old pupils		
Code	Number of occurrences:	Mistakes/1000 words
CE:	34	0.311
FF:	3	0.027
ID:	394	3.608
NE:	288	2.637
SE:	773	7.080
VC:	474	4.341
NOR	836	7.657
Total:	2802	25.66
Total number of tokens:	207714	
Total number of words:	109179	109.179

Table 1 Number of codes/labels by the learners aged 12-13

5.1.1.1 Calque Errors 12-13-year-old pupils

From *Table 1*, we can observe that the 12-13-year-old pupils have made 34 calque errors. This then equals one mistake per 3211 words, or 0,311 mistakes per 1000 words. This is not a frequent number of errors, but the mistakes seen here are clearly connected to the Norwegian language in the sense of the pupils have directly translated both elements of the word. Below there are some examples of exactly this.

Example 10: the police and then the murder was taken. We found uss a new *cleaninghelp* , and everything went good.

Example 11: e brigade with my new cellfone. Five minutes later a big red *firecar* came and took the snake, they put him in a bag and one of t

Example 12: noting. I go in tu the badroom there was noting! I go to the *sleproom* and there was books and t-shirts and shoe and socks, yes th

In *Example 10,* the pupil most likely translated the Norwegian word "Vaskehjelp" into "cleaninghelp" by translating each element of the word. The pupil has translated both elements "vaske" and "help" into English to create an English translation of the word. What should have been written by the pupil is the word "cleaner". This is an example of how some CE mistakes occur and how the pupils are influenced by Norwegian.

In *Example 11* we can observe the same happening thing happening as in the first example. The word "firecar" have been directly translated by the pupil, who has translated each of the elements "brann" (fire*)* and "bil" (car) from the Norwegian word "brannbil".

In *Example 12* it is possible to observe that the pupils have done the same mistake with directly translating each element from the Norwegian word "soverom*"*. In this example, we can also observe that the pupils also have a spelling error in the word/element of the word "sleep".

5.1.1.2 False Friends (FF) 12-13-year-old pupils

When it comes to the false friend label, we could only observe 3 occurrences of this type of mistake within the 12-13-year-old pupil's texts. This then tells us the mistakes only occur 0,027 times for every 1000 words. An example of a false friend mistakes can be found below.

Example 13: olice becuas the sceary man was waiting downstears. I was so *lucky* when I heard that the police came. We went downstears and I

Example 14: h a goldfish in a boil. The dog is very tired. It is a shous *back* the TV, and a football behind the TV. Hello!! I have that w

In *Example 13*, the pupil chose to use the word "lucky" as in the Norwegian word "lykkelig", which translates to "happy". With the context in this sentence, we can see that the pupil mixed

the words, and did not mean to write that they were lucky that the police came, but rather happy.

In *Example 14:* we can see the pupil using the adverb *back*, where it should have been used preposition *behind* instead, and therefore this mistake can be labelled as a false friend. The Norwegian word for "behind" is "bak", and this can cause confusion for the pupils. The Norwegian preposition "bak" and the English adverb "back" have similar meanings, which can mean that the function of "bak" may have been transferred to "back".

5.1.1.3 Indeterminate errors (ID) 12-13-year-old pupils

The indeterminate label we had to use 394 times during this study. This then equals 3,608 times per 1000 words. What this can tell, is that quite a huge part of the matches gathered from the corpus were difficult to classify for several reasons. The indeterminate label is quite broad, since it contained words that were difficult to fit in other codes, were not understandable, or contained different signs not regularly used. Examples from the corpus can be found below.

Example 15: clock on the wall to see how much time I had left before my *carpool* arrived.

Example 16: at home naw saw if you gow home naw nowt be skird it is the *suer* man.

In *example 15,* which is from one 15-16-year-old pupil, we see that the pupil has written the word *carpool,* from the context surrounding the word it would be sufficient rather use the word *car*. Why it has been written *carpool* in the text is unknown and the word is hard to classify, and therefore it has been put under indeterminate.

In *example 16,* the pupil has used the word *suer*, which is not in the Norwegian language or the English language. What the pupil could have meant was to translate the Norwegian word *sur* or the pupils could just misspell the English word *sour.* What the pupil meant by the word is unknown and therefore we have classified this mistake as an indeterminate error.

5.1.1.4 Norwegian Elements (NE) 12-13-year-old pupils

As shown in *Table 1*, the number of Norwegian element errors created by the 12-13-year-old pupils are 288. This then equals 1 mistake per 379 words written by the pupils, or 2,637 mistakes per 1000 words. Below there are three examples gathered from the corpus.

Example 17: a nice plan. And so we started Adam begin to hent planks to <u>bank</u> in. And Adam hent glas we can bank in in the planks so we g

Example 18: ome clothes and runs to school. Today we should get back our <u>examen</u>. As I walks into our classroom everyone eies falls on me, b

Example 19: ut that we should build a treehouse. I had a hammar and some <u>spikers</u>, Jack had an orange saw and John had the planks. We went ho

By looking at *Example 17,* it is possible to observe that the pupil has used the word *bank,* which is an English word that is used wrong here. When looking at the context the pupil is not talking about a bank, they are talking about hammering nails into wooden planks. The pupil has used the Norwegian verb "å banke", which in this context means "å hamre" and tried to make it sound more English by removing the last letter in the word.

In *Example 18*, we can observe that the pupil has done the opposite as the pupil from *example 17.* This pupil has used an English word and put a typical Norwegian ending on the word. The correct way would be to write *exam* and not *examen*, so this is an example of Norwegian Elements found in the pupil's English.

Example 19 is another example of how some pupils try to make Norwegian words "more" English. In this example, the pupil has written the word *spikers*, "spiker" is the Norwegian word for nails. What the pupil has done in this example is to use the Norwegian word "spiker" and add the English plural -S at the end, to try to make the word more English.

5.1.1.5 Vocabulary confusion 12-13-year-old pupils

The vocabulary confusion section is the third biggest category found in the age group 12-13. We found 474 occurrences where we believe the pupils have used the wrong word or should have used another word. This then equals to 4,341 mistakes per 1000 words.

Example 20: ed sofa. The man is watching TV. On the purple tabel it is a <u>cup</u> with fruit. There is a goldfich on a green bookshelf back i

Example 21: see a books. I can see a TV. I can see a window. I can see a <u>blossom</u>. I can see a gold fish. I can see a people. And many niece

Example 22: mp (ned Fra twree. I are so skerw of (slange) She is 3 meter <u>big</u>. end haw a yellow (Prikker pÃ¥ seg) We (gikk) hom too mee, e

In *Example 20* we can observe using the word "cup", this word does not seem to fit in the sentence. By context (looking at the task given to the pupils), we can observe that on the table there is a bowl filled with fruit, and not a cup, so therefore, we assume that the pupil has mixed up the words "cup" and "bowl".

In *Example 21,* we can see the pupil using the verb "blossom", where it should have been used the subject "flower" instead, therefore this mistake has been labelled as vocabulary confusion (VC). The word "blossom" is often associated with the word "flowers" and that is probably the reason why the pupil has made this mistake.

In *Example 22*, the pupil has used the wrong adjective for describing the person. Where the pupil has written *big*, where it should instead be written the word *tall,* which is better when it comes to describing a person's height. Therefore, the mistake from *Example 22* is also labelled as a VC mistake.

5.1.1.6 Norwegian words (NOR) 12-13-year-old pupils

The largest group of labels we have found is the group labelled NOR. This label contains all the words written in Norwegian from the age group 12-13. In total, we found 836 examples of pupils using Norwegian words in their texts. This then becomes 7,657 Norwegian words per 1000 words from their texts.

Example 23: They sad I <u>skulle</u> open the present, and I do that. But the was a football there, not a bombe, the was a football as I

Example 24: acks father is a painter so we got paint of him. John was hammering spikker. Because his father is a <u>snekker</u>, my fatehr is a mailman.

Example 25: tings to do it was a litel plac som a can play wid my football and a dislike ol the mygg sov wi use <u>*myggspray*</u>. THE TREHOUSE I, Dag Haakon and Anders tray at bild op a trehouse on one of the big tre in Dag Haak

Example 23 and *Example 24* both contain Norwegian words, the reason for why these mistakes is occurring is probably due to the pupils not knowing the English word for what they are trying to say. In *Example 23* the pupil had written the word *skulle* in Norwegian, when he should have written the English word *should.* In *Example 24,* the pupil had written the word *snekker* in Norwegian, while the pupil should have written *carpenter* in English.

In *Example 25*, the pupil has used the Norwegian word "myggspray", probably since he or she did not know the English term "mosquito spray/repellent". We assume the pupil did not know these words since they are not a high frequent word and are content words.

5.1.2 Labels from the 15-16-year-old pupils

In this section of the thesis, the focus in on all the individual labels for the ages 15-16.

Labels 15-16-year-old pupils		
Code	Number off occurrences:	Mistakes/1000 words
CE:	20	0.219
FF:	12	0.131
ID:	154	1.691
NE:	107	1.174
SE:	237	2.602
VC:	381	4.183
NOR:	61	0.669
Total:	972	6.853
Total amount of tokens:	141832	
Total amount of words:	91065	91.065

Table 2 Number of codes/labels by the learners aged 15-16

5.1.2.1 Calque Errors (CE) 15-16-year-old pupils

The pupils aged 15-16 made 20 calque errors, which is not a lot. This equals to one mistake per 4554 words or 0.219 mistakes per 1000 words. These mistakes we have found are containing elements from the Norwegian language, since we can see that the pupils have translated each element of some words, from Norwegian directly to English, and so creating calque errors.

Example 26: he kids or the adults? If a little kid see an adult toss his *gumpaper* on the ground. What would he do? He would toss it on the gr

Example 27: uys arive Norway with love. Paul Johnsen I forgot my english *wordbook*   and diddent understand the word environment. Bush is a bad

In *Example 26* we can observe that the pupil has probably tried to write the words "gum wrapper", which in Norwegian would translate to "tyggegummipapir". The pupil has here

translated each of the elements "tyggegummi" and "papir" directly into English in the hopes of getting the correct word.

In *Example 27*, the pupil has done the same as the pupil in the previous example, he or she has directly translated both elements "ord" and "bok" from the Norwegian word "ordbok" into English. By doing this the pupil then created the word *wordbook*, where instead the pupil should have written the word "dictionary".

5.1.2.2 False Friends (FF) 15-16-year-old pupils.

When looking at the number of false friends, we can see that there is not a lot of them. In the age group 15-16 we only found 9 occurrences of this type of mistake. This equals 0,131 mistakes per 1000 words.

Example 28: ve moved out, and my little brother can not talk I hope. The *character* on a test I have last week, was wrong. I could not answer i

Example 29: buy some nice clothes and new make-up. No, it was 15 minutes *again*. She sreemd for a taxi and jump in. When she came from she

Example 30: 21 years now, and my sister, mother and father died on this *fly*. I think in this thing ho happent with us wen we just was o

In *Example 28*, we see the pupil has misunderstood the meaning of the word "character". The word "character" might from a Norwegian stance look like the word can be the English translation of the Norwegian word "karakter", which it is not. Because of the similarities between the word in Norwegian and English, the pupil has probably therefore been "tricked" into believing they were the same, but instead, the pupil should have written the word *grade*.

In *Example 29*, we see the pupil using the word "again" wrong. The pupil has probably confused the English word "again" with the Norwegian word "igjen". The Norwegian word "igjen" can be used as the English words "again" and "left". This causes confusion for the pupil, and they make the mistake.

In *Example 30*, the pupil mistakes the English word "fly" for the Norwegian word "fly" which translates to "plane". The English word "fly" translates to the Norwegian word "flue",

45

and this will confuse the pupil because the words are the same, but they have different meanings across the languages.

5.1.2.3 Indeterminate errors (ID) 15-16-year-old pupils

In total, there were 154 cases where we struggled with finding a fitting label for the matches from the group of 15-16-year-old pupils. This equals 1,691indeterminated matches per 1000 words.

Example 31: the bad people come after oss whit this Guns, and they begynte *Ã¥* skyte then they traff 3 of my friends, i started to cry.

Example 32: ator and turned left in on a road in all to high speed. So i *kuraces* into a car and hurt my leg a quiet bad and I broke my arme.

In *Example 31*, the word that came up as a match was *Ã¥,* the meaning behind this word and from context is difficult to find, therefore, it has been labelled indeterminate. In *Example 32*, the pupil has written the word *kuraces,* which is neither a Norwegian word nor an English word. What the pupil had intended to write is also difficult to know by looking at the context, and therefore is labelled indeterminate (ID).

5.1.2.4 Norwegian elements (NE) 15-16-year-old pupils

In the corpus, we could find 107 occurrences of Norwegian elements from the texts from the 15-16-year-old pupils. This equals 1,174 Norwegian elements per 1000 words.

Example 33: bach would help, but when I came outside it suddenly stared *regning*   .
Why? Why today? On the evening, when I so gone sleep, I th

Example 34: ara A! Ida co to hor rom and skriming. " The day can not bee *werre* "But yes then can! Wen she came in to hor rom, she see, th

Example 35: ne hour untill it started. I couldent walk home, cause I had *wricked* my ankle when I jumped off the road so I wouldent get hit b

In *Example 33*, we can see the pupil has written the word "*regning",* what the pupil wanted to say was the English word *raining* (can be understood through context). What we believe to have happened in this case, is that the learner has used the Norwegian verb "regne" and added

46

the English suffix "-ing". In other words, she or he has used Norwegian vocabulary combined with English grammar, and thus this mistake has been labelled as a Norwegian element (NE).

In *Example 34,* the pupil has tried to make the Norwegian word "verre", into English by using the letter *W* instead of the letter *V.* The correct sentence would be "The day could not become worse", thus the pupil should have used the word "Worse".

In *Example 35*, we can see the pupil doing the same thing as the pupil from *Example 26.* The pupil has taken the Norwegian word "Vrikke" and tried to make it sound more English by adding letters that are more commonly used in English. What the pupil should or could have written in this example is "because I *sprained* my ankle…".

5.1.2.5 Vocabulary confusion (VC) 15-16-year-old pupils

The 15-16-year-old pupils have made 381 vocabulary mistakes, where they have used the wrong word. This then equals 4,183 mistakes per 1000 words. Examples of Vocabulary mistakes can be seen below.

Example 36: place you simply must see, " GaldhÃ¸ppiggen ". It is Norway's <u>*biggest*</u> montain. On the way up you can by the way see the beautiful

Example 37: enough, so I went down to make breakfast. Suddenly the phone <u>*called*</u>. It was Lisa. She told me to get over at her house as quick

Both *Example 36* and *Example 37*, can be used to show that sometimes the pupils can be confused by the meaning of some words. In *Example 36*, the pupil tried to tell us that the mountain Galdepiggen is the biggest mountain in Norway, which can mean a lot, such as: is it the biggest mountain when it comes to mass (volume) or is it the biggest mountain when it comes to height. A better suggestion here would be to describe the mountain as the tallest mountain in Norway, as it would be a more specific and clear description.

In *Example 37*, the pupil has written "suddenly the phone called", which is not the correct word to use in this situation. One can use the word call, when a person is calling someone, but when someone calls, the phone rings at the other end. Therefore, a more correct word to use here would be the word "rang" as in "the phone rang". In this case, the pupil might have believed that the words *call* and *ring* had the same meaning since in Norwegian these words,

can be replaced by just one word "ringe". Hemchua and Schmitt (2007, p. 21) point out that a single word in L1 can correspond with two or more in L2 and this can cause confusion for the L2 learner, they called this divergent polysemy. This is what could have happened in *Example 37.*

5.1.2.6 Norwegian words (NOR) 15-16-year-old pupils

The pupils aged 15-16 had just 61 occurrences where we could find whole Norwegian words. This then equals 0,669 Norwegian words per 1000 words. Examples of Norwegian words we found can be seen below.

Example 38: And many kids collect money to different objects. -So <u>*derfor*</u> I think it is wrong to say things like that and write it in the news- paper. The day everything wen

Example 39: John Arne Riise has played a good <u>*sesong*</u> and has goaled lot off time. He is playing very good and is making supporters happy and is helping

Both *Example 38* and *Example 39,* show how some pupils tend to use Norwegian words in their English texts, which probably is due to lack of vocabulary. In *Example 38,* the pupils have used the Norwegian word "derfor", where instead the pupils should have written the English word "therefore". In *Example 39*, the pupils had written the word "sesong" in Norwegian, where instead it should be written "season" in English.

5.1.3 Comparing the differences between the two age groups

This section of the Master thesis has a focus on the differences between the two age groups of pupils. The focus lies on identifying and highlighting the differences in labels per 1000 words and in percentage between the two groups. This will help give insight into the differences in lexical mistakes made by the 12-13-year-old pupils and the 15-16-year-old pupils.

Diagramtittel

Label	Age 12-13	Age 15-16
CE:	0,311	0,219
FF:	0,027	0,131
ID:	3,608	1,691
NE:	2,637	1,174
VC:	4,341	4,183
NOR	7,657	0,669
Total	18,584	8,071
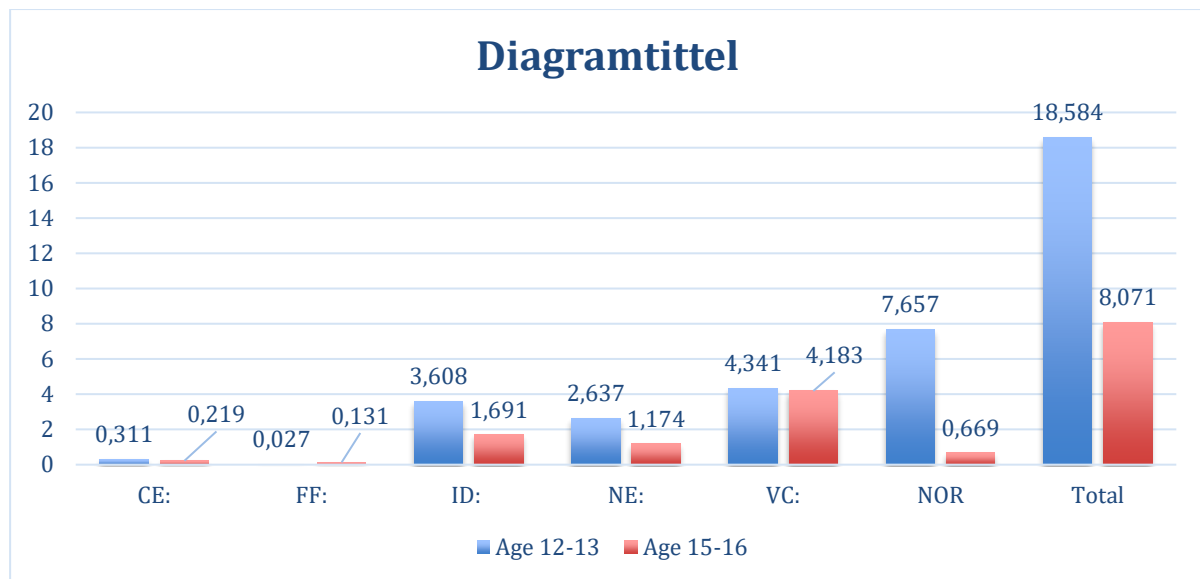
Figure 1 Comparison lexical errors 12-13 and 15-16

As shown in *Figure 1*, the 12-13-year-olds make lexical mistakes more frequently than the 15-16-year-olds, and this concerns most of the error types. In total, we found 18,5 matches we had to label for per 1000 words they had written, while the 15-16-year-old pupils had much fewer, only 8 matches that needed to be labeled for per 1000 words.

When looking at the tables and then the labels that contain Norwegian elements (NOR, CE, and NE) we can observe that the 12-13-year-old pupils have a lot higher frequency of mistakes than the 15-16-year-old pupils. The biggest difference is the NOR label, here it can be observed that the 12-13-year-old pupils used 7,65 Norwegian words per 1000 words (this equals one mistake per 130 words), while the pupils aged 15-16 used 0,66 Norwegian words per 1000 words (This equals one mistake per 1492 words), which is a huge difference.

Words containing Norwegian elements (NE) have also a lot higher frequency within the group of 12-13-year-old pupils. The 12-13-year-old pupils made 2,63 mistakes that contained Norwegian elements per 1000 words (one mistake per 379 words), while the 15-16-year-old pupils made just 1,17 mistakes per 1000 words (one mistake per 851 words). This means that the pupils aged 12-13 have more than doubled the number of mistakes compared to the pupils aged 15-16.

When it comes to the calque errors, we can read from *Figure 1* that the pupils aged 12-13 made 0,31 calque errors per 1000 words (one mistake per 3211 words) they had written and that the pupils aged 15-16 made 0,21 mistakes per 1000 words (one mistake per 4553 words) they had written. The pupils aged 15-16 made fewer calque errors than the pupils aged 12-13 by approximately 30%.

The VC label does not directly contain Norwegian elements, such as we are looking for, but is relevant for our discussion later in this master thesis. When looking at the VC produced by the two age groups, the results are quite even. The 12-13-year-old pupils produce 4,341 vocabulary mistakes per 1000 words (one mistake per 230 words), and the 15-16-year-old pupils produce 4,183 vocabulary mistakes per 1000 words (one mistake per 239 words). The error frequency is surprisingly even between the two age groups, given that the frequency of the other error types is much lower among the 15-16-year-olds. However, the proportion of L2 vocabulary confusion errors is much larger among the 15-16-year-olds than among the 12-13-year-olds.

The false friend label is quite small, we could only find three matches containing false friends for the age group of 12-13, while we found 12 matches containing false friend mistakes among the 15-16-year-old. The frequency of false friend mistakes is extremely low for both groups. In the age group 12-13 it only occurs 0,02 false friends' mistakes per 1000 words (one mistake per 36 393 words), and in the age group 15-16 only occurs 0,131 mistakes per 1000 words (one mistake per 7588 words). From this data, it is difficult to be able to interpret the differences between the two age groups, since they occur too rarely.

Below there are presented two pie charts, these show the proportions of the various error types for the two groups. *Figure 2* represents the proportion of labels found within the age group of 12-13 and *Figure 3* represents the proportion of labels found within the age group of 15-16. The information these two charts can give when comparing them is that they tell something about how the change ratio in percentage between the two groups is.
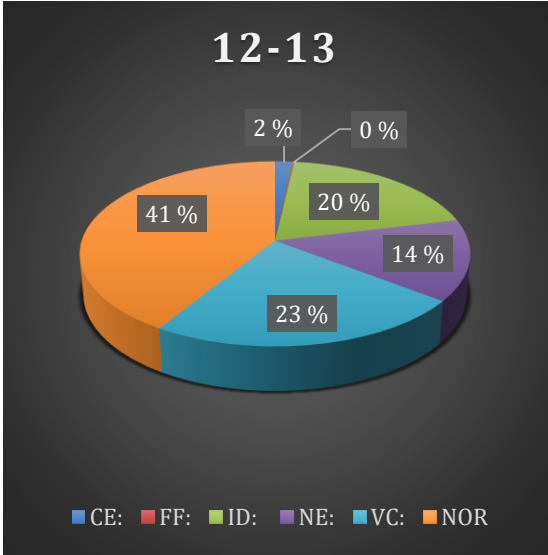

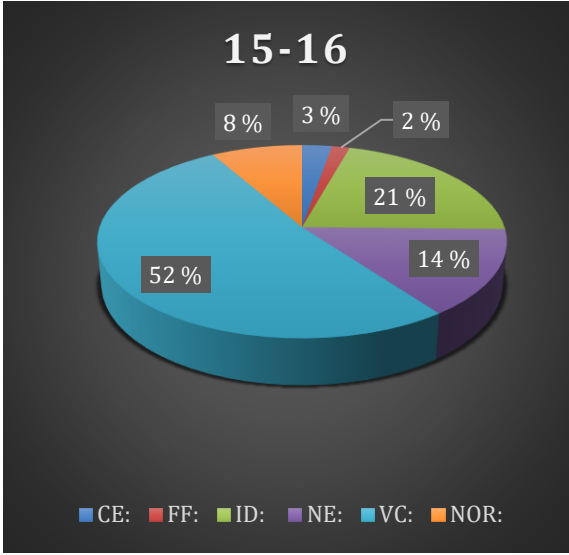*Figure 2 Distribution of mistakes among the 12-13-year-olds*

Figure 3 Distribution of mistakes among the 15-16-year-olds

51

Figure 2, which focuses on the labels for the 12-13-year-old pupils, shows that the largest label in percentage is the Norwegian Word label (NOR). This label makes up 41% of the pie chart, which is surprisingly much. The second most frequent error type among the 12-13-year-olds is the vocabulary confusion label (VC). This error makes up 23% of the chart. The third most frequent category is the indeterminate label (ID), this makes up 20% of the chart. The Norwegian Element (NE) label makes up 14% of the chart, while the calque errors (CE) make up 2% of the chart. The false friend (FF) errors are so few that they only make up for a few decimals of the chart.

When looking at *Figure 3* which focuses on the error types for the 15-16-year-old pupils, we see that the pie chart is different when comparing it to the pie chart of the 12-13-year-olds. The most frequent mistake here is by far the vocabulary confusion label, this makes up 52% of the pie chart, thus over half of the mistakes that the 15-16-year-olds make. The second most frequent mistake by looking at the pie chart is the indeterminate label (ID), this makes up 21% of the chart. The third largest group of mistakes is the Norwegian elements errors (NE), this type of mistake represents 14% of the chart. The Norwegian word label (NOR) makes up only 8% of the chart, while the calque error (CE) makes up 3% of the chart. The least frequent mistakes are the false friend (FF) mistakes, they only make up for 2% of the chart.

By comparing these two pie charts, it is easy to observe that there is a huge difference between the results of the data. The biggest differences seen in percentages can be found when looking at the Norwegian word (NOR) label. This label makes up 41% of the chart from pupils aged 12-13, while it only covers up to 8% in the chart of pupils aged 15-16. The difference between the groups is then as high as 33% when comparing the pie charts.

The second largest difference in percentages can be found when looking at the vocabulary confusion (VC) label. The vocabulary confusion label makes up 23% of the pie chart for 12-13-year-old pupils, while it makes up 52% of the chart when it comes to 15-16-year-old pupils. This is then a difference of 29% between the groups. When looking at *Table 1* above, we see that most of the other labels when it comes to the 15-16-year-old are lower in numbers than the 12-13-year-old labels, except the vocabulary confusion label. This label is quite even

between the groups and can be used to explain why the vocabulary confusion mistakes have a higher percentage than the 12-13-year-olds.

The third largest difference between the two groups can be found when comparing the false friends (FF) label. It is not a signific difference in percentage between the groups, the difference in percentage when comparing the charts is just 2%. The false friend label makes up just a few decimals of the mistakes created by 12-13-year-old pupils, while it makes up 2% of the mistakes created by 15-16-year-old pupils. The total amount and percentage of false friends increase with age when comparing these two pie charts. It should be mentioned that most of the false friend errors that were made by the 15-16-year-old pupils were when they wrote "fly" instead of "plane". The 12-13-year-old pupils did not have the same task, which mean that they were not expected to write about traveling and had no reason to make use of the word plane as the 15-16-year-old pupils did.

There is not much difference in percentage when it comes to the Calque error (CE) and indeterminate (ID) labels. Here it is only a 1% difference from the 12-13-year-olds to the 15-16-year-olds. The indeterminate label within the age group of 12-13 is at 20%, while the group of 15-16-year-olds is at 21%. The calque errors among the 12-13-year-olds only made up 2% of their mistakes, while at the 15-16-year-olds the calque errors made up 3% of their mistakes. Lastly, the Norwegian elements (NE) label ended up with no difference between the 12-13 and 15-16 age groups when talking about percentage.

The lexical errors that the two different age groups make are caused by either interlingual or intralingual influence. How is interlingual and intralingual influence distributed between the two age groups? First, we must look at the labels that we have used, we have divided our labels into interlingual and intralingual influence. The interlingual errors consist of the calque error label, Norwegian word label, Norwegian element label, and the false friend label. Meanwhile, the intralingual errors consist of only the L2 vocabulary confusion errors and the indeterminate label is miscellaneous and does not consist of either interlingual or intralingual influence.

When looking at *Figure 2* and *Figure 3*, we can see that the 12-13 age group has a combined 57% interlingual influence, while the 15-16 age group has a combined 27% interlingual influence. This informs us that the younger age group relies more on their first language than

the older age group. We also see that the 12-13-years-olds have 41% Norwegian word errors, which means that they just write the Norwegian word instead of trying to write English. The 15-16-years-olds only have 8% Norwegian word error, which tells us that they are way less likely to write a Norwegian word in an English text. The younger pupils both rely on their first language and use it while writing English way more than the older age group. If we just look at the age difference it seems that when pupils reach 15-16-years-olds they have learned to avoid Norwegian while writing English, thus making the difference in interlingual influence between the 12-13 age group and the 15-16 age group double in size.

Figure 2 and *Figure 3* show that the 12-13 age group has 23% intralingual influence, while the 15-16 age group has 52% intralingual influence. In our data, we only had L2 vocabulary confusion as a label to determine intralingual influence. By looking at both age groups we see that the older age group has more trouble with errors within the English language. There can be many reasons why this happens, and it will be discussed later in our study. However, it can be noted that the group of 12-13-year-olds has more problems than the group of 15-16-year-olds when it comes to L1 and L2 confusion, thus making relatively fewer errors than the older age group when it comes to intralingual influence. A percentage of 52% intralingual errors for the 15-16 age group might seem worrying, but it might be a positive development.

The group of 12-13-year-olds had 20% indeterminate and the group of 15-16-year-olds age has 21% indeterminate. We chose not to count the indeterminate label as an interlingual or intralingual influence. This is because we feel that the indeterminate label contains a lot of errors that cannot be categorized as an interlingual or intralingual influence. The indeterminate label contains a lot of nonsense and errors where the source of the problem is unclear. This label is used to make the other labels more valid and not spread unclear errors throughout our data.

5.2 The most frequent mistakes

This section is divided in two, the first consists of the most frequent errors influenced by Norwegian among the 12-13-year-olds. Then secondly, we will present the most frequent errors influenced by Norwegian among the 15-16-year-olds.

5.2.1 Most frequent mistakes among the 12-13-year-olds

The table below shows the three most common lexical errors among the 12-13-year-olds influenced by Norwegian. The table shows the name of the mistake, the number of mistakes, and it shows what label we have given the mistake.

Nr.	Mistake	Number of mistakes	Label given
Nr. 1	The *The*-mistake	106	NE
Nr.2	The *Hytte*-mistake	31	NOR
Nr.3	The *Opp*-mistake	24	NOR

Table 3 The most frequent mistakes among the 12-13-year-old pupils

Table 3 shows the most that the most frequent mistake among the 12-13-year-olds was the *the*-mistake. This mistake occurs 106 times, and considering that there are 472 pupil texts, we can say that one in five has made this mistake at least once. The pupils used the *the*-mistake where it should have been written the word "*it"* instead. We believe this is because the word "*the*" is used in the same way/place as the Norwegian word "*det*" is used in Norwegian, and the reason for this might be that they sound phonological almost the same. One interesting thing found when looking for the incorrect use of "*the*" is that when it came to the 12-13-year-olds we found 106 incorrect uses, while when it came to the 15-16-year-olds we found only 2 incorrect uses of the word.

The *Hytte*-mistake and the *opp*-mistake are both Norwegian words and are probably due to a lack of a fitting vocabulary in English. The Norwegian word "hytte" was used in total 31 times and the word "opp" 24 times. This means that these mistakes were quite frequent among the pupils. The hytte-mistakes we believe are due to one of the tasks given to the

pupils where they were asked to write 3-4 paragraphs about a picture (found in the appendix: task 3B, 7th grade, 2005). In the picture was a treehouse, some kids, and a snake, and we believe that most of the *hytte*-mistakes are related to this task due to the context around the *hytte*-mistake. This then probably means that the pupils were struggling with finding the correct word to use when it came to the word "treehouse" and substituting it with the Norwegian word "hytte".

5.2.1.1 *The*-mistakes

One of the most frequent mistakes we found among Norwegian pupils, was mistakes related to the use of "the". As explained earlier in the theory section, this is one of the most frequent words in English. In total, we found 106 examples of where "the" was used incorrectly among the 12-13-year-old pupils, and we believe these mistakes are highly influenced by the Norwegian language. Examples 34-36 below show the incorrect use of "the" found among the 12-13-year-old.

Example 40: I have <u>the</u> verry god I like my verry god her.

Example 41: I go to my friend. <u>The</u>   was none some skriiking help. I go up, I now the are none here. the was a mouse here. the are a sna

Example 42: Now are I in campingien. Her are <u>the</u> boring. I have (FÃ¥tt) now friends. They's name is John, Peter ands Mary. We playing footboll and we

In *Example 40,* the pupil had used the word "the" not as a definite article, but rather as a pronoun. Here the learner has apparently word-by-word translated the Norwegian idiom "ha det bra", which is not idiomatic in English. In *Example 41*, "the" translates to the Norwegian presentative "det", which in English corresponds to the existential "there". The same thing can be said about *Example 41*, in this example, the pupil has also used the word "the" as a subject.

In *Example 42* we have found the same mistake as the two previous examples, but the sentence also has the wrong word order in the sentence. The pupil has here directly translated each word from Norwegian to English without fixing the word order. In Norwegian the sentence would be (if it is interpreted correctly) "Her er det kjedelig."(Here is it boring), which is possible to say in Norwegian. In English, the correct saying would be "It is boring

here". The word order in Norwegian would be Adverbial + verb + subject + adjective, while the correct English word order would be subject + verb + adjective + adverbial.

What can be seen happening here is that the sentences start to be clearer and more understandable (except nr. 3, because of the word order). This shows that the pupils do mix the words "it" and "the" to some degree. This led us to wonder if this could have a more phonological explanation for why pupils use the word "the". What we do suggest has happened in these situations are that the pupils believe the word "the" and the Norwegian word "det" have det same meaning, as they do sound familiar. If one were to pronounce the "th" in "the" as /d/, "the" would sound like the Norwegian word "det". This would almost be like a false friends mistake but in a more verbal/phonological sense. This is then an interlingual mistake made by the pupils are the mistake are created by a language mix-up between Norwegian and English.

5.2.1.2 Usage of the Norwegian word "Hytte" among 12-13-year-olds

The usage of the Norwegian word "Hytte" is quite big, when looking into the corpus. In total, we found 31 examples of where the whole Norwegian word "Hytte" was used. In addition to that, there are also several examples of pupils mixing Norwegian and English words when it comes to compound words such as treehouse, wooden cabin, log house, etc. Below are examples of both using the Norwegian word "hytte" and mixing the languages when it comes to compound words.

Example 43: Lots of Love One day I and my friends found out at we will have a picnik. So we begin and bygge the *hytte* or house in the tree and 2 oures later was we ferdig so we took a kurv with a Icecrem and much more

Example 44: raing (PrÃ¸vd) a big boat and i (Fikk) it (til) we sees. Hilsen Sindre! My two friends and i made a (*Hytte*) in a three and i eat pizza and drink cook (Cola) and had it good wen (mens) my friends worked. We

Example 45: mother and dad go to shot the snake. Then was the snake angry longer and we can go up to the house (*Hytta*) agan (igjen) and work! we shift (Skifta) name and cold it the (SjÃ¸rÃ¸verhytta) to the three friend.

Example 46: We want have it green. It hud be a clubbhous. We gow op the *tree-hytta* and we see a sneick. It was big, orange and yellow. It has

Example 47: at shcool we heard a boy named Adam talking about building a *wood-hytte* in a big tree in his garden. John, Peter and I don't like A

In examples 43, 44, and 45, we can observe the pupils using the Norwegian "hytte", where instead it should have been written "treehouse" or "wood cabin". In *Example 43*, the pupils showed some uncertainness when it comes to writing the word "treehouse". The pupils wrote "Hytte or house in the tree", when trying to refer to the word "treehouse". First, the pupil wrote the Norwegian word "hytte", which the pupil probably knew was incorrect, and then he or she tried to give an English explanation of the "three house" by writing "house in the tree".

Another example of this can be seen in *Example 44* above. The pupil used the Norwegian word "hytte" and the pupil seemed to show that he or she did not know the English equivalent of the word. The pupils have indicated this by putting the word "hytte" in quotation marks.

In *Example 45*, the pupil used the Norwegian word "hytta", where it should have been written, "treehouse" (we assume this is from task 3B, from the national test from 2005 for 7th grade pupils). The pupil in this case also indicated that using this word is incorrect by putting the word in quotation marks, and this can be seen in several instances in this example.

In *Example 46* and *Example 47*, the pupils have translated the first element of the compounds. In *Example 46,* the pupil wrote "tree-hytta", the pupil translated the first element "tree" from Norwegian into English, but probably did not know the English word that would be the correct equivalent for the Norwegian "hytte". The same thing goes for *Example 47*, but in this example, the pupil wrote "wood-hytte", where the first element is "wood", and the second is the Norwegian word "hytte". In this case, the most suitable word to use would again become "treehouse", if we look into to context of the rest of the sentence, where the pupil mentioned that the "wood-hytte" is in a large tree.

These are all just some of the examples when it comes to the usage of the Norwegian word "hytte", among 12-13-year-old English learners. But since these mistakes are so frequent, it can indicate that the pupils are lacking vocabulary equivalents of the Norwegian word "hytte" in their English, and therefore use the Norwegian word. Something that is surprising is that

the pupils are self-aware of this mistake. Several pupils have used quotation marks or parenthesis around the word, to indicate that this is not correct.

5.2.1.3 Usage of the Norwegian word "opp" among 12-13-year-old

In total, we found 24 instances where the pupils have used the Norwegian word "opp" in the corpus. The English translation of the word "opp" from Norwegian would be the word "up". "Up" can be used as an adjective, adverb, and as a preposition in English, depending on the context around. The 24 examples we have found, show how the pupils have used the word "opp" as adjectives, adverbs, and prepositions. Below there can be found examples of the use of "opp".

Example 48: e claw and punch the head of the stupid Girl. The head fly away and land in my fang! and then i woke *opp* in my own bed it was a nithmear â€¦

Example 49: ling. And he Say yes but you are doing thet to morrow. The next day. We Startet to male, And we hang *opp* one pirrat flagg. Peter tok his meal now. thet was pitZZa.. thet koming a Snake, deady kome and hel

 In *Example 48*, the pupil used the Norwegian word "opp" where it should be the adverbial "up" in English. In *Example 49*, the pupil has also used the Norwegian word "opp" and should have used the English preposition "up". This is because the flag is hung "up".

We did not find any adjective usage of the word "up", only adverbial, and prepositional uses of the word. The results are higher than we would expect, as we believed the English word "up" should have been known by the pupils since it is frequently used. What also could have happened in these instances is that the pupils believe this word is written "opp" as it has phonetical similarities to the word "up", so the pupils might have the word "up" in their English vocabulary but have rather written it in Norwegian because of the phonological similarities. If this is the instance it would be an interlingual mistake, but in this case it can be difficult to interpret what exactly the underlining meanings have been in each of the situations.

5.2.2 Most frequent mistakes among the 15-16-year-old

Below is a table that shows the most frequent mistakes made by the pupils aged 15-16 influenced by Norwegian.

Nr.	Mistake	Number of mistakes	Label given
Nr.1	The *fly*-mistake	9	FF
Nr.2	The *at*-mistake	5	NOR
Nr.---- (Shared third with many others)	The *again*-mistake	2	NE

Table 4 The most frequent mistakes among the 15-16-year-old pupils

The group of 15-16-year-old pupils had much fewer frequent mistakes than the 12-13-year-old pupils. Their most frequent mistakes were the *fly*-mistake which occurred 9 times. In this case, the pupils used the Norwegian word "fly" when they should have used the English word "plane". The English word "fly" translates to the Norwegian word "flue". This false friend mistake is probably due to the pupils believing the word had the same meaning in both Norwegian and English.

The *at*-mistake occurred 5 times. The pupils used the Norwegian word "at" when they should have used the English word "that". The pupils have here used the Norwegian demonstrative instead of the English, which probably is because they were unsure of which word to use. This mistake was not that frequent since it only occurred 5 times, and the word "that" is a frequent word in English, and most pupils aged 15-16 should be familiar with the word.

 The *again*-mistake like many other mistakes among the 15-16-year-olds only occurred twice. This mistake in question occurred when the pupils used the word "again" the same as they would the Norwegian word "igjen". For example, in Norway one can say "Jeg har 20 minutter igjen", but you cannot say "I have 20 minutes again" in English. In English one can say, "I must complete the task again", and in Norwegian one can say "jeg må fullføre oppgaven

igjen". This mistake probably occurred because the Norwegian word "igjen" has several equivalents of the word in English.

5.2.2.1 The *fly*-mistake among 15-16-year-old

The fly-mistakes were found in total 9 times in our data and were the most repetitive false friend mistake by far. The two examples below are taken from our data and it shows how the 15-16 age group uses the English word "fly" wrong. The pupils in the two examples have used the English word "fly" wrong because of the influence of Norwegian. The Norwegian word "fly" is the same as the English word "plane". This is where the confusion comes from in these examples, the pupils from the 15-16 age group most likely have the knowledge that "fly" is an English word because they have heard it during their lessons. When the pupils know that "fly" also is in the English vocabulary it causes confusion because we also have that word in Norwegian, but it means something entirely different. The English word "fly" means the same as the Norwegian word "flue", and they end up using the wrong word because of the association between Norwegian and English. This can be seen in the examples below.

Example 50: 21 years now, and my sister, mother and father died on this *fly*. I think in this thing ho happent with us wen we just was o

Example 51: ive. Castle lay in Oslo, Karl Johannsgate. We have to take a *fly* in a few times to came there. They standing on the terrace

5.2.2.2 *At*-mistakes among 15-16-year-olds

The next two examples which can be found below show two pupils from the 15-16 age group using the word "at" wrong. The two examples are using the same word, but the error is not the same in these examples. In the first example, we see a pupil writing Norwegian when he does not know the English words. In the example, we can see Norwegian words such as "dyrbare, "slik at dere" and "gikk". This pupil does not seem to try to find English words when unsure and starts writing Norwegian when they are in trouble. This happens more often in the 12-13 age group, but it happens in the 15-16 age group as well, and it could be seen as a sign of a pupil giving up. The pupil knows that the teacher reading their text knows Norwegian, which is why they write Norwegian when they are unsure, to let the teacher know what words they meant to write. The problem with this is when they try to communicate with

someone else that does not know one word of Norwegian because then they cannot do this tactic. If they start talking to someone from England for example, they will not understand if there are suddenly Norwegian words in a message. Therefore, this error and way of writing English will cause communication errors.

In the other example the pupil writes "Who are saying _at_ the young people …". This is an error that involves an English word that is used the same way as it does in Norwegian, but the word has different meanings depending on which language is being used. In Norwegian the sentence would be «Hvem sier _at_ de unge menneskene …», which is what causes the confusion. The Norwegian word for "at" is translated to "that" in English, but the English word "at" is translated to "på" and "hos" in Norwegian. The error caused by the pupil could have multiple reasons for happening. It could be a transfer error where the pupil is mixing the meaning of the Norwegian word "at" with the English word "at". On the other hand, it could be a vocabulary error where the pupil ends up writing the Norwegian word for "that", but since "that" is one of the most frequent English words it would seem more of a transfer mistake than a vocabulary mistake. Examples of this type of mistake can be seen below.

Example 52: And we must kill everybody in this mall, slik _at_ dere not say this to the police. " I was scared, oh God help i think.

Example 53: See your later. Who are saying _at_ the " young people " care of the environment. It are some children who not care of the environment

5.2.2.3 _Again_-mistakes among 15-16-year-olds

Below there are two examples where the pupils use the English word "again" incorrectly. The reason that the pupils are using "again" incorrectly is because of the influence of Norwegian. In Norwegian the word "igjen" translates into "again". The word "again" always expresses repetition, while "igjen" also expressed the meaning "left". The pupils used "again" when the word "left" would be correct, but in Norwegian the word "igjen" is used in both circumstances. In both examples, they are talking about the amount of time left until they had to do something else. This is a translation mistake because in Norwegian the word "igjen" is also used when talking about the amount of time left. This means that the pupils have translated the Norwegian word "igjen" to "again" because they have the knowledge that

"again" is "igjen", but they lack the knowledge that English has another word when talking about an amount of time left. Examples of these errors can be seen below.

Example 54: buy some nice clothes and new make-up. No, it was 15 minutes *again*. She sreemd for a taxi and jump in. When she came from she

Example 55: k, she had to bought sow new clothes. And it was only 1 hour *again* to meet up on best school in New York. She run to a big cen

5.2.2.4 Calque errors among 15-16-year-olds

The 15-16-year-old group did not a lot of the same frequent mistakes when it came to the same words, but we could observe that there was a high frequency of some types of mistakes such as Calque errors. The calque errors made up 3% of the mistakes created by the 15-16-year-olds. These last two examples from the 15-16 age group have two different words, but the error is the same. Both examples are a result of a calque error, where they have translated the examples from Norwegian directly into English.

In *Example 56* the pupil writes "half-ready", which can be translated to "halvferdig" in Norwegian. Which in this context, means that they were almost ready before their mom came out of the bedroom. "Halvferdig" is an adjective in Norwegian, and if each element is translated directly, it gives us the word "halfdone".

"Eating places" in *Example 57* is directly translated from "spiseplasser", which is often said in the Norwegian language when discussing places to eat. In English, it would be more common and correct to say, "There are a lot of great places to eat in Norway." These examples are examples where the calque error does not cause a communication error. It is comprehensible for an English speaker to understand what one means to say when saying, "eating places" and "half ready", it is just not a correct way to say it. These types of lexical mistakes are something that Johansson (1978, p. 71) probably would consider to be an annoyance for the native speaker of the language. The examples used are shown below.

Example 56: Doch Ness, ran away with alot of ballons. And before we were *half-ready* our mom came out of the bedroom. She came down to the livin

Example 57: lot of merry-go-rounds there. Its fabelous. Its lot of good *eating places* in Norway. I will show you them. I hope you will lik

6 Discussion

In this section, we will discuss the results that we found, previous research, and theory. We will discuss each of the research questions in their own section, which is found below.

6.1 How is the distribution between lexical errors caused by interlingual and intralingual influence among pupils in Norway aged 12-13 and 15-16?

The results show that the relation between interlingual errors and intralingual mistakes differs between the ages. In this study, we have found that the 12-13-year-old pupils are struggling mostly with interlingual mistakes influenced by their L1 (Norwegian). The biggest issue concerning this group is that they tend to use Norwegian words when they lack the L2 equivalent or equivalents. The older group of 15-16-year-old pupils tend to struggle more with intralingual mistakes, which also might be because they mostly stopped using Norwegian words and Norwegian elements in their writing. They rather tend to create more vocabulary mistakes. This is something when compared to other studies also seem to verify, where the older pupils/students tend to avoid using L1 word and seem less influenced by their L1 than the younger groups.

6.1.1 Less Interlingual errors among the older group (aged 15-16)

Compared to the study by Hemchua and Schmitt from 2007, who studied lexical mistakes among university students with Thai background (mostly between the ages of 19-20), we can observe that there is a difference in lexical mistakes among the different age groups. While they found zero examples of L1 words or coinage, they did find examples of calque errors among their students (of Thai origin). The correlation between our studies is that the older the learner is, and the more English experience the learner has, the fewer lexical mistakes connected to the L1 the learner has.

Even though Hemchua and Schmitt's students were older than the groups we studied, we see a correlation with L1 influence in both of our studies, they estimated that 23,75% of their student's mistakes were influenced by their L1 (see: Hemchua and Schmitt, 2007, p. 20), while in our study, we found that the pupils have a higher percentage of L1 influence among both our groups. This is because both our groups (the 12-13-year-olds and the 15-16-year-olds) are younger and therefore tend to lean more on their L1 knowledge than Hemchua and

Schmitt's' older and more experienced group. In Mahan and Brevik's article (2013, p.37) they argue that 58% of the word errors found among their target group had been influenced by Norwegian. This is a higher degree of influence than found in Hemchua and Schmitt's study, but this is also quite predictable because of the age and experience difference between the groups.

The amount of Norwegian influence among the interlingual errors can be explained by misformations. As explained in section 2.2, misformations often result from misapplying word formation rules and produce non-existent words which often show clear influence from the learner's mother tongue (Thornbury, 2020, p. 29). We looked at misformations that showed clear influence from Norwegian and found a lot of words in our data. One of the words was "helpsome", which contains Norwegian influence from the Norwegian word "hjelpsom". The correct word in English is "helpful", but the pupil apparently did not have the word "helpful" in their vocabulary or did not remember it at the time, and instead produced a non-existent word. Another example from our data was the word "clatring", which also is mentioned in 3.3 when it is used as an example of a Norwegian element. It has Norwegian influence from the Norwegian word "klatring" which means "climbing" in English. Some pupils probably did not know what the English word for "climbing" was, or they had forgotten it in the moment. When they were struggling to find the correct word, they decided to produce a non-existent word that look like an English version of a Norwegian word.

As mentioned in section 2.2, according to Agustín-Llach (2017, p. 65) learners tend to borrow L1 words, create lexical adaptations, and calque L1 words or expressions when they lack exact lexical knowledge of the L2 equivalents. This can be used to explain the differences in results between in two groups in our study. The 15-16-year-old pupils have had two years more English in their respective schools, and therefore have more experience and have been more exposed to English than the 12-13-year-old pupils. These two years have probably helped them gain a larger vocabulary, better lexical and linguistic awareness. Therefore, they would have a smaller need to tend to borrow L1 words and create lexical adaptation based on L1 knowledge or calque words. For this reason, we see a smaller number of L1 words, lexical adaptations, and calque cases among the older group of 15-16-year-olds. The same would go for the Thai students that Hemchua and Schmitt studied, they were mostly three to four years

older than the pupils we studied, and they were at a university level, so they would have a better lexical understanding than our group of pupils.

Muñoz (2006, p. 2) argues that older children acquire L2 faster than younger children (again, in the early stages of syntactic and morphological development where time and exposure are held constant). We believe such a development can be seen through the differences in the two groups we have studied, the 12-13-year-olds, and the 15-16-year-olds. We believe the pupils that are aged 15-16 are learning much faster because of more experience, increased exposure to English, and because of maturity. For this reason, their language is using less Norwegian words and Norwegian elements because they now acquire English faster than when they are younger, in the same way as Muñoz mentioned. The older learners were observed to be especially good at syntactic and morphological rule acquisition, and at metalinguistic ability and vocabulary (Muñoz, 2006, p. 4). The older learners seemed better fit to understand different aspects of the L2, and therefore create fewer mistakes, in addition to having a better vocabulary, this also could be part of the influence of why the pupils are quitting using Norwegian elements in their L2 language. As the pupils get a better understanding of rule acquisition the use of for example Norwegian suffixes at the end of words will diminish. We can see the diminishing of Norwegian elements between the two age groups in our study. In our study, we can observe that the number of Norwegian elements is lower among the pupils aged 15-16 than the group of pupils aged 12-13.

As found in our study, we see that the group aged 12-13 has 57% of their errors because of L1 influence. Mahan (2013, p. 91) has gotten some different results in her older group of 14-15-year-olds. She states that 62% of the errors in her group were because of L1 influence. This goes against what we have found, but we believe this can be explained by a difference in methods. Mostly when we have been uncertain of where to put a "match" we have put this in the indeterminate category. In our study, we see that we have a much higher percentage of indeterminate errors than Mahan has found. Next up our group of 15-16-year-olds, we see a drastic change in L1 influence. We can observe that the number of L1 influence goes down to 27%. This is more similar to Hemchua and Schmitt's study, where their group had 23% L1 influence in their language. This seems to show that there is a change in L1 influence (interlingual) mistakes when as pupils grow up.

6.1.2 A higher number of intralingual mistakes among the older group

In section 5.1.3 we could observe that there is quite a big difference between the two age groups when it came to vocabulary confusion. We can observe percent vice that the group of 15-16-year-olds struggled the most with VC-mistakes, while the younger group struggled more with interlinguistic errors. Percentage vice we can observe that there is a big difference, but in reality, the number is quite even between them as shown earlier. Why is it the case that the frequency of vocabulary confusion mistakes is much higher in the older age group than in the younger one?

To answer this, we must look further into our data. In our data set, we can observe that the difference comes mostly from interlinguistic mistakes. We see that the older group (aged 15-16), creates fewer interlinguistic mistakes such as calque errors, Norwegian word errors, false friend errors, and Norwegian element errors. While seeing this, we can also observe that the number of vocabulary mistakes remains quite even. Therefore, we can see that the frequency of vocabulary confusion is much higher among 15-16-year-old than the 12-13-year-olds. Now, the next question we need to ask, is why we see such a change between interlingual and intralingual mistakes.

The transition between first and second language can be confusing for pupils. The pupils can find it difficult to learn another language because they mix different words with each other. Making sentences can also be difficult for pupils, because of subject-verb agreement and word order. Every grammar rule that you learn in Norwegian will get replaced with other English grammar rules. This may be confusing because the pupils might think that the different rules are for both languages, and they might use a Norwegian rule in English for example. On the other hand, it can also be helpful for pupils to already have experienced learning a language from before. When a pupil is learning a second language, they can use associations that link between the languages. There are lots of words that are written the same in both Norwegian and English, "egg", "hammer", and "storm" to name a few. Some words are very close to each other as well, the English word for "flag" is "flagg" in Norwegian for example. This information can help the pupils that have Norwegian as their first language to learn English. A lot of time is saved when learning a second language because they do not have to relearn different concepts, someone will not confuse a dog with a cat for example (Thornbury, 2020, p. 18). First language learners must learn concepts and categorize them. A

second language learner already has knowledge of these concepts and just must put the different words and labels in the category. For example, both a cat and a dog go into the animal category, but second language learners know that they are not the same.

Therefore, it seems like the pupils aged 12-13 are making more Norwegian word errors because they do not have the required vocabulary yet. When pupils do not know the word that they want to write, they decide to just write it in Norwegian. Instead of trying to find another word for it, or attempting to make it English, they just write it in plain Norwegian. Meanwhile, the 15-16 age group does not have a lot of Norwegian word errors because they are trying to use their English vocabulary and at least they try writing in English. Most of the pupils aged 15-16 try to find an English word, even it if is not correct. Therefore, the gap between the age groups is so big when it comes to the Norwegian word label. We also suggest this is one of the reasons why the vocabulary confusion error is so even between the two groups. When the pupils are quitting using Norwegian words and struggle with finding the correct vocabulary, they tend to use words that are either similar in writing (mis-selection) or create meaning-related errors.

In Hemchua and Schmitt's study is that their largest group of findings were the "Near synonym" mistakes among their students (aged 19-20), which can be compared to our group of vocabulary confusion. Below is one example of a near synonym by Hemchua and Schmitt.

 "You will <u>get</u> up in the morning because of the sound of birds. The intended meaning of the underlined words in the context was 'to become awake after sleeping' not 'to leave the bed'. Therefore, 'wake up' was required." (Hemchua and Schmitt, 2006, p. 18)

This mistake looks quite like our label "vocabulary confusion", where the learner has used the wrong lexical word for a situation. Our group of 15-16-year-olds seems to struggle with the same as the student in Hemchua and Schmitt's study. The older groups, seem to struggle mostly with choosing the correct words for the correct situation. Mahan and Brevik (2013, p. 36-37) when referring to Mahan's study, say that 39% of the word errors created by the pupils aged 15-16 in Norway were because they tended to simplify their language by using verbal English instead of formal English.

After comparing our data with Mahan's (2013) data, we noticed that we have both similarities and differences. Mahan used effects when writing about what was wrong with a sentence or what was wrong with the words that were used, while we used labels to determine what was incorrect. Mahan decided that when a learner used informal words, they got marked as an error. In our study, we put the informal words in a label for itself because we did not see the informal use of words as a lexical mistake. While Mahan (2013) noted down the informal usage of words as word errors, we did not.

As mentioned in 2.3.2, Stylistic/connotational is way above the rest of the effects, and one of the reasons for that is the verb "to get". Hasselgreen & Sundet (2017) had a list of the top ten wrongly used words in the CORYL corpus. The word that is used wrongly the most according to Hasselgreen & Sundet (2017, p. 203) was "got", where 2,3% of the total number of wrong words were "got".

When we sorted through our data, we decided to place all the "get" and "got" into the non-lexical mistake label. We decided to do this because using "get" and "got" is perfectly fine and legitimate, and we did not want to expect 12-13- and 15–16-year-olds to write perfectly formal English as their L2. Below are a few examples of the "got" error, where the CORYL corpus marked it as wrong, but we decided to not include it in our data as a lexical error.

Example 58: my day than this. But I went with the horrible teacher and I <u>got</u> the test, it was about the homework we had to to day and I

Example 59: nd faled down the rest of it, that did quite hurt. And wen I <u>got</u> into the kichen I maneged to break one of my favorite glase

Example 60: ckboard and solve one of his unsolveble math questions. As I <u>got</u> up I tripped and fell on my face. The class started to laug

Example 61: 992 whit a solo-rap album and the hit " Real Slim Shady " He <u>got</u> world famous after this, and everyone wanted a piece of him

These sentences are fine to write at an intermediate level of English. All these examples were placed in the non-lexical mistake label by us, while Mahan and Hasselgren would classify these as word errors. In our study, we put the informal words in a label for themselves

because we did not see the informal use of words as a lexical mistake. Our reason for this was that we did not want to expect our age groups to write perfect formal English as their second language. This is also since younger pupils who learn an L2, might not have learned the difference between good formal English and verbal English.

Another element of vocabulary confusion we found in addition to the simplification were instances of mis-selection. As explained in section 2.2, a mis-selection is when a learner uses an existing word that is similar in sound or spelling to the correct form (Thornbury, 2020, p. 29). When looking for mis-selection it is important to read the context of the whole sentence. Even if a pupil manages to write a word correctly, does not mean that they know how to use the word or what it means. Two examples of mis-selection in our data are the use of "sad" and "sat" instead of "said". Both words are very similar to "said", but they have two different meanings, which can confuse the pupils. When looking at the contexts of the sentences we can see that they meant to write "said" but wrote the wrong word instead. One of the sentences was about two friends talking, and the pupil wanted to write "and he said" but instead wrote "and he sad". The other sentence was about someone who called 911 and said that he called the wrong number, but the pupil wrote that he "sat to the 911 that I call wrong nummber". Mis-selection occurred often in our data, and it happens when the pupil gets confused and mixes up words in their L2 vocabulary.

As noted above, the younger pupils tend to make more interlingual errors, and this is what they struggle with in our dataset as well. If we look at the 15-16-year-olds we can see that they make 52% intralingual errors, which can be interpreted as alarming. However, it can also be seen as a positive development as the pupils age. When the 12-13-year-olds struggle with interlingual errors, it would be great to try to remove as much interlingual influence as possible. This can be seen just by looking at the 15-16-year-olds, they have developed and do not make that many interlingual errors anymore. This might be because the teacher and pupils focus more on interlingual errors because using a different language (L1) is worse than making vocabulary mistakes within the target language. This means that the group aged 15-16 has managed to learn and remove most of their interlingual influence, thus making more intralingual errors by comparison to the younger age group aged 12-13.

71

As explained in section 2.2, "Meaning-related errors typically occur when words that have similar or related meanings are confused, and the wrong choice is made." (Thornbury, 2020, p. 29). We could easily track meaning-related errors by looking at vocabulary confusion in both age groups. One of the most frequent meaning-related errors that we got in our data was the use of "see" instead of "look" and "watch". In English, the words "see", "look" and "watch" is something that you do with your eyes, but they are not synonyms. The word "watch" is used when you pay attention and look at something for an amount of time, for example when you want to watch a football game. The word "look" is when you direct your eyes in a direction, for example, you look out the window, or you look at a house.

Lastly, the word "see" is used when you become aware or notice something. For example, you watch a football game and see a popular football player, or you look out the window and see a man walking his dog. The pupils in our data often used "see" when either "look" or "watch" would be correct. This is probably because in Norwegian they only have one word which is "se", which can combine with "på" to express the meanings "look at" or "watch" and it is easier for pupils to remember the word "see" because it is very close to the Norwegian word. Pupils then end up writing sentences such as "They starting to see around in the house" when it would be correct to use the word "look" instead. An example of a sentence where "watch" would fit is "I can see a man see tv", here the pupil first used see correctly when they wrote "I can see", but since the pupil only uses the word "see" they end up using it instead of "watch" at the end of the sentence as well. As mentioned in section 2.2, meaning-related wrong-choice could also derive from the learners L1.

In our data, we noticed that the pupils often used the English word "good" a lot, even when "well" would be correct. The pupils use these words when they want to translate the Norwegian word "bra". This causes confusion because "good" is an adjective, while "well" is an adverb. This means that context matters when deciding which word to use when the pupil wants to translate the Norwegian word "bra" into a sentence. In English, you would write "this car is good, and it drives well", but in Norwegian you can write "denne bilen er bra, den kjører bra" and it is correct. The reason for the confusion could therefore be that Norwegian only uses one word, but in English, you have to use two words. This is something that the pupil must learn and separate the two languages to not cause confusion.

6.1.3 Difficult cases to determinate

Some cases were hard to determine, as they could be classified as both an interlingual error and an intralingual error. In Mahan's study she had stylistic/connotational errors as an effect, in our study we put the stylistic/connotational errors inside the L2 vocabulary confusion label. These types of errors happened in both the 12-13 age group and the 15-16 age group. One error that Mahan also found a lot of was the use of "good". This error also appeared frequently in both our 12-13 and 15-16 age groups. Below is one example from each of our age groups.

Example 62: taken. We found uss a new cleaninghelp, and everything went *good*. But one day I found the door wide open again, but it was j

Example 63: airs and broke my leg. That night i wasn't actually sleeping *good*, because i woke up and saw my bed was broken. so that day w

Example 62 is from the 12-13 age group, while *Example 63* is from the 15-16 age group. In both examples, the word "good" is used instead of "well". Mahan & Brevik (2013) mentioned that the overuse of common and generic adjectives should be avoided while writing English, which especially applies to the English word "good". As we can see, it also shows that the pupils in our data have problems with writing "good" instead of "well". On the other hand, the English word "good" is close to the Norwegian word "godt", which may confuse some pupils and mix the words when writing. In Norwegian you would say "sov godt (sleep well)", and this confusion might make the pupils instead write "sleep good". This can also be because of a lack of knowledge of the lexical word well, and therefore the pupils have used the simpler word "good". The word "well" is not that frequently found in the corpus, and this could indicate the pupils do not have that much knowledge about the word, which we believe is the case here.

6.2 To what extent do pupils aged 12-13 and 15-16 make lexical errors which are influenced by Norwegian?

What we see is that the younger pupils (aged 12-13) struggle mostly with their vocabulary, in the sense that they tend to use Norwegian words when they do not know the English equivalent of the word. This L1 influence almost disappears when it comes to pupils aged 15-16. When it comes to Norwegian influence in that age group, they tend to have more Norwegian elements in their language than whole L1 words. This can be both coinage mistakes (creating new words by mixing L1 and L2) or for example writing an English word with a Norwegian suffix (or the other way around), or other types of L1 influences. We can observe that different age groups tend to have different types of L1 influence, but overall, the amount of L1 influence decreases with age.

The largest group of L1 influences found was among the 12-13-year-old group, they tend to use whole Norwegian words to a large degree and created more than twice as many Norwegian elements as the older group of pupils. One error type we found to be surprisingly low was the false friend error. We expected to find much more of this mistake since we believed similar words in both Norwegian and English would be confusing for the pupils. The L1 influence among the 12-13-year-olds seems in huge degree to indicate the lack of lexical understanding of the words. Agustín-Llach who looked into a study of 12-15-year-olds found also out that these pupils struggled mostly with L1 influence, and their biggest mistake was also L1 word usage. These mistakes are clear evidence that the pupils lack vocabulary to substitute the L1 words and that they also might struggle with conceptualization of these.

The older group has a lower extent of lexical errors which are influenced by transfer from Norwegian. They do tend to not use Norwegian words and have much fewer elements of Norwegian in their language. We believe there might be many reasons for this, for example, increased cognitive development, increased vocabulary, and L2 experience can play a role in this. As Munoz (2006) mentioned older kids also acquire faster than younger children. We believe that we can see this when comparing these two groups, the older group has probably acquired a better skill set when it comes to English rules and a better vocabulary, so their need to lean on Norwegian is therefore decreased. For example, under the Norwegian element label, we can observe cases where pupils use Norwegian suffix endings in English words.

This tells us that the pupil that has created this error probably has not acquired the skills of choosing the correct tense.

The younger groups' influence on their L1 is clearly influenced by their lack of English vocabulary. As Thornbury mentioned, an educated native speaker has a vocabulary of around 20 000 words, while a second language learner is lucky if they have acquired 5000 words over several years (2020, p. 20). Both our groups are in the process of acquiring new vocabulary, and it seems that the extent of L1 influence decreases with age, lexical knowledge, and increased vocabulary.

6.3 Which errors are most frequent among the lexical errors caused by transfer from Norwegian?

In this study, we have found some mistakes that were more frequent than others. In this section, we will discuss if these findings actually are frequent. To do this we must first define what frequent is. Frequency can be defined as something that happens between short intervals.

In this study, we have found some mistakes that were frequent among the 12-13-year-olds. The most frequent mistake was the interlingual *The*-mistake, where the pupils used the word "the" instead of the word "it". This mistake is probably due to the phonetical similarities between the word "the" and the Norwegian word "det". As mentioned in the findings section, this error is then made at least once by one out of five pupils. This tells us that this mistake is at least to some degree frequent among the 12-13-year-olds.

The *the*-mistake mistake we could only find twice among the older group aged 15-16, and this tells us the older group has gotten a better understanding of the word and its meaning. What this tells us is that this type of mistake has been diminished between the ages of 12-13 to 15-16. This might be because the older group is better at conceptualizing the word. The 12-13-year-olds probably believe that the L1 word "det" and the English word "the" are the same in concept. The older group sees probably sees that these two are different concepts and that the L1 word in Norwegian needs a Norwegian concept, while the L2 word needs an English concept (as the determiner in this case). This would then work as the different conceptualizing modules Weinreich mentioned earlier, the compound model and the co-ordinate model.

The other mistakes such as the "Hytte" and "opp" mistakes, we do not dare to call frequent, as they are only made by some of the pupils. They also occur way to few times to say that this is a general issue among the pupils. The pupil probably also felt pressured to use the word "hytte" where they were supposed to write the word "treehouse" because one of the tasks depicted a treehouse. Since the word "treehouse" is not that frequent and a core word in English, some pupils probably did not know it and wrote the Norwegian word "hytte" instead.

The older group, the 15-16-year-old, did not have any mistakes because of transfer from Norwegian which we can say is frequent. They had some mistakes that were more frequent, but not generally frequent in their group. Their most frequent mistake was the "fly" mistake, where they mixed the Norwegian word "fly" with the English word "plane". This mistake only occurred a few times among such a big group of pupils it cannot be called frequent. The same goes for the other errors related to Norwegian transfer.

6.4 Implication for teaching English in Norwegian schools

Our study shows that a lot of pupils tend to lean on their L1 (Norwegian), especially when they lack an L2 equivalent to the Norwegian word they are thinking of. We also see that pupils in Norway often choose the wrong word for the situation they are thinking of, which we assume is because they do not have the correct word in their vocabulary. This then could create for example the calque errors such as "cleaninghelp" or English words that contain Norwegian elements within them, or Norwegian words that contain English elements in them.

To avoid the types of mistakes like those mentioned above it would be important that schools continue to focus on learning new vocabulary, in the correct context. For example, through the different conceptualizations that Weinreich mentioned. If the new vocabulary taught is just forgotten or not learned well enough the pupils would continue creating VC-mistakes, it would be like Mahan and Brevik (2013) said, that the lack of knowledge about a word created word errors and the learner would go back to their L1 knowledge. The Norwegian curriculum in English states that "learning the pronunciation of phonemes, and learning vocabulary, word structure, syntax, and text composition gives the pupils choices and possibilities in their communication and interaction" (The Ministry of Education and Research, 2020, p. 2). The curriculum already mentions the importance of vocabulary, and therefore with a smaller vocabulary one's "choices" of words to use would diminish when it comes to creating clear

and understandable communication. As mentioned earlier in this thesis, lexical mistakes such as using the wrong word can be an annoyance for other English speakers, as it complicates communication. For this reason, it is important to continue and increase the pupil's vocabulary set.

Schools should probably also try to focus on the most frequent mistakes that pupils make. Removing the most frequent mistakes is an essential step toward being fluent and efficient in a secondary language. Some of the reasons to remove the most frequent mistakes are found below.

1: Effective communication, the fewer mistakes, the clearer the communication gets. It would also be less of an annoyance for the native speaker or the other English L2 speaker.

2: Confidence, removing frequent mistakes can be a good confidence boost as one sees the texts he or she is creating are improving.

3: Credibility, removing the most frequent mistakes will help create a clearer text and show that one can master the L2. Having some types of mistakes that are frequent in a text can undermine the credibility of the text.

In our study, we can say the group of 12-13-year-olds had some mistakes that were quite frequent. For example, we could see that some pupils struggled with understanding the word "the", where they believed it had the same meaning as the Norwegian word "det". If frequent mistakes like these were removed the texts and sentences of the pupils would seem clearer be easier to understand. Frequent mistakes like these among 12-13-year-old would be wise to fix. Here it has to be mentioned that this was a frequent mistake from the 2004-2005 texts among pupils, and if it is a frequent mistake today, we cannot tell.

Also, a lot of mistakes might be avoided if the pupils have a good understanding of the most frequently words used. Thornbury (2020, p. 21) said that the most important words to learn are the words that are often in use. Since the word "the" is a word some pupils struggle with, this would be an example of a word that the pupils could need a deeper understanding of. This is also one of the most frequently used words used in English, and therefore the pupils should be able to master using the word correctly.

7 Conclusion

In this study, we have tried to provide an answer to the following thesis question: Which types of vocabulary errors do pupils in Norway aged 12-13 and 15-16 make?

Based on our findings, theory, and comparison with other studies we see that the 12-13-year-olds and the 15-16-year-olds tend to struggle with different vocabulary errors. The 12-13-year-olds seem to lean more on their L1 than the older group aged 15-16. The 12-13-year-olds use a substantial number of Norwegian words, as well as Norwegian elements in their English. The older group aged 15-16 does not have the same issue, as they rather struggle with vocabulary confusion. Both groups have the same amount of vocabulary errors, but the older group seems to put away the usage of Norwegian words and have fewer Norwegian elements in their English.

What we suggest is the difference between the two groups is that the 12-13-year-olds tend to use more Norwegian in their L2 language because of:

1) Lack of English equivalents to the Norwegian word they are thinking about/lack of vocabulary, which leads to them using their L1.
2) They have less lexical knowledge about words and English grammatical rules, which leads to lexical adaptations (Mixing L1 and L2).
3) They have fewer English lessons/education and thus less English exposure (Input and output).

We can see in our study and by comparing our study to others, that to be less dependent on one's L1 in L2 one has to try to increase one's vocabulary. Lack of vocabulary creates a lot of different lexical errors in the L2, such as different types of L1 influence (borrowing L1 words, using L1 elements in one's English, false friends, and Calque errors) and it creates more vocabulary confusion. We can also see that with age and experience the amount of L1 errors diminishes. In our two groups, we could not see a difference in the number of intralingual mistakes, but we suggest this because the pupils are quitting using whole Norwegian words and replacing them with the wrong words instead. We believe this is because they still lack the correct equivalent of the word they intend to use. Thus, they create mis-selection and

meaning related errors instead of using Norwegian words or words that are influenced by Norwegian.

Suggestions for further research

Few studies focus on Norwegian influence in L2 learning.

When creating this study, we found that there were few other studies that focused on Norwegian L1 influence when learning English as an L2. There were a couple of other studies that were in the same field, but there is still a huge gap in knowledge when it comes to Norwegian L1 influence when learning English. Our study had to some degree look into how the influence of Thai (L1) was influencing learning English as an L2, because of the lack of more relevant research.

Comparing to other age groups.

Data we would find interesting to read would be how the data would look if compared to other age groups, both younger learners and older. There is not much information about the age group younger than 12 and older than 16-18. A study of older or younger Norwegian English learners would contribute a lot to the field. Our study focuses on the 12-13-year-old and the 15-16-year-old, Mahan and Brevik focused on 14-15 and 19-21-year-olds, so more information about different age groups land English learning is something that needs further research.

A modern set of data.

Most studies done in the same field as our study, that investigate issues connected to L1 when learning English as an L2 are not the most up-to-date. The same goes for our study, which is created in 2023, but the data set is from 2004-2005. Between 2004-2005 and today a lot of things could have happened with the pupil's English skills. We believe that pupils today are possibly better in English because of more input, but there are not many studies that prove this today. An issue with this is that it is difficult to find a modern set of data to create a study for this, especially since the national test in writing is not tested today. Research like our study with more modern data would tell us a lot more about the English situation today and could be used to say if the English of pupils has improved over the years.

Works cited

Agustín-Llach, P, M. (2017). Vocabulary Teaching: Insights from Lexical Errors. *TESOL International Journal, 12(1), 63-74.*

Atar, C. & Erdem, C. (2019). THE ADVANTAGES AND DISADVANTAGES OF CORPUS LINGUISTICS AND CONVERSATION ANALYSIS IN SECOND LANGUAGE STUDIES. *ResearchGate.* *https://www.researchgate.net/publication/337858444_THE_ADVANTAGES_AND_DISADVA NTAGES_OF_CORPUS_LINGUISTICS_AND_CONVERSATION_ANALYSIS_IN_SECOND_ LANGUAGE_STUDIES*

Clarino Centre Bergen. (without year). *CORYL (Corpus of Young Learner Language).* Retrieved from 06.03.2023 https://clarino.uib.no/korpuskel/corpora

Cohen, L., Manion, L. & Morrison, K. (2018). *Research Methods in Education (8TH edition).* Routledge.

Garshol (2019). *I JUST DOESN'T KNOW;Agreement errors in Englishtexts by Norwegian.* [Doctoral dissertations]. University of Agder.

Gleiss, M, S. & Sæther, E. (2022). *Forskningsmetode for lærerstudenter.* Oslo: CAPPELEN DAMM AKADEMISK.

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

Heaton, J. (2008). *Secondary analysis of qualitative data; an overview. Historical social research*, 31 (3), p. 33-35.

Janda, L. A. (Ed & writer.). (2013*). Cognitive linguistics - the quantitative turn: The essential reader*. De Gruyter, Inc.

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition.* Routledge.

Johansson, Stig. (1978). *Studies of Error Gravity: Native reactions to errors produced by Swedish learners of English*. Sweden: Acta Universitatis Gothoburgensi.

Mahan, K, R., & Brevik, L, M. (2013). «I can English very good» – engelske ordfeil blant norske elever og studenter. *Bedre Skole* (nr. 3), 34-39. https://www.utdanningsnytt.no/files/2019/08/22/Bedre%20Skole%20%203%202013.pdf

Mahan, K, R. (2013). *Lexical Errors in Norwegian Intermediate and Advanced Learners of English.* https://www.duo.uio.no/bitstream/handle/10852/37040/MahanxMaster.pdf?sequence=1&isAllowed=y

Miceli, S. (2019). *Reproducibility & Replicability in RESEARCH. THE NATIONAL ACADEMIES INFOCUS,* 18(2), 12-14. http://infocusmagazine.org/portable/18.1.pdf

Munden, J. & Sandhaug, C. (2017). *Engelsk for Secondary School.* Oslo: Gyldendal Norsk Forlag.

Muñoz, C. (Ed.). (2006). *Age and the rate of foreign language learning. Multilingual Matters*. The Cromwell Press Ltd.

Schmitt, N., & Hemchua. S. (2006). An analysis of lexical errors in the English compositions of Thai learners*. Prospect* (Adelaide, S. Aust.), 21(3), 3–25.

The Ministry of Education and Research. (2020). *Curriculum in English* (ENG01-04) [Curriculum for English]. Established as regulations. The National curriculum for the Knowledge Promotion 2020. https://data.udir.no/kl06/v201906/laereplaner-lk20/ENG01-04.pdf?lang=eng

Thornbury, S. (2002). *How to teach vocabulary.* Harlow: Longman

University of Bergen. (2023, 2. February). *Corpus research.* University of Bergen. https://www.uib.no/en/rg/test/90615/corpus-research

Weinreich, U. (1974*). Languages in Contact: Findings and Problems*. Mouton

Appendix

Nasjonale prøver i engelsk

February 1st 2005

7th grade

Writing test

Time: 60 minutes

Answer all three tasks. Spend half the time on task 3.

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

_____

_____

_____

_____

_____

_____

_____

_____
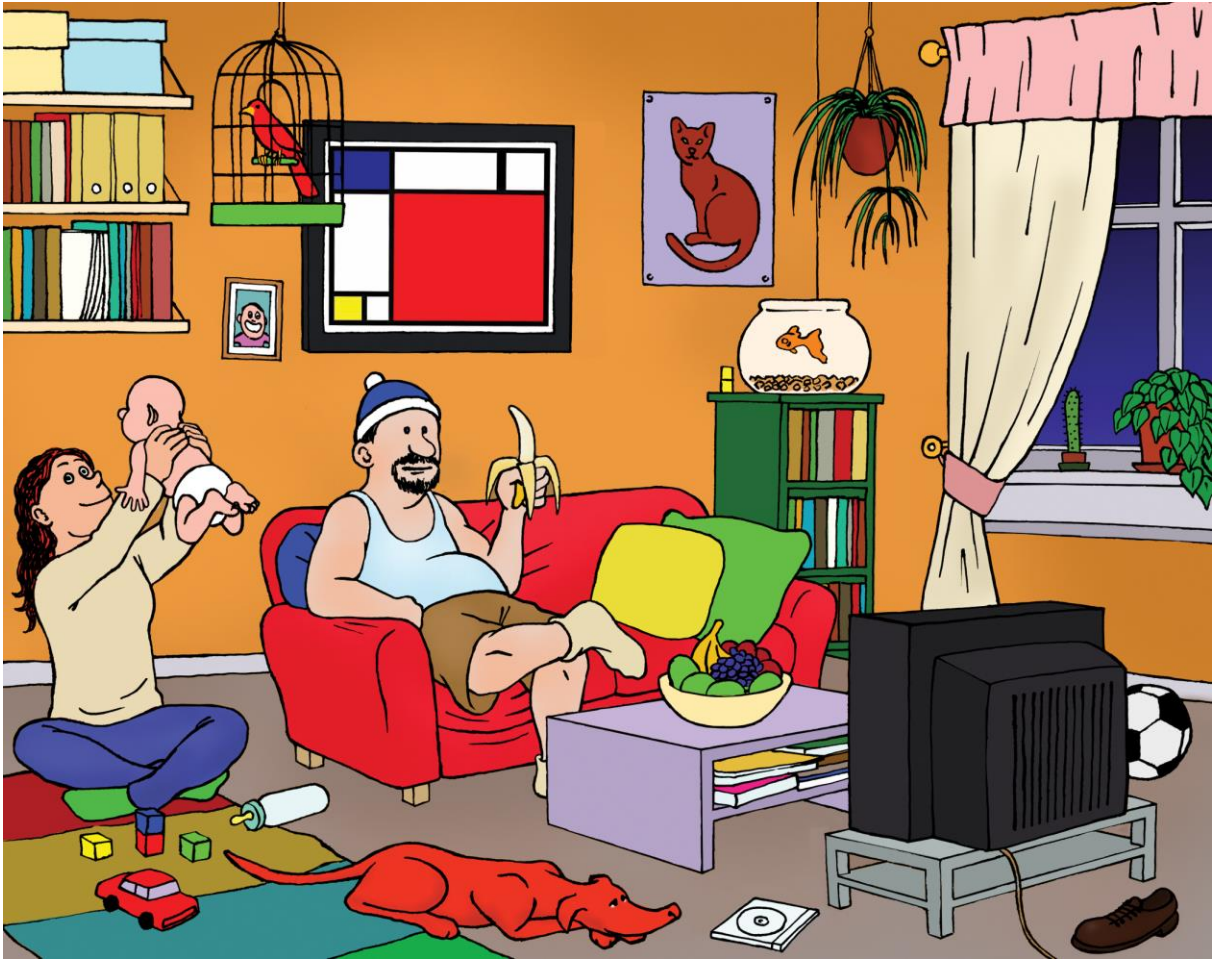
2. You are on holiday in one of these places. Write a postcard to a friend and tell him or her, for example:
WHERE you are. WHAT you are doing. WHAT you like and/or dislike.

3. DO TASK a) OR b):

a) Write a text, 3 or 4 paragraphs. Start with the sentence:

One day when I came home from school, I found the front door wide open.

Give your text a title.

OR

b) Write your own text, 3 or 4 paragraphs, using this picture.

Write your text on the next page.

Nasjonale prøver i engelsk

March 31st 2004

10th grade
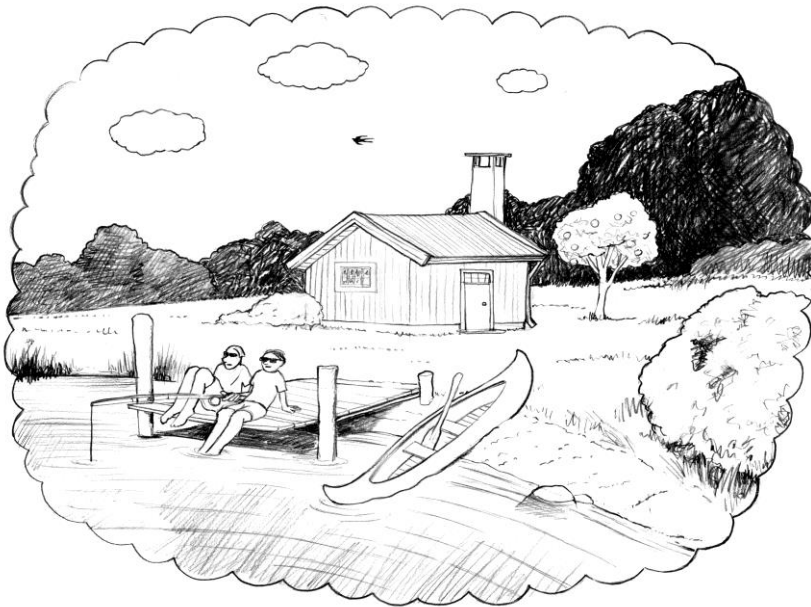
Writing test

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.

You are going away for a weekend. Write an e-mail inviting some friends to join you.

Tell them for example:

 WHERE…. WHEN…. and WHAT TO BRING.

Mixed classes: bad for the learning environment?

Your school is considering introducing separate classes for boys and girls. You have

strong feelings on the subject. Write 1 or 2 paragraphs for the school newspaper,

arguing for your view, but showing that you can understand both sides of the issue. Give

your text a title.

EITHER

a) Write a text of 5 – 8 paragraphs with the title "The next ten years of my life".

OR

b) Write a text of 5 – 8 paragraphs, finishing with the sentence:

…. That's what friends are for.

Give your story a title.

Appendix 3 – Norwegian National Test in English year 2005, 10th grade.

Nasjonale prøver i engelsk

January 25th 2005

10th grade

Writing test

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"