INLAND NORWAY
UNIVERSITY
OF APPLIED SCIENCES

LUNA

Henrik Skifjell Eliassen

Master's thesis

The use of *I* as subject in mental and verbal processes in texts by Norwegian advanced learners of English

Lecturer Education in English

Spring 2018

Consent to lending by University College Library   YES ☒ NO ☐
Consent to accessibility in digital archive Brage   YES ☒ NO ☐

Acknowledgements

I would like to thank my two excellent supervisors, Sylvi Rørvik and Hege Larsson Aas. Their constructive feedback from the beginning till the end of this process has been extremely helpful. I would not have been able to complete this thesis in time if it wasn't for them. From the very beginning, they encouraged me to ask for help whenever I needed it.

I would also like to thank my girlfriend Aina for her personal support during the last year.

Abstract

Title: The use of *I* as subject in mental and verbal processes in texts by Norwegian advanced learners of English

The main aim of this thesis is to investigate the hypothesis that Norwegian advanced learners of English overuse *I* as subject in mental and verbal processes, and in the recurrent word-combinations *I think*, *I feel*, *I know*, and *I would say*, due to transfer from Norwegian and developmental factors. The investigated material in this study comes from the Norwegian component of the International Corpus of Learner English (NICLE). In addition to that, four other subcorpora are included in this study, and these are put together by texts written by Norwegian and English writers in their first language (L1). Two of these subcorpora consist of texts written by Norwegian and English experts, and the comparison of these two subcorpora represents the standard the NICLE writers want to reach. The other two subcorpora consist of texts written by Norwegian and English novices, as they are at the same skill level as the NICLE writers, and at the same developmental stage.

The method used to investigate the hypothesis is through use of the Integrated Contrastive Model. The first step of this model is to carry out contrastive analyses between the Norwegian and English L1 subcorpora. The next step is to compare the NICLE texts to the L1 texts from the English experts. After that, explanations are sought for the differences in usage. In this study, influence from transfer and developmental factors are investigated.

The findings from the present study correlate partly with the hypothesis as overuse of *I* as subject in mental and verbal processes are taking place among NICLE writers, as well as in the four recurrent word combinations. Transfer from Norwegian does not seem to impact the NICLE writers overuse, with the exception of *I* in verbal processes. Developmental factors do not seem to influence the NICLE writers' usage of *I* in clauses with mental processes, but influence can be seen in clauses with verbal processes. Developmental factors seem to influence the NICLE writers in the usage of the recurrent word-combinations, except for in the combination *I would say*.

Norsk sammendrag

Tittel: Bruken av *I* som subjekt i mentale og verbale prosesser i tekster skrevet av norske engelskstudenter

Hovedmålet med denne oppgaven er å undersøke hypotesen om at norske engelskstudenter har overforbruk av ordet *I* som subjekt i mentale og verbale prosesser, og de tilbakevendende ord-kombinasjonene *I think*, *I feel*, *I know*, og *I would say*, grunnet overføring fra norsk, og utviklingsfaktorer. Det undersøkte materialet i denne studien er hentet fra korpuset NICLE, som er en sammensetning av tekster fra norske engelskstudenter. I tillegg er det i denne studien lagt ved fire andre subkorpora som er satt sammen av tekster skrevet av norske og engelske skrivere på deres L1. To av disse subkorpora inneholder tekster skrevet av norske og engelske eksperter, og sammenlikningen av disse representerer målstandarden NICLE skriverne ønsker å oppnå. De andre to subkorpora inneholder tekster skrevet av norske og engelske skrivere som er på samme ferdighetsnivå som NICLE skriverne, og som dermed representer det samme utviklingsstadiet.

Metoden for å undersøke hypotesen er gjennom bruk av Integrert Kontrastiv Modell. Det første steget i denne modellen er å utføre kontrastive analyser mellom de norske og engelske L1 subkorpora. Etter den kontrastive analysedelen blir NICLE tekstene satt opp mot tekster fra engelskspråklige eksperter. Deretter blir det forklart om forskjeller i bruk kan forklares grunnet overføring fra norsk og utviklingsfaktorer.

Funnene I denne oppgaven stemmer delvis med hypotesen, ettersom overforbruk av *I* i mentale og verbale prosesser, og overbruk av de fire tilbakevendende ord-kombinasjonene finner sted blant NICLE skriverne. Overføring fra norsk virker til å ha lite innvirkning på NICLE skriverne sitt overbruk, med unntak av *I* i verbale prosesser. Utviklingsfaktorer virker ikke til å ha innflytelse på NICLE skriverne sitt bruk av *I* i mentale prosesser, men innflytelse kan sees i bruken av *I* i verbale prosesser. I bruken av de tilbakevendende ord-kombinasjonene virker NICLE skriverne å være påvirket av innflytelse fra utviklingsfaktorer, med unntak av ord-kombinasjonen *I would say*.

Table of contents

Tables:

Figures:

1. Introduction

Mastering the English language is an increasingly important concern for Norwegian students at university level. Hellekjær (2005, p. 14) shows that the proportion of English titles on college and university reading lists were considerable, ranging from about 65 per cent in the Natural sciences to 50 per cent in Humanities and Social Sciences. Norwegian advanced learners of English would of course like to resemble native speakers in their written output, as they represent the target standard to follow. The objective of the present study is to contribute to the expansion of our knowledge about Norwegian advanced learners' writing, focusing on the use of *I* as subject in their argumentative English L2 texts. Studying *I* as subject while seeking to explain what might cause the differences in usage between the learners and native speakers will hopefully provide some answers that can contribute to one of the central aspects in English teaching, which is to develop better L2 writers.

Corpus-based studies have suggested that Norwegian students of English at university level, further referred to as *advanced learners of English* in this study, tend to express personal feelings and attitudes in their texts to communicate with the reader. Petch-Tyson (1998, p. 108) defines this interpersonal function of language within the text as *writer/reader (W/R) visibility*. The use of *I* as pronoun in texts is one element of W/R visibility researchers have investigated and the results show general overuse of *I* from advanced learners compared to native speakers (McCrostie, 2008, p. 107; Petch-Tyson, 1998). Studies that specifically target Norwegian advanced learners of English have been conducted, and the findings show overuse of *I* as subject in texts compared to native speakers (Fossan, 2011; Hasselgård, 2009, 2012; Paquot, Hasselgård, & Ebeling, 2013). Hasselgård's (2009) and Paquot et al's. (2013) findings indicate that Norwegian advanced learners predominantly overuse *I* as subject in mental and verbal processes. Mental processes are verbs that express reflection of our own experience of the world, while verbal processes are verbs that express reaction in the form of saying. These processes are further described in section 2.1 of this thesis.

Many of the studies conducted so far suffer, however, from the fact that they have compared the Norwegian component the International Corpus of Learner English (NICLE) (Granger, Dagneaux, Meunier, & Paquot, 2009) to text types that are not necessarily justified for comparison. The key problem with this approach is that the NICLE texts are argumentative, and the writers were given prompts that asked for their personal opinions and feelings. The purpose of this present investigation is to examine Norwegian advanced learners' usage of *I*

as subject in mental and verbal processes by comparing them to other argumentative texts from native-speakers of English.

The present study will also investigate three recurrent word-combinations that are connected to clauses with mental processes, namely *I think*, *I feel*, and *I know*. Lastly, the study will investigate the recurrent word-combination *I would say* as that is connected to verbal processes. The reasoning behind including these four recurrent word-combinations is because Ädel and Erman (2012) have suggested that non-native speakers overuse *I* framed constructions in their texts. Investigating these four combinations will hopefully bring some answers as to whether this is true for Norwegian L2 learners.

One subcorpus comprising native speaker novice writer texts and one subcorpus comprising native-speaker expert texts have been used for comparison with the learner texts from NICLE. In addition to that, the present study separates itself from other studies in that it includes one novice and one expert subcorpus from the learners' native language. This is because little is known about what causes potential overuse of *I* from Norwegian advanced learners of English. The present study will investigate two categories that might influence the Norwegian advanced learners' usage. *Transfer* from their L1 is the first category, while *developmental factors* are the second category. The category of developmental factors is used in instances where the novice subcorpora share characteristics which are different from the expert subcorpora. There is also a third category named *other*, which refers to cases where neither transfer from L1 or developmental factors are detected. However, this category is outside the scope of the present project and will not be investigated.

The method used to detect transfer and developmental factors in this study is called the *Integrated Contrastive Model* (ICM) and will be outlined in the method and material section (see section 3.1-3.2).

1.1 Research questions and hypothesis

The present project's hypothesis is as follows: ***Norwegian advanced learners of English tend to overuse I as subject in clauses with mental and verbal processes, and in I framed constructions, due to transfer from Norwegian and developmental factors.*** In order to shed light on this hypothesis, the following three research questions will be explored.

Research question 1:

    a)   **Do Norwegian advanced learners of English overuse *I* as subject in clauses with mental and verbal processes in their texts?**

The thesis explores this research question by analyzing all clauses with *I* as subject in comparable argumentative texts. The hypothesis in this study is that overuse from the Norwegian advanced learners is taking place.

    b) **Do Norwegian advanced learners of English overuse the recurrent word-combinations *I think*, *I feel*, *I know*, and *I would say* in their texts?**

This research question will be investigated by comparing the frequency of usage between a subcorpus from the Norwegian advanced learners and a subcorpus from English expert L1 writers.

Research question 2:

    Is there any evidence that transfer from Norwegian influences the Norwegian advanced learners of English, and, if so, how?

The present study investigates whether L1 transfer from Norwegian can explain overuse of *I* in advanced learners' writing. The hypothesis is that overuse is taking place partly because of transfer, due to the challenges of writing in a foreign language.

Research question 3:

    Is there any evidence that developmental factors influence the Norwegian advanced learners of English, and, if so, how?

The category of developmental factors refers to novice-expert differences in the subcorpora included. Developmental factors occur if all three novice subcorpora share characteristics, but

also in cases where the NICLE writers resemble either the L1 Norwegian or English novice writers. The present study's hypothesis is that overuse is connected to developmental factors.

1.2 Plan for the thesis

The structure of the thesis is as follows: chapter 2 introduces the theoretical framework of this study and present previous research. In chapter 3, the methodological choices made for this thesis will be outlined, and the material used will be presented in detail. Chapter 4 will present the results from the comparison between the L1 expert writers. Chapter 5 follows the same pattern as chapter 4, but here, L1 novice writers are compared. Chapter 6 compares the Norwegian advanced learners to L1 expert English writers to check whether overuse take place. In chapter 7, all the subcorpora are compared to see whether the Norwegian advanced learners are influenced by transfer from Norwegian and developmental factors. Chapter 8 includes a conclusion.

2.   Theoretical framework and previous studies

The first section of this chapter (2.1) will explain in detail what mental and verbal processes are with concrete examples from NICLE. Section 2.2 will investigate the term *humble subjectivity* and outline why it is used in the investigation of the recurrent word-combinations *I think/jeg tror* and *I feel/jeg føler*. In section 2.3 previous conducted studies on *I* as subject in texts from ICLE and NICLE writers are outlined. Section 2.4 will explain in detail what recurrent-word combinations are and outline previously conducted studies. Lastly, section 2.5 will provide a summary of the chapter.

2.1 Mental and verbal processes

The terms "mental process" and "verbal process" have been taken from systemic-functional grammar (SFG). In SFG, a clause is considered to contain three kinds of meaning, known as "metafunctions": the experiential, the interpersonal, and the textual (Thompson, 2014, p. 30). The language used to talk about the world is the experiential; the language used to interact with other people is the interpersonal; and organizing language to fit in its context is the textual (Thompson, 2014, p. 30). In the experiential metafunctions, processes (verbal groups) are the main part of the clause, because the clause is primarily centered around the action, state, or event that the participants are involved in (Thompson, 2014, p. 92).

A basic rule applies to mental and verbal processes. Mental processes can be described as verbs expressing our own experience of the world, our consciousness, while verbal processes can be described as words of saying. Together, mental, and verbal processes are *inner* experiences that represent reflection and reaction (Holmes & Nesi, 2009, p. 58). The following section will give a more detailed overview of these two categories.

2.1.1 Mental processes

There are clear distinctions between something that happens in the external world and something that happens in the internal world of the mind (Thompson, 2014, p. 97). Examples of verbs that express mental processes are: *thinking*, *imagining*, *wanting*, *seeing*, *liking* etc. The clause "She could hear his voice" illustrates a typical clause with a mental process, as the person who hears is not acting out anything, but undergoing the process of hearing. Mental

processes are different from other processes in terms of semantics, but there are grammatical reasons for putting them in a separate category. First, there needs to be at least one human participant involved for a mental process to take place (further referred to as the *Senser*). Secondly, the kind of entity that fills the role of the other participant (further referred to as the *Phenomenon*) can be a person, a concrete object, and other things. Thirdly, mental processes usually differ from other processes in terms of tense, as the simple form is most commonly used. Examples of mental processes in the NICLE subcorpus are illustrated in table 1:

Table 1: Mental processes

Senser	Process: Mental	Phenomenon
I	think	that historical and philosophical subjects are a lot different from civil engineering
I	feel	sorry for people with no dreams
I	believe	that the world is still full of dreams just waiting to be exposed

2.1.2 Verbal processes

Typical verbs that express verbal processes are: *say*, *reply*, *argue* etc. (Thompson, 2014, p. 106). There is one participant, referred to as the *Sayer* in every verbal process, and that is a person. Hence, *I* works as the sayer in this study. Another participant that can be involved in a verbal process is a *receiver*; the person whom the saying is addressed (Thompson, 2014, p. 106). In other cases where the process is directed at, rather than addressed to, the participant is referred to as the *target*. What differentiates the target from the receiver is that the target does not need to be a human being. The message itself is referred to as *Verbiage*. Verbal processes are exemplified in tables 2 and 3 below with material from NICLE. The reasoning behind having two tables is to show that both sayer and verbiage can appear first in a clause.

Table 2: Verbal processes

Sayer	Process: Verbal	Verbiage
I	would say	that crime does not pay
when I	mention	untrue violence

Table 3: Verbal processes

Verbiage	Sayer	Process: Verbal
All of the examples	I	mentioned here

2.2 Humble subjectivity

Granger (1998) identified differences in literary traditions, comparing writers with Anglo-American background to people of continental European background. Readers of American or British background are likely to use *I* as a subjective stance when expressing authority and emphasis. Granger (1998) also identified that readers from continental European backgrounds usually introduce their opinions with statements of attitude (like *I think*, *I feel*) to indicate *humble subjectivit*y. This term is connected to *hedging*, which is a type of language used to protect our claims. Using caution when expressing claims can protect them from being easily dismissed, and it helps to indicate how certain we are in relation to the opinions and statements we present (University College London, n.d.). Vass (2004) illustrates the use of *humble subjectivity* by referring to American law review articles, where the writers need to reflect the fact that no one can be absolutely sure of a proposed hypothesis. They do this most notably through the use of modal verbs, as this prevents potential negative reactions if their claims and predictions made in their articles do not prove to be correct. Two of the recurrent word combinations included in the present project, namely *I think,* and *I feel* contain verbs that are not part of the core modal verbs, but they can express modal meaning. The examples 2.1 and

2.2 from NICLE below illustrate how *I feel*, and *I think* can be used to reflect humble subjectivity:

(2.1) <u>*I think*</u> *the reason we are afraid of our dreams and hopes today are our fears of failing* (NICLE, 27).

(2.2) *But although <u>I feel</u> that the teacher training should contain more methods and practise, I also think it important to know your subject well as a teacher* (NICLE, 100).

As can be seen from both these examples, the writers are using the verbs to express modal meaning, in that they are making a claim without being too assertive, and therefore open for people to disagree with their opinions.

Granger (2008) mentioned that humble subjectivity was a trait found among readers from continental European countries. It is therefore interesting to explore whether this is a pattern found among NICLE writers, and not among the English L1 writers. If this is so, cultural differences might partly explain the differences in usage of *I think* and *I feel*.

2.3 Studies conducted on the use of *I* in texts.

According to Petch Tyson (1998, p. 108), writer visibility can be defined as an interpersonal function of language within the text which is used to express personal feelings and attitudes to communicate with the readers. Writer visibility features occur in several pronoun groups like first person plural pronouns *(we, us, our)*, second person pronouns (*you, your*), and first-person singular pronouns *(I, me, my, myself)*. Petch Tyson's article explores writer visibility, including first-person singular pronouns, by comparing the French, Dutch, Swedish, and Finnish section of the International Corpus of Learner English (ICLE) to a comparable US corpus. Petch-Tyson points out that the type of environment together with the function in which *I* occurs can be revealing (Petch-Tyson, 1998, p. 111). For instance, in the comparable US corpus, around half of all occurrences of *I* were found together with verbs in the past tense. Many of these represented personal experiences, like the examples from the article (1998, p. 113) show below.

(2.3)   I honestly ran out of my room

(2.4)   I looked through the rest

In the ICLE corpora, however, only the Finnish corpus recounted personal experience or feelings (Petch-Tyson, 1998, p. 111). The main function of *I* in the French, Dutch and Swedish non-native speaker data was about the writer within the context of the piece of discourse, saying something about the function of the writer or what the writer thinks, as shown in the examples below (Petch-Tyson, 1998, p. 114)

The function of the writer:

(2.5)   I want to explain that

(2.6)   I will give my arguments for

What the writer thinks:

(2.7)   I think that imagination is a

(2.8)   I think I have made it clear

Petch-Tyson argues that differences in W/R visibility may have to do with different perceptions of the writing task or to the topics of the essays (Petch-Tyson, 1998, p. 117).

Hasselgård (2009, p. 122) studied the frequent use of *I* as subject, comparing texts from the NICLE with the British component of the International Corpus of English (ICE-GB), which is a corpus comprising expository texts. The results from her study of *I* as subject is presented in table 4 below:

Table 4: Results from Hasselgård's (2009) study of *I* as subject. Frequency per 10.000 words

The use of I as subject	Frequency per 10.000 words
NICLE	53.2
ICE-GB conversation	285.4
ICE-GB student essays	13.6
ICE-GB academic writing	5.3
ICE-GB press editorials	0

It is apparent from this table that Hasselgård's findings show overuse of *I* as subject from Norwegian advanced learners in written texts. The use of *I* as subject occurs approximately four times as often in NICLE than in ICE-GB student essays which is the most comparable text type in Hasselgård's study.[1] In her study, Hasselgård (2009, p. 132) found that in NICLE, the majority of self-reference subjects were found with mental and verbal processes, and the verbs *think* and *say* stood out as typical examples. A possible explanation for the overuse found among NICLE writers compared to the other native speaker written genres is that the essay questions given to the NICLE writers "might invite personal views, thus triggering the thematization of the writer as the source of information" (Hasselgård, 2009, p. 132)

Petch-Tyson (1998) and Hasselgård (2009) both acknowledge that their results of *I* as subject might have been influenced by how the ICLE and NICLE writers perceive the presented task descriptions. The suggested task descriptions from ICLE (Granger et al., 2009*)* below clearly invite the writer to share personal opinions and feelings:

---

[1] Hasselgård (2009) was in this study looking to manually detect thematic structures. Thus, a selection of 23 texts that dealt with crime was made.

(2.9)   Some people say that in our modern world, dominated by science and technology and industrialization, there is no longer a place for dreaming and imagination. What is your opinion?

(2.10)   Most university degrees are theoretical and do not prepare students for the real world. They are therefore of little value.

Norwegian pupils in lower and upper secondary school and students at university level are often called out by teachers and lecturers for being too personal in their text. Recski (2004) argues that there lies a great bit of irony in accusing learners of being too personally involved in their texts when the task descriptions ask for their personal opinions. Personal references and subjective attitudes are tough to avoid in those cases according to Recski (2004).

Hasselgård (2009, p. 122) mentions that the ICE-GB texts are more expository rather than argumentative, and one of the characteristics of a well-written expository text is that the writer leaves out personal opinions and emotions. Because of this, Hasselgård (2009, p. 137) mentions in her concluding remarks that "the comparison with Norwegian suffers from a mismatch of genres. There is a particular need for investigating more genres of spoken and written Norwegian, including argumentative prose by professionals as well as students". Hence, the present project will compare NICLE texts to argumentative texts from English and Norwegian L1 novice and expert writers to further examine to what extent overuse of *I* as subject in clauses with mental and verbal processes is taking place. As both Petch-Tyson's (1998) and Hasselgård's (2009) studies are to a large degree quantitative where explanations of overuse are based on careful perceptions, investigating factors that might influence the NICLE writers' usage of *I* as subject is needed. This present study seeks to explore this by investigating traces of *transfer* from Norwegian and *developmental factors*.

Usage of *I* as subject from NICLE writers is far less frequent compared to NS conversation. Hasselgård (2009, p. 132) argues that in press editorials, a clearly opinionated genre, *I* does not occur in her material, and suggests that the learners use features from conversational English, such as *I*, in their written output. Since the present project covers written material exclusively, influence from spoken language will not be examined, but this is an important issue for further research.

Heidi Fossan (2011) studied the use of the pronoun *I*, comparing 100 NICLE texts to 100 texts from the Louvain Corpus of Native English Essays (LOCNESS) (Université catholique de Louvain, 2017). LOCNESS comprises argumentative texts and is therefore suitable for comparison with NICLE. Fossan's (2011, p. 130) findings reveal that in total, *I* occurs 90.41 times per 10.000 words in NICLE, while it occurs 30.15 times per 10.000 words in the LOCNESS. Thus, according to Fossan's findings, Norwegian advanced learners use *I* three times as often compared to native speakers in argumentative texts. This overuse seems very high, and there is no obvious explanation as to why Norwegian learners feel the need to guide the reader through the relatively short texts as often as they appear to do (Fossan, 2011, p. 130).

Fossan did investigate whether transfer from Norwegian L1 influence the NICLE writers in their usage of 1st person singular pronouns by comparing their texts to a Norwegian L1 corpus called NOESS and LOCNESS. NOESS is the abbreviation for NOrwegian ESSays, a small corpus of 13 argumentative essays written by Norwegian high-school students (Fossan, 2011, p. 81). Her findings indicated that the high frequency of 1st person pronouns among NICLE writers was not caused by transfer from Norwegian L1, as the frequency found in NOESS was not significantly higher than the frequency found in LOCNESS (Fossan, 2011, p. 111).

Fossan (2011) also investigated the influence of transfer from Norwegian by examining the differences in frequency among four corpora comprising academic texts. Here, the Norwegian part of the Varieties of English for Specific Purposes database (N-VESPA), an L2 corpus comprising English essays written by Norwegian learners were compared to a Norwegian L1 corpus comprising texts from bachelor students, named The Norwegian Bachelor Assignments (NOBA) (Fossan, 2011, p. 82). In addition to that, two corpora from British students were included, named The British Academic Written English Corpus (BAWE and BAWE-ling). The difference is that BAWE-ling is a subcorpus within BAWE, consisting of text from the linguistics section. Fossan (2011, p. 114) argues that transfer seems like a possible explanation for the overuse of 1st person pronouns (singular and plural) found in N-VESPA. This is because her results showed that both N-VESPA and NOBA had a significantly higher frequency of 1st person pronouns compared to BAWE and BAWE-ling. In her final remarks on the L1 transfer issue, Fossan (2011, p. 120) suggests that further research needs to be conducted with more suitable L1 Norwegian corpora, which ideally would include both students and professional writing in different genres and disciplines. The present project aims to investigate whether transfer is influencing Norwegian L2 advanced learners in the

argumentative genre. Hence, subcorpora comprising argumentative texts from Norwegian and English L1 novices and experts are compared to NICLE.

Paquot et al. (2013) found that Norwegian learners overuse features of W/R visibility. There are, however, fewer features of W/R visibility in discipline-specific texts, thus suggesting that learners to some extent adapt to genre requirements. In the NICLE corpus, which consists of argumentative essays, 2.550 per 100.000 words were first-person pronouns, while in the Norwegian part of the Varieties of English for Specific Purposes database (N-VESPA), which consists of discipline-specific texts, 1.366 per 100.000 words were first-person pronouns. *I* occurs 767 times per 100.000 words in NICLE, and 703 times per 100.000 words in VESPA. LOCNESS was also included in that statistic and contained 241 occurrences per 100.000 words (Paquot et al., 2013, p. 4).

Fossan (2011) and Paquot et al. (2013) suggest that NICLE writers use *I* as subject more frequently than LOCNESS writers. These findings are valuable as LOCNESS is the native-speaker component of ICLE, and hence, a justified choice for comparison with NICLE. However, Hasselgård's (2009) findings with regards to overuse of *think* and *say* in self-reference clauses, and Paquot et al. (2013) findings on the overuse of *I think*, are not sufficient enough results to draw any conclusions with regards to overuse of *I* in mental and verbal processes. This is because there are various other words that can be used together with *I* in clauses with mental and verbal processes, as I outlined in section 2.1. For that reason, the present study examines all the occurrences of mental and verbal processes connected to *I* as subject in the selected subcorpora.

2.4 Recurrent word-combinations

Recurrent word-combinations are continuous strings of words that occur more than once in identical form (Altenberg, 1998, pp. 101-122). Typical examples of recurrent word-combinations are; *and the*, *in a*, *I think*, *I think that*, *out of the*, *painfully clear*. Altenberg (1998, p. 102) found that in the spoken London-Lund corpus, an estimated 80 per cent of the words in the corpus were part of a recurrent word-combination. This shows how recurrent word-combinations is a necessity in our everyday language. Recurrent word-combinations occur overall more frequently in native speaker production compared to non-native

production. This is because native speakers have a broader repertoire of types, and display variety of form to a greater extent (Ädel & Erman, 2012, p. 83). Additionally, specific groups of recurrent word-combinations are found to be underused among non-native speakers. Examples of this are recurrent word-combinations, such as *in the context of*, and combinations of adverb and adjective, such as *painfully clear* (Ädel & Erman, 2012, p. 83). However, non-native speakers overuse other groups of recurrent word combinations, such as *I* / *We*-framed constructions, like *I claim that/we could say that*, and combinations like *all over the world* (Ädel & Erman, 2012, p. 83).

The four recurrent word-combinations included in the present project are all *I* framed constructions, which means that according to findings from Ädel and Erman (2012), there is a high chance of overuse taking place among NICLE writers as well in these constructions. However, if the results from these four combinations do show overuse when compared to native speakers, it cannot be safely concluded that NICLE writers overuse recurrent word-combinations with *I* as subject in general when compared to native speakers. This is because there are various other combinations that can be used together with *I* as subject as well that are not included in this study, such as *I believe*, *I might*, *I would*, *I could*.

2.4.1 The complexity of the recurrent word combination *I think*

It could be claimed in the light of present-day semantic theory, *I think* is organized as a prototype with Cogitation (Aijmer, 1997, p. 12). Cogitation is the prototypical meaning of the verb *think* (*to think of*, *I am thinking*). Aijmer (1997, p. 10) argues that there is an "epistemological distinction between two functions of language: indicating facts and expressing the state of the speaker". These two distinctions are referred to as *expressive* and *fact indicating*/*objective style*. When *I think* functions as expressive, the aim is to express the speaker's emotions (cogitation), as example 2.11 below from NICLE shows:

(2.11) The reason to this is that I'm working part-time as a teacher at a high school close to where I live. I think I've learned a lot from this, and I get the experience that I want   (NICLE, 88).

Here, the person is reflecting on the experience she has gained by working as a teacher in an anticipating way.

When *I think* functions as fact indicating or objective in style, the speakers refer to their opinions and beliefs in an objective and informative manner. Often, the complementizer *that* comes after *I think* to indicate this style, but not necessarily. Example 2.12 and 2.13 below from NICLE illustrate that *I think* can be used in the fact indicating/objective style with or without the *that*-complementizer:

> (2.12) <u>I think that</u> the television is a very good tool for manipulating the masses, and the pictures from the 11th September crash in New York going over the TV-screens all over the world was the terrorists way of manipulating us (NICLE, 29).

> (2.13) What <u>I think</u> should be done is to make a better selections of the boys already from the first day they get in touch with the army, this will also make less people be in the army from day one   (NICLE, 2).

In both these examples, the writer is giving his/her opinions on the subject they are writing about, and there is no expression of actual thinking as a mental activity (cogitation) involved in the process.

In the present study, *I think* has been directly translated into *jeg tror* in Norwegian, which is not free of implications. As we have seen, there are several meanings connected to the recurrent word-combination *I think*, which means that other lexical items can be used to convey the same message to the reader. An example of this is the combination *I believe*, that is commonly used among native speakers in the fact indicating/ objective style. The situation is similar in the Norwegian language where for instance *jeg syntes* can replace *jeg tror* in the fact indicating/ objective style. Thus, comparing *I think* to *jeg tror* can be problematic as other lexical items can replace them, even though the same meaning is conveyed. However, since both *I think* and *jeg tror* can be used in the expressive and fact indicating/ objective style in their respective L1 languages, comparing them seems justified. A more thorough approach would have been to investigate all the lexical items that can replace *I think* and *jeg tror* while conveying the same meaning, but that is outside the scope of the present project.

2.4.2 Studies conducted on the use of *I think* in written and spoken discourse

Paquot et al. (2013) studied the usage of the recurrent word-combination *I think* among NICLE writers and found that *I think* occurs 108 times per 100.000 words in NICLE, while only 18 times in VESPA which is close to LOCNESS with 16 occurrences (Paquot et al., 2013, p. 5). Based on this study, the NICLE writers are likely to show overuse in the present study as well. However, if overuse is found, it remains to be seen whether transfer from Norwegian L1 or developmental factors influence the NICLE writers.

Ringbom (1998, p. 41) compared the usage of the recurrent word-combination *I think* between seven western European learner corpora from the ICLE and LOCNESS. All the essays were argumentative and dealt with a variety of different topics. The table below compares the results of the recurrent word-combination *I think* from the eight corpora: [2]

Table 5: *I think*: frequency per 10.000 words in Ringbom's (1998) study

Corpora:	LOCNESS	FRE	SPA	FIN	FINSW	SWE	DUTCH	GERM
Frequency:	3	10	5	7	16	16	6	9

It is apparent from table 5 that writers with Swedish as their L1 show a predilection for the recurrent word-combination *I think*. French and German advanced learners seem to have a preference towards the clause as well, although they had a lower frequency than the Swedish advanced learners. The Swedish writers use *I think* over twice as much as the Finish writers, which is interesting, considering the fact that Sweden and Finland are neighboring countries. This makes it difficult to predict the frequency in NICLE, but it is reasonable to assume that the usage is close to the Swedish writers as the Norwegian and Swedish language share many similarities, and people from both countries learn English from a young age at school.

---

[2] FINSW refers to Finish-Swedish writers.

Baumgarten and House (2010) compared the usage of *I think* between a group of English L1 speakers from the US and two groups of L2 English speakers. It is important to note here that this study examined the usage in spoken language. The participants were involved in one-on-one conversations with the researchers, and they talked about specific topics of general interest, like "the role of English in the world" and "men and women in the contemporary arts scene" (Baumgarten & House, 2010, p. 1186). Table 6 presents the findings from the conversations:

Table 6: *I think* in relation to the total of I-utterances (Baumgarten & House, 2010, p. 1188)

Speakers:	L1		L2 group 1		L2 group 2	
Function:	**Percentage**	**Raw frequency**	**Percentage**	**Raw frequency**	**Percentage**	**Raw frequency**
Total utterances of *I*	100 %	120	100 %	80	100 %	188
Total utterances of *I think* in relation to *I*	17.5 %	21	53.7 %	43	25.5 %	48

As can be seen from table 6, *I think* is used more frequently in the L2 groups compared to the L1 group. The difference in percentage between the L1 group and the L2 group 2 is not that considerable, however, this is due to the fact that the L2 group 2 had many more instances of *I* overall, which again caused the percentage of *I think* to drop in relation to *I* as subject.

Since the present project investigates Norwegian advanced learners' usage of *I think* in written discourse, it might seem strange to compare them to English L1 native speakers and English L2 speakers' usage of *I think* in spoken discourse. There are differences between written and spoken discourse on a general level, also in terms of formality, but whether those differences affect the usage of the recurrent word-combination *I think* is unknown. Therefore, it is quite interesting to see whether the NICLE writers will resemble the results from the L1 native speakers or the L2 speakers. However, since the spoken material is compared to written material in this case, great caution will be applied when interpreting the results.

2.4.3 Hasselgård and Johannson's study on *I would say*, and the complexity of the recurrent word-combination

Hasselgård and Johansson (2011, p. 49) found that NICLE writers overuse the recurrent word-combination *I would say* compared to LOCNESS writers. Their results are shown in table 7 below:

Table 7: Raw frequencies and relative frequencies *of I would say* per 100.000 words in Hasselgård and Johansson's (2011) study

Raw frequencies and relative frequencies per 100.000 words.		
Corpus:	**NICLE**	**LOCNESS**
Instances:	27	5
Relative frequency:	12.8	1.5

In LOCNESS, instances of *I would say* were found functioning as a stance marker and as an introduction to a conclusion (Hasselgård & Johansson, 2011, p. 50). The same pattern was found among the NICLE writers who also used *I would say* in conclusions and as stance marker.

One difference between *I would say* and the closest Norwegian translation *jeg vil si* (the translation used in the present project) is that the English modal *would* has the past tense form, while the Norwegian modal *vil* has the present tense form (Hasselgård & Johansson, 2011, p. 51). This is a potential source of transfer errors, but Hasselgård and Johansson (2011, p. 51) found that the expression *I will say* only occurred three times in the NICLE corpus, which means that NICLE writers do not directly translate the recurrent word-combination *jeg vil si* from Norwegian into English.

Hasselgård and Johansson (2011, p. 51) argue that the overuse of *I would say* among NICLE writers could be influenced by transfer from Norwegian, as the Norwegian recurrent word combination *jeg vil si* "flags the following proposition as the speaker's opinion, but not necessarily as tentative". In English, however, the past-tense form of the modal provides a

tentative ring to *I would say* (Hasselgård & Johansson, 2011, p. 51). Thus, the NICLE writers might, through transfer from their L1, use the English expression *I would say* with more assertiveness than what it actually seems to have in native speaker writing. Hasselgård and Johansson (2011, p. 52) further argue that developmental factors also might play a role among NICLE writers, since LOCNESS writers use the recurrent word-combination as well, although with less assertiveness. They conclude that further research is needed that investigate the influence of developmental factors. The present study will try to further investigate whether transfer and developmental factors influence the NICLE writers through use of the Integrated Contrastive Model. This will hopefully expand the knowledge on NICLE writers' usage of the recurrent word-combination *I would say*.

2.5 Summary

This chapter has in section 2.1 given an overview of Systemic Functional Grammar and explained the differences between mental and verbal processes (Thompson, 2014; Holmes & Nesi, 2009). In section 2.2, the term *humble subjectivity* was outlined and its relevance for the present project was discussed. Section 2.3 examined previously conducted studies of *I* as subject (Petch-Tyson, 1998; Hasselgård, 2009; Fossan, 2011; Paquot et al., 2013). The discussion showed that the there is a need for further studies that compare NICLE writers to argumentative texts from English and Norwegian L1 experts and novice writers. Fossan (2011) and Paquot et al. (2013) show that NICLE writers overuse *I* as subject when compared to native novice writers that also write in the argumentative genre. However, Hasselgård's (2009) findings with regards to overuse of *say* and *think* in self-reference clauses, and Paquot et al's. (2013) findings with regards to overuse of *I think*, are not adequate evidence to suggest that NICLE writers overuse *I* as subject in mental and verbal processes. The present study, therefore, analyses all the instances of clauses with *I* as subject in mental and verbal processes to check for overuse. Fossan (2011) suggested that transfer from Norwegian might influence L2 Norwegian writers.

Section 2.4 explained in detail the function of recurrent word-combinations and described how a previously conducted study from Ädel and Erman (2012) has found that non-native speakers of English tend to overuse *I* framed constructions. Section 2.4.2 showed that previously conducted studies (Ringbom, 1998; Baumgarten & House, 2010) have found overuse of the recurrent word-combination *I think* among non-native speakers in written and spoken

discourse. Lastly, in section 2.4.3, a study conducted by Hasselgård and Johansson (2011) was described, which showed that NICLE writers overuse the recurrent word combination *I would say* compared to LOCNESS writers.

3. Material and method

This section will explain in detail the material and the methodological choices made for this thesis. First, in section 3.1, a brief history of the development of the corpus-linguistics field will be presented to show how the focus has shifted from contrastive analysis to the Integrated Contrastive Model (ICM). Secondly, the ICM will be presented in section 3.2 to show how the model works and why it is relevant for the present study. Thirdly, in section 3.3, the material used will be presented in detail before discussing the favorable and unfavorable factors of each corpus. Then, in section 3.4, the categories used to sort the instances of *I* as subject will be presented. Lastly, in section 3.5, an overview of the quantitative and qualitative working methods used in this study will be presented.

3.1 Language transfer and the development of the Integrated Contrastive Model

According to Osborne (2015, p. 333) "transfer is the influence that previous knowledge or skills have on future learning". Specifically applied to the field of linguistics, Odlin (1989, p. 27) defines *language transfer* as "the influence resulting from similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired". Traditionally, transfer has been seen as having either positive effect which can facilitate learning or negative effect which can inhibit learning. However, as Osborne (2015, p. 333) points out, the phenomenon of transfer is more complex than what a positive-negative opposition suggests.

Robert Lado (1957) was one of the founders of modern contrastive linguistics and helped develop the term *contrastive analysis* (CA), which claims that one can predict issues encountered by leaners of a foreign language (L2) by comparing this language with the learners' first language (L1) (Gilquin, 2000/2001). Lado was careful to point out that individual differences would undoubtedly take place, but he remained confident that the learning difficulties identified by CA would prove to be stable and predictable for each language background (Osborne, 2015, p. 335).

Benson (2002, p. 68) argues that contrastive analysis was "rooted in a behaviourist theory of language learning whereby learning was equated with 'habit forming': the habits of the L1

were believed to be 'transferred' and regarded as 'interfering with' the newly-acquired habits of the L2". This hypothesis gained a lot of followers during the 1960s which led to success and a rapid development of the discipline. It later turned out that CA's predictable power and its practical applications for language teaching were far more restricted than what was initially thought (Gilquin, 2000/2001). However, with the advent of computer-based corpora, CA has regained interest among researchers because of its empirical solidity that is no longer lacking.

Interestingly, the same methods and aims that CA applies, can be used in cases with two different varieties of the same language. In the present study, Norwegian L2 English and native-speaker L1 English are the two varieties. This type of comparison is referred to as Contrastive Interlanguage Analysis (CIA) The development of such an approach is also due to the advent of computer-based technology which makes it possible to create extensive corpora databases (Gilquin, 2000/2001). Nowadays, there exists a large number of well-designed corpus databases that consist of material by foreign learners – such as ICLE.

The two disciplines outlined above, CA and CIA, can be combined in what Granger (1996) named the Integrated Contrastive Model. Starting with the CA approach of comparing two different languages (in our case, novice Norwegian L1 vs. novice English L1, and expert Norwegian L1 vs. expert English L1), one can predict areas that might be problematic for foreign language learners. The CIA approach is, on the other hand, diagnostic, and here, explanations are sought for potential overuse among the NICLE writers compared to the English expert texts. In this present study, transfer and developmental factors are analyzed in the diagnostic phase.

3.2 The Integrated Constrastive Model: the tool used to help explore the research questions

Granger (1996, p. 46-47) first introduced the integrated contrastive model (ICM) before Gilquin (2000/2001, p. 100-101) further developed the model. The purpose of the model is to shed new light on how contrastive corpus data can assist in explaining characteristics of learners' interlanguage (Gilquin, 2000/2001, p. 95).

A recapitulation of the research questions is important at this stage, as the ICM works as the tool to investigate them.

Research question 1:

a) **Do Norwegian advanced learners of English overuse *I* as subject in clauses with mental and verbal processes in their texts?**

b) **Do Norwegian advanced learners of English overuse the recurrent word-combinations *I think*, *I feel*, *I know*, and *I would say* in their texts?**

Research question 2:

Is there any evidence that transfer from Norwegian influences the Norwegian advanced learners of English, and, if so, how?

Research question 3:

Is there any evidence that developmental factors influence the Norwegian advanced learners of English, and, if so, how?

The material in the present project consists of five different subcorpora. In the investigation of *I* as subject in mental and verbal processes, the material has been limited to 30 texts from each subcorpus due to practical constraints, as every instance of *I* as subject need to be analyzed manually. In the investigation of the four recurrent word-combinations, 100 texts from each subcorpus have been included, as the sorting process is more practical.

In order to follow the principles of the ICM, two of the selected subcorpora labeled ENG NEWS and NOR NEWS consist of argumentative newspaper articles written by Norwegian and English "expert" writers in their respective L1. Additionally, two "novice" writer subcorpora labeled NOR ESSAYS and LOCNESS are also included, and these consist of argumentative texts written by Norwegian and English non-professional writers in their L1.

Lastly, the most central subcorpus included is the Norwegian component of the International Corpus of Learner English (NICLE). These argumentative texts are written by Norwegian advanced learners of English, making NICLE the only subcorpus comprising L2 texts in this study. The five subcorpora will be described in detail in section 3.3.

The figure below is an adapted version of Gilquin's model, developed by Rørvik (2013, p. 17) and demonstrates the chosen approach for the present study:

Figure 1: Rørvik's (2013) adapted version of the Integrated Contrastive Model
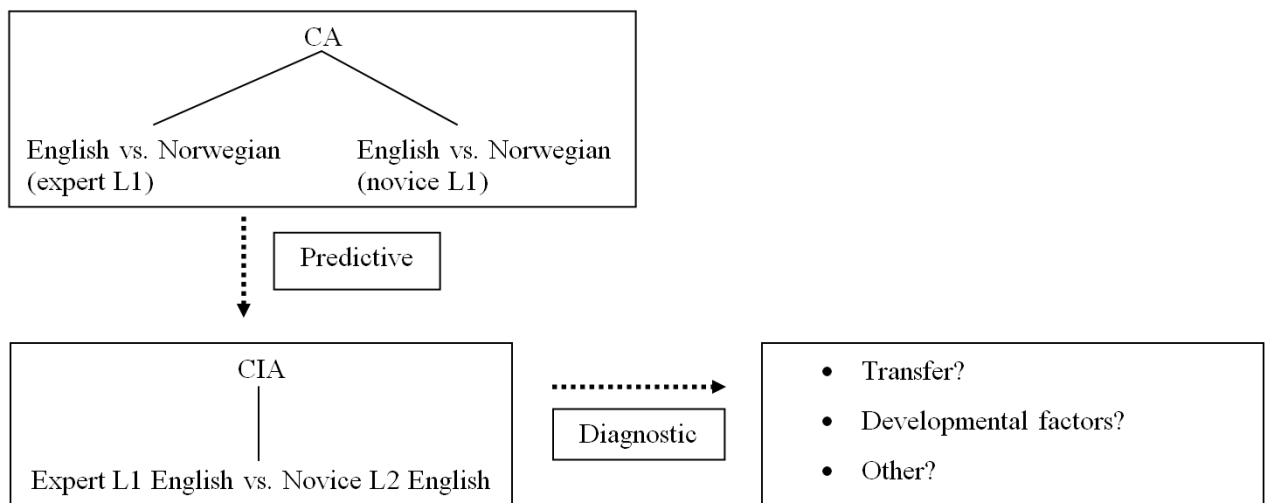
The process begins with a contrastive analysis (CA) which involves comparing ENG NEWS to NOR NEWS and LOCNESS to NOR ESSAYS with the purpose of predicting whether the two languages differ in usage in both novice and expert writing. Then, a contrastive interlanguage analysis (CIA) will be performed, testing the predictions made based on the CA results. Here, the NICLE texts will be compared to the ENG NEWS texts to see whether overuse is found among Norwegian advanced learners, and if so, to what extent.

The process ends with the diagnostic phase where different reasons for overuse in the NICLE texts compared to the ENG NEWS texts are analyzed. In this part of the ICM model, three factors that influence overuse are included, namely, *transfer*, *developmental factors*, and *other*. Traces of transfer from Norwegian are checked by comparing the NICLE texts with the NOR NEWS texts. The next category is developmental factors, and the focus here is to check for differences between novice and expert writers. If the novice texts differ from expert texts in their respective L1s, and the NICLE texts differ from the L1 expert texts in similar ways, we have indications suggesting that developmental factors influence novice writers. The last

category, *other*, is used to account for instances where neither evidence of transfer or developmental factors is found. However, other influencing factors will not be investigated as that is outside the scope of the present project.

3.3 The material

The following sections will present the five different subcorpora included in this study. Each subcorpus will be assessed by looking at both favorable and unfavorable factors.

3.3.1 NICLE

NICLE is the Norwegian component of the International Corpus of Learner English (ICLE) (Granger et al., 2009). These argumentative texts are not discipline-specific texts; they discuss various topics, often controversial ones like crime, climate change, and feminism. The ICLE contributors are of approximately the same age (around 20 years old), they are all advanced learners in that they are university undergraduates in English Language and Literature (Granger, 1998, p. 9). The learners vary in features such as experience, sex, mother tongue, region, and other foreign languages (Granger, 1998, p. 9).

The criteria set for the L2 subcorpus are (a) that the texts are argumentative and that (b) the learners home country is Norway with (c) Norwegian as their L1. All these criteria are met in the NICLE subcorpus which makes it a satisfying choice for this present study.

3.3.1.1 Favorable factors

The NICLE learner group is in many ways homogenous. All the contributors attend university, and 93 percent have studied English for less than a year at university level, indicating similar levels of English proficiency (Fossan, 2011, p. 73). The importance of comparing findings with other similar studies is more relatable when the same corpus is used. Meyer (2002, p. 103) supports this claim by saying: "[I]t is therefore most desirable to work with a corpus already available not just to decrease the work time but to add to the growing body of work that has been based on a given corpus".

3.3.1.2 Unfavorable factors

Although all the texts are argumentative, the task descriptions given to the students are not available. However, as I outlined in the discussion of previous studies (see section 2.3), the ICLE website (Granger et al., 2009) has listed suggested essay titles that clearly invites the writer to share personal opinions and feelings.

3.3.2 ENG NEWS and NOR NEWS

These two subcorpora contain texts from English and Norwegian newspapers, namely *The Guardian* for the English material, and *Dagbladet* for the Norwegian material.[3] The texts are argumentative in the form of opinion columns and editorials. By incorporating Norwegian expert texts, potential transfer from Norwegian can possibly be detected in the NICLE texts. The English expert texts are valuable as they represent the norm the NICLE writers want to follow, and they are needed for comparison with the L1 novice subcorpora to check for influence of developmental factors.

3.3.2.1 Favorable factors

The two subcorpora contains only texts between 600 and 1,000 words in length since that matches the average word length in NICLE (661 words). Both *The Guardian* and *Dagbladet* are respected newspapers in their home country, which heavily indicates that the writers are on a high level of writing proficiency in their respective L1. Hence, the two subcorpora correlate well with one another in terms of expert L1 writing.

3.3.2.2 Unfavorable factors

It can be argued that comparing NICLE texts to expert L1 texts and using that as a yardstick is unfair to the Norwegian students. Lorenz (1999, p. 14) criticizes expert writing in learner corpus research, arguing that it is "both unfair and descriptively inadequate". Hyland and Milton (1997, p. 184) also take a stand against this approach, pointing out the "unrealistic

---

[3] The website of *The Guardian* can be accessed at www.guardian.co.uk, whereas *Dagbladet* can be accessed at www.dagbladet.no

standard of 'expert writer' models". However, while Gilquin and Paquot (2008, p. 45) say that native student writing is arguably better for comparison with EFL learner writing to evaluate interlanguage, they add that it is "highly questionable whether findings from such comparisons can make their way into the classroom". Furthermore, Granger (2004, p. 133) stresses that advanced learners (NICLE writers included) are "getting close to the end point of the interlanguage continuum" and are "keen to get even closer to the NS [native-speaker] norm". Leech, (1998, p. xix) follows up by arguing that "native-speaking students do not necessarily provide models that everyone would like to imitate". To elaborate on this, native students have on a regular basis produced more dangling participles than English as a Foreign Language (EFL) learners (Granger, 1997), and spelling mistakes have been found (Cutting, 2000). The best way to approach the argument of what "norm" is most suitable, is to consider the aim of the comparison. Paquot (2010, p. 72) says that expert writing has a major role to play in learner corpus research when pedagogical applications are the aim of the comparison between learner and native speaker material. Ädel explains the rationale behind using expert texts as a yardstick in the following way:

> On the one hand, it can be argued that in order to evaluate foreign learner writing by students justly, we need to use native-speaker writing that is also produced by students for comparison. On the other hand, it can also be argued that professional writing represents the norm that advanced foreign learner writers try to reach and their teachers try to promote. In this respect, a useful corpus for comparison is one which offers a collection of what Bazerman (1994, p. 131) calls 'expert performances' (Ädel, 2006, p. 206-207).

Given the arguments from (Ädel, 2006; Paquot, 2010; Leech, 1998; Granger, 1994, 1997; Gilquin & Paquot, 2008) together with the fact that this present study is intended for didactical purposes, expert L1 English seems to be justified as the ideal standard to follow.

3.3.3 LOCNESS and NOR ESSAYS

The Louvain Corpus of Native English Essays (LOCNESS) is the native-speaker component of ICLE and consists of British pupils' A-level essays, and texts from British and American university students (Université catholique de Louvain, 2017). However, in the present study, the selection only contains American texts. NOR ESSAYS consists of essays written by Norwegian upper-secondary students. These argumentative texts were downloaded from a Norwegian internet site called *daria.no* where upper-secondary students can post their essays.

The reasoning behind having two novice L1 subcorpora is to check for developmental factors that might influence the NICLE writers.

3.3.3.1 Favorable factors

The text length of these two subcorpora varies more than the others, fluctuating between 400-2300 words. However, the average text length is 1028 words which works well when compared to the NICLE texts. Granger (1998, p. 13) points out that LOCNESS is directly related to ICLE. This is due to the fact that writers of both NICLE and LOCNESS are of nearly the same age and are at the same study level (Fossan, 2011, p. 75).

The LOCNESS texts included in this present study have been taken from the American part of this corpus because they wrote in response to similar topics as to those given to the NICLE writers. In some cases, the task descriptions have been adapted to better fit a US context, like instructing the students to write about gun control or capital punishment.

3.3.3.2 Unfavorable factors

It would have been preferable to access texts from Norwegian L1 university-age novices instead of upper secondary students, but this was not possible within the timeframe of this present study. This means that the NOR ESSAYS writers are younger than both LOCNESS and NICLE writers.

The task descriptions given to the LOCNESS and NOR ESSAYS writers are unavailable, which means that some students might be given questions where they are asked to give their own opinions, while others might receive less personal task descriptions. However, like all the other subcorpora included in this present study, the LOCNESS and NOR ESSAYS texts are argumentative and answer to contemporary topics.

Granger (1998, p. 13) stresses the disadvantages of LOCNESS in that it is a small corpus containing written material from non-professionals. However, the aim of including non-professional subcorpora for comparison with NICLE is to look for developmental factors, which means that including non-professional writing is an advantage in this case.

3.3.4 Comparable information about the subcorpora

Table 8 shows comparable information in the five subcorpora. The information presented in this table has already been discussed in the previous sections of this chapter, however, the idea here is to present an overview that summarizes key information.

Table 8: Comparable information from the included subcorpora

Subcorpus:	ENG NEWS		NOR NEWS		NICLE		LOCNESS		NOR ESSAYS	
Text language:	L1 English		L1 Norwegian		L2 English		L1 English		L1 Norwegian	
Number of texts:	30	100	30	100	30	100	30	100	30	100
Number of words:	23433	79388	26451	88010	19836	67332	33376	87430	28346	102499
L1:	English		Norwegian		Norwegian		English		Norwegian	
Home country:	UK		Norway		Norway		US		Norway	
Genre:	Argumentative		Argumentative		Argumentative		Argumentative		Argumentative	
Topic:	General argumentative topics		General argumentative topics		General argumentative topics		General argumentative topics		General argumentative topics	

The reason the "Number of texts" category is split into 30 and 100 texts is because, in the investigation of *I* as subject in clauses with mental and verbal processes, 30 texts have been examined, while in the study of the four recurrent word-combinations, 100 texts have been examined.

3.4 The categorization of *I* as subject

The process began by extracting every clause that has *I* as subject from each subcorpus. Then, the clauses were analyzed individually before being sorted into three different categories, namely *mental processes, verbal processes,* and *other.* Clauses with mental and verbal processes have already been described in section 2.1. However, many processes in clauses did not fit under these two categories, which means that a third category has been necessary. Since the focus of this thesis centers around mental and verbal processes, it is irrelevant to have separate categories for other process types. Hence, all the instances where one of these processes occurred were placed in the *other* category. The table below illustrates, with examples taken from NICLE, how the clauses have been sorted into the three categories:

Table 9: Illustrating the sorting of clauses with *I* as subject

Mental process	Verbal process	Other
I believe that the police	I am sorry to say that this kind of	when I went to the university in Trondheim
I think this all depends on	I would say that crime does not pay	I am writing this essay
I feel sorry for people with no dreams	I mention untrue violence	I was working in the mental institution

In some cases, however, it has been difficult to sort out what category is most suited for a given occurrence. O'Donnell, Zappavigna, and Whiteclaw (2009, p. 49) say that authors in the field of systemic functional linguistics differ somewhat in the criteria used to classify processes, sometimes because the author's own interpretation differs. Occurrences of *I* that have been difficult to categorize in this present study have been discussed with my two supervisors. A more thorough approach would have been to prepare all the difficult clauses to code in a survey and have several English grammar experts assign a process type to each clause. However, this was not possible due to the time span of this present project.

3.5 Quantitative and qualitative analysis

The present study mainly consists of quantitative analysis given the fact that a large proportion of the study is based on statistics provided by the subcorpora material. The comparing of subcorpora statistics is a quantitative approach, and the process of calculating whether differences are statistically significant or not is quantitative as well. However, while corpus studies in general are heavily associated with quantitative research methods, there are necessary qualitative elements included in this project.

The analytical process in this present study can be summarized in the following steps. The first step is qualitative as the different categories of *I* as subject and the recurrent-word combinations have been manually selected. In the second step, the instances of *I* from each subcorpora were categorized manually, which is also a qualitative process. Then, the process entered the quantitative analysis stage where the results were subjected to statistical scrutiny. Here, the 30/100 texts from each subcorpora were merged into one unit. The next step in the process was to look for statistical significance in the differences found between the subcorpora. An online log-likelihood calculator (Rayson, n.d.) has been used to verify whether the results are statistically significant. This tool has been essential in the present study because it provided supporting evidence concerning potential overuse from Norwegian advanced learners. The log-likelihood calculator is directly intended for frequency-based corpus studies (Rayson, n.d.). In the present study, the log-likelihood calculation was set at a 5 percent level ($p < 0.05$), meaning that the approximate significance had to be 95 percent certain or more. The log-likelihood value needed to be 3.84 or more in order to prove statistical significance. The last step of the process was qualitative where explanations were sought for the differences.

From this overview, it is apparent that quantitative and qualitative have depended on each other in the present project.

4. English and Norwegian L1 expert texts

In this chapter, the results from the L1 expert texts are presented. Section 4.1 shows the results from clauses with *I* as subject in mental and verbal processes, and here, 30 texts from each subcorpus have been used. The results from the recurrent word-combinations, namely *I think*, *I feel*, *I know*, and *I would say* (section 4.2 – 4.5) have been gathered by comparing the frequency in 100 texts from each subcorpus. As the five subcorpora differ in size, frequency per 10.000 words has also been included, but is important to note that all the frequency calculations are carried out by comparing the raw frequency from each subcorpora. The frequency per 10.000 words approach correlates well with previous studies that have also compressed the results into this format. The main limitation of the frequency per 10.000 words approach, is that the potential for *I* as subject depends on the number of clauses, not the number of words. A more thorough approach would have been to look at the frequency of *I* as subject per 1.000 clauses. However, since previous studies have operated with frequency per words, and not clauses, comparing the results would have been difficult if frequency per clause was the approach in this present study.

4.1 Clauses with mental and verbal processes

The results obtained from the CA, where ENG NEWS was compared to NOR NEWS, are presented in table 10 below:

Table 10: Frequency of *I/jeg* as subject in ENG NEWS and NOR NEWS

	ENG NEWS		NOR NEWS	
	Raw frequency in 30 texts	**Estimated frequency per 10.000 words**	**Raw frequency in 30 texts**	**Estimated frequency per 10.000 words**
***I/jeg* in mental processes**	63	26.88	26	9.83
***I/jeg* in verbal processes**	5	2.13	10	3.78
***I/jeg* in other processes**	60	25.60	26	9.83
Total instances of *I/jeg*	128	54.61	62	23.44

As can be seen from the raw frequency results in table 10, the ENG NEWS writers have significantly more instances of *I* as subject in total compared to the NOR NEWS writers.[4] The raw frequency also shows that English expert writers use *I* in clauses with mental processes significantly more often than Norwegian experts.[5] Example 4.1 and 4.2 from ENG NEWS illustrate the usage of *I* as subject in clauses with mental processes, as the writer is expressing experience from his/her own consciousness.

(4.1)   *I imagine that many would expect the Science Media Centre, established to improve the quality of science reporting, to be leading the assault* (ENG NEWS, 28).

---

[4] The difference between ENG NEWS and NOR NEWS was found to be 99.99 % certain ($p < 0.0001$).

[5] The difference between ENG NEWS and NOR NEWS was found to be 99.99 % certain ($p < 0.0001$).

(4.2)   *I've never seen those either, partly because I̲ feel guilty about not having watched the Bergman films first*   (ENG NEWS, 22).

The difference in verbal processes was too minor to prove statistical significance, but according to the present material, the Norwegian experts use *jeg* as subject in verbal processes more frequently than English expert writers.[6] In example 4.3 and 4.4 from NOR NEWS, usage of *jeg* in clauses with verbal processes are illustrated, as the writer is telling the reader about a verbal exchange between herself and other individuals.

(4.3)   *Et lite problem ble det likevel da j̲e̲g̲ ba om å få sjekket bagasjen helt frem til Trondheim.* [Translation: *A small problem nevertheless came when I asked to get the luggage checked all the way to Trondheim*][7]   (NOR NEWS, 3).

(4.4)   *J̲e̲g̲ vil også diskutere mulighetene for et tettere samarbeid mellom Norge og USA.* [Translation: *I will also discuss the possibilities for a closer cooperation between Norway and USA*][8]   (NOR NEWS, 6).

In other processes, the raw frequency shows that English experts use *I* as subject significantly more often than Norwegian experts.[9] Two examples of *I* as subject in other processes from ENG NEWS are illustrated below:

(4.5)   *With what rudimentary political acumen I̲ possessed at the time, it seemed like the right thing to do*   (ENG NEWS, 5).

(4.6)   *my writing comes from a place I̲ don't have total control over*
(ENG NEWS, 5).

---

[6] The difference between ENG NEWS and NOR NEWS was not found to be significant.

[7] The translations from Norwegian to English are as literal as possible, and the aim is to keep the word order from the original text. Because of this, the word order might not be idiomatic.

[8] The writer in example 4.4 is not outlining an upcoming discussion in her text. She talks about her upcoming meeting in the future.

[9] The difference between ENG NEWS and NOR NEWS was found to be 99 % certain (p < 0.01).

4.2 *I think* and *jeg tror*

The results in frequency of the recurrent word combination *I think*/ *jeg tror* are presented in this section. One major factor for including this combination is that studies have shown that L2 advanced learners, including Norwegians, tend to overuse *I think* in their writing (see section 2.4). Hence, it is interesting to see, through use of the ICM, whether the NICLE writers are influenced by transfer from their L1 or developmental factors.

In table 11 below, the usage of *I think/jeg tror* is compared between the L1 expert writers:

Table 11: Frequency of *I think/jeg tror* in ENG NEWS and NOR NEWS

	ENG NEWS		NOR NEWS	
	Raw frequency in 100 texts	**Estimated frequency per 10.000 words**	**Raw frequency in 100 texts**	**Estimated frequency per 10.000 words**
I think*/*jeg tror	8	1.00	1	0.11
Total instances of *I*/jeg	329	41.42	170	19.31

It can be seen from the data in table 11 that the L1 expert writers rarely use the recurrent word-combination *I think*/*jeg tror* in their argumentative texts when compared to the total instances of *I*/*jeg*. Calculations of the raw frequency reveal that the English L1 experts use *I think* significantly more often compared to Norwegian L1 experts use of *jeg tror*.[10] Examples 4.5 and 4.6 from ENG NEWS show how the L1 English expert writers use *I think* in their argumentative texts:

> (4.5) *These services are ways we have devised to formalise, to enshrine, our connectedness to one another. I think this is why some Britons still balk at being called "customers" when travelling by train or waiting for a blood test* (ENG NEWS, 89).

---

[10] The difference between ENG NEWS and NOR NEWS was found to be 95 % certain (p < 0.05).

(4.6)   *And every time <u>I think</u> about complaining, I want a minotaur to punch me in the kidneys and remind me how it was before*   (ENG NEWS, 22).

In example 4.5, the writer uses *I think* to express his/her belief on why Britons do not like to be called "customers" by other people. Humble subjectivity is present here (see section 2.2) as *I think* makes the writer's own belief sound less assertive than if *I know* or *I am sure that* were to be used. The reader is invited to either agree or disagree with the writer's own belief, which indicates humbleness in the text.

Example 4.6 illustrates that *I think* is not used to express an opinion or giving a statement, which means that humble subjectivity is not the intended aim. Here, the writer is directly telling the reader that he/she is undergoing the process of thinking about complaining, which means that the prototypical meaning of the word connected to cogitation is present. Example 4.7 below is the one occurrence of *jeg tror* in NOR NEWS.

(4.7)   *Dette er ikke bare til skade for grunnforskningen, ble det påpekt, men like mye for den anvendte forskingen. <u>Jeg tror</u> en liknende beskrivelse ville vært dekkende for mange norske fagfelt.* [Translation: *This is not only damaging for the basic research, it was pointed, but as much for the applied research. I think a similar description would be covering for many Norwegian disciplines*]
(NOR NEWS, 40).

In this one instance of *jeg tror* among L1 Norwegian experts, the writer is, like in example 4.5 from ENG NEWS, expressing an opinion with humble subjectivity. However, this being the only occurrence in a sample of 100 texts indicate that NOR NEWS writers avoid being subjective through use of *jeg tror* in their texts.

4.3 *I feel* and *jeg føler*

The recurrent word-combination *I feel/jeg føler* was chosen as it is commonly known in the English and Norwegian language, and because it is very emotionally loaded. It will be interesting to see whether the L1 and L2 writers are reluctant to using the combination as it is more personal compared to the others.

Table 12 shows the usage of *I feel/jeg føler* in ENG NEWS and NOR NEWS:

Table 12: Frequency of *I feel/jeg føler* in ENG NEWS and NOR NEWS

	ENG NEWS		NOR NEWS	
	Raw frequency in 100 texts	**Estimated frequency per 10.000 words**	**Raw frequency in 100 texts**	**Estimated frequency per 10.000 words**
I feel/jeg føler	**2**	**0.25**	**2**	**0.22**
Total instances of *I*/jeg	**329**	**41.42**	**170**	**19.31**

As shown in table 12, the expert writers avoid the recurrent word combination *I feel/jeg føler* to a large degree.[11] When examining the two instances in ENG NEWS, it becomes apparent that *I feel* is not necessarily used to express humble subjectivity, as shown in examples 4.8 and 4.9 from ENG NEWS:

(4.8) *What <u>I feel</u> about their personalities is irrelevant* (ENG NEWS, 81).

(4.9) *I've never seen those either partly because <u>I feel</u> guilty about not having watched the Bergman films first* (ENG NEWS, 22).

In example 4.8, the writer is using the recurrent word combination *I feel* to tell the reader that his/her subjective feelings are irrelevant. Therefore, the writer's intention is not to express

---

[11] The difference between ENG NEWS and NOR NEWS was not found to be significant.

48

humble subjectivity on a topic. In example 4.9, the writer uses the recurrent word combination to express his/her embarrassment of not having watched the Bergman movies, so humble subjectivity can be ruled out here as well as it is not used to tone down assertiveness.

The two examples 4.10 and 4.11 from NOR NEWS below show that the writers here too are not trying to express humble subjectivity in the usage of *jeg føler*:

> (4.10) *Jeg kaller fortsatt Iran hjem, for uansett hvor lenge jeg må eller ønsker å bo i Frankrike, og selv om <u>jeg føler</u> meg fransk etter alle disse åra, betyr hjem bare én ting: Iran.* [Translation: *I still call Iran home, for regardless of how long I must or wish to live in France, and even though I feel French after all these years, home means only one thing: Iran*]   (NOR NEWS, 16).

> (4.11) <u>*Jeg føler*</u> *dyp uro når prinsessen er blitt okkultismens høye beskytter i Norge, ved ikke bare å bekjenne sin tro på paranormale evner, engler og overnaturlige krefter, men ved aktivt å utbre denne troen med støtte i sin spesielle status.* [Translation: *I feel deep unrest when the princess has been occultisms high protector in Norway, by not only professing her faith on paranormal abilities, angles, and supernatural forces, but by actively spreading this faith with support in her special status*]
> (NOR NEWS, 77).

In example 4.10, the writer is using *jeg føler* to express what nationality he/she feels most connected to. The writer is instance 4.11 describing his/her own feelings to the reader. Like the two instances from ENG NEWS, the NOR NEWS writers are not using the recurrent word combination *jeg føler* to express humble subjectivity in their opinions and statements on topics, but to express how they feel about themselves.

4.4 *I know* and *jeg vet*

The recurrent word-combination *I know/jeg vet* has been chosen as it represents assertiveness to a larger degree than the other *I* framed constructions included in the present study. Hence, it will be interesting to see if there are differences in the way novices and experts express subjective certainty in their texts. However, *I know/jeg vet* is one of many combinations that can be used to express subjective certainty, so no conclusion can be drawn that for instance suggests higher subjective certainty among NICLE writers if they happen to have higher usage of *I know* compared to the other subcorpora.

Table 13 compares the frequency of *I know*/*jeg vet* between ENG NEWS and NOR NEWS:

Table 13: Frequency of *I know/jeg vet* in ENG NEWS and NOR NEWS

	ENG NEWS		NOR NEWS	
	Raw frequency in 100 texts	Estimated frequency per 10.000 words	Raw frequency in 100 texts	Estimated frequency per 10.000 words
I know/jeg vet	4	0.50	1	0.11
Total instances of I/jeg	329	41.42	170	19.31

It becomes apparent from table 13 that *I know/jeg vet* is rarely used among L1 expert writer in argumentative texts. Although the frequency is somewhat higher in the English texts, the difference is not significant.[12] It might be assumed that *I know* is used by the author to express assertiveness in the text, but a closer look at the instances reveal that this is not the case in these particular texts, as can be seen in examples 4.12 and 4.13 from ENG NEWS:

> (4.12) *In other words, she pretends to be the character and then she recites the script.*
> *But what do I know?* (ENG NEWS, 2).

---

[12] The difference between ENG NEWS and NOR NEWS was not found to be significant.

(4.13)   *I know of one health journalist who argued vociferously on Wednesday to stop her editors splashing with "ban this killer vaccine"*   (ENG NEWS, 28).

In example 4.12, the writer not being assertive as he/she admits to not having enough expertise on that topic to draw a conclusion. In example 4.13, the writer is simply telling the reader that he/she knows a health journalist. Hence, *I know* functions as a way of showing familiarity rather than assertiveness. The only example of *jeg vet* from NOR NEWS (4.14) is also not intended as an assertive statement, as the writer argues that he/she does not know, rather than knowing, as shown below:

(4.14)   *Digitaliseringen av litteratur vil gi oss mer tekst og flere forfattere, men fattigere forlag skjønt jeg vet ikke hva motstykket til en innbringende konsert er for en skrivende sjel.* [Translation: *The digitalization of literature will give us more text and more authors, but poorer publishers though I do not know what the counter piece to a grossing concert is for a writing soul*] (NOR NEWS, 75).

4.5 *I would say* and *jeg vil si*

This recurrent word-combination has been investigated as Hasselgård and Johansson (2011) have already studied it and found overuse amongst NICLE writers (see section 2.4.3). However, the authors are very careful in their conclusions as to why overuse is taking place, but they do consider both transfer from Norwegian L1 and developmental factors as possible influencing factors. Hence, the aim here is to contribute with more research on this recurrent word-combination, and investigate the influence of transfer from Norwegian L1 and developmental factors. Hopefully, this will further increase our knowledge of NICLE writers' usage of *I would say*.

Table 14 below compares the frequency *I would say* in ENG NEWS and *jeg vil si* in NOR NEWS:

Table 14: Frequency of *I would say*/*jeg vil si* in ENG NEWS and NOR NEWS

	ENG NEWS		NOR NEWS	
	Raw frequency in 100 texts	**Estimated frequency per 10.000 words**	**Raw frequency in 100 texts**	**Estimated frequency per 10.000 words**
I would* say/jeg *vil si	**0**	**0.00**	**1**	**0.10**
Total instances of *I*/ *jeg*	**329**	**41.42**	**170**	**19.31**

As can be seen from this table, ENG NEWS writers do not use *I would say* at all in their argumentative texts, while one example of *jeg vil si* is found in NOR NEWS.[13] Example 4.15 is the one occurrence found in NOR NEWS:

> (4.15) *Jeg vil si noe om mitt forhold til journalister. Jeg hater journalister.*
> [Translation: *I will say something about my relationship to journalists. I hate journalists*]   (NOR NEWS 24).

---

[13] The difference between ENG NEWS and NOR NEWS was not found to be significant.

In this example, the writer is being very honest about his/her thoughts on journalists. As the first sentence in this example is the very first sentence of that particular argumentative text, it seems like *jeg vil si* is being used to introduce her upcoming opinions and statements. Since this is the only example found in NOR NEWS, no conclusions can be drawn as to whether this is common practice or not.

4.6 Discussion of the findings from the expert texts in relation to previous research

This chapter has investigated and discussed differences in usage of *I* as subject in mental and verbal processes, and four different recurrent word-combinations between English and Norwegian L1 expert writers. ENG NEWS writers' usage of *I* as subject in mental processes is significantly higher than NOR NEWS writers' usage of *jeg*. In verbal processes, the difference was not found to be statistically significant, but NORNEWS contained more occurrences than ENG NEWS. In previous studies, Fossan (2011) found that NICLE writers use *I* as subject 90.41 times per 10.000 words, while Paquot et al. (2013) found a frequency of 76.7 per 10.000 words among NICLE writers. Surprisingly, the ENG NEWS frequency of *I* as subject (54.61 per 10.000 words) is closer to NICLE than what was expected. NOR NEWS has a frequency of 23.44 per 10.000 words in the use of *jeg* as subject, which is far below NICLE in *I* usage. Thus, the experts have like expected, a lower frequency of *I* as subject compared to previous research on NICLE, but surprisingly, ENG NEWS have a significantly higher usage compared to NOR NEWS (see section 4.1).

The recurrent word combination *I think* appears significantly more often in ENG NEWS compared to *jeg tror* in NOR NEWS. This was expected as Norwegian L1 experts use *jeg* less frequently in their texts than English L1 experts use of *I*. However, the ENG NEWS writers use *I think* less frequently than the LOCNESS writers and all the L2 learner groups from Ringbom's (1998) study. The frequency of the recurrent word-combinations *I feel*/*jeg føler*, *I know*/*jeg vet*, and *I would say*/*jeg vil si* was found to be very low in both ENG NEWS and NOR NEWS. This was somewhat expected as Ädel and Erman (2012) found that non-native speakers overuse recurrent word-combinations with *I* as subject compared to native speakers. It would be surprising if native speakers had a high frequency of recurrent word-combinations with *I* as subject to begin with based on the findings from that study.

4.6.1 Implications for the investigation of the novice L1 writers

The same areas that were investigated in sections 4.1-4.5 will be explored in chapter 5, but this time in English and Norwegian novice texts. It is expected that the Norwegian L1 novice writers will have a higher usage of *I* as subject in mental and verbal processes compared to the Norwegian L1 experts. This hypothesis is based on the fact that expert writers have much more experience in writing compared to novices, as well as more knowledge on the topics they discuss. Where the experts have referred to other situations and individuals to prove a point, it is expected that the Norwegian L1 novices will refer to their own opinions through the use of *I* to a much larger degree. This thought is based on the fact that NOR ESSAYS writers are at high-school level. The LOCNESS writers, on the other hand, have already been studied by several researchers, including Fossan (2011) who found that their frequency of *I* as subject is 30.15 per 10.000 words. This number might of course vary, as other texts from the same corpus are investigated, but they are not expected to have a higher frequency of *I* in clauses with mental and verbal processes compared to English L1 expert writers in the present study. The LOCNESS writers are at university level, which means that they are expected to be closer to the expert L1 norm than the NOR ESSAYS writers. As for the four recurrent word-combinations, it is difficult to predict whether L1 novices will show higher frequencies than L1 experts. This is because studies to date, like the one from Ädel and Erman (2012) have focused on differences between L1 and L2 writers.

5. English and Norwegian L1 novice texts

In this chapter, the L1 novice texts are investigated through the CA approach. Thus, LOCNESS will be compared to NOR ESSAYS in usage of *I* as subject in clauses with mental and verbal processes, and in the usage of the four recurrent word-combinations, namely: *I think*/*jeg tror*, *I feel*/*jeg føler*, *I know*/*jeg vet*, and *I would say*/*jeg vil si*.

5.1 Clauses with mental and verbal processes

In the following CA, the two novice-writer L1 subcorpora have been compared in the usage of *I*/*jeg* as subject in mental, verbal, and other processes. The results are presented in table 15 below:

Table 15: Frequency of *I*/*jeg* as subject in LOCNESS and NOR ESSAYS

	LOCNESS		NOR ESSAYS	
	Raw frequency in 30 texts	**Estimated frequency per 10.000 words**	**Raw frequency in 30 texts**	**Estimated frequency per 10.000 words**
***I*/*jeg* in mental processes**	67	20.07	83	29.28
***I*/*jeg* in verbal processes**	5	1.49	17	5.99
***I*/*jeg* in other processes**	21	6.29	30	10.58
Total instances of *I*/ *jeg*	93	27.86	130	45.85

It is apparent from table 15 that Norwegian L1 novice writers' usage of *jeg* as subject is higher than the English L1 novice writers' usage of *I*.[14] The difference in usage of clauses with *I* as

---

[14] The difference between LOCNESS and NORESSAYS was found to be 99.9 % certain (p < 0,001).

subject in mental processes was statistically significant as well. [15] In verbal processes, the difference was also found to be statistically significant.[16] Lastly, in other processes, Norwegian novice writers seem to have a higher usage of *jeg* as subject compared to LOCNESS, however, the difference was not found to be statistically significant. These results will be discussed in detail in section 5.6.

5.2 *I think* and *jeg tror*

In table 16 below, the usage of *I think*/*jeg tror* is compared between the L1 novice writers:

Table 16: Frequency of *I think*/*jeg tror* in LOCNESS and NOR ESSAYS

	LOCNESS		NOR ESSAYS	
	Raw frequency in 100 texts	Estimated frequency per 10.000 words	Raw frequency in 100 texts	Estimated frequency per 10.000 words
I think/*jeg tror*	25	2.86	46	4.48
Total instances of *I*/*jeg*	399	45.60	563	54.89

As can be seen from table 16, NOR ESSAYS has a higher frequency of the recurrent word-combination *jeg tror* compared to LOCNESS in the usage of *I think*, but the difference is not significant.[17]

Both LOCNESS and NOR ESSAYS have a majority of instances where the recurrent word combination *I think* and *jeg tror* is used to express humble subjectivity when giving opinions or statements. This can be seen from the instances below:

[15] The difference between LOCNESS and NORESSAYS was found to be 95 % certain (p < 0.05).

[16] The difference between LOCNESS and NORESSAYS was found to be 99 % certain (p < 0.01).

[17] The difference between LOCNESS and NOR ESSAYS was not found to be significant.

(5.1)   *An invention of the 20th century that <u>I think </u>has significantly changed people's lives is television*   (LOCNESS, 68).

(5.2)   <u>*Jeg tror*</u> *at OL er en viktig tradisjon som vedlikeholder interessen for sport.* [Translation: *I think that the Olympic Games is an important tradition that maintains the interest for sports*]   (NOR ESSAYS, 23).

In example 5.1, the writer is using *I think* to express his/her subjective opinion on the invention of the television. The writer's use of *I think* tells the reader that his/her opinion is not necessarily the objective truth, and people are free to agree or disagree. In example 5.2, the writer is expressing his/her belief on the important tradition the Olympic Games bring to sport, and like the first example, the reader is indirectly told through *jeg tror* that this is the writer's own belief, and other individuals are free to believe something else on the tradition of the Olympic Games. The excessive usage found among L1 novice writers show that they express humble subjectivity in their texts to a larger degree than L1 expert writers. These results challenge Granger (1998), who argues that humble subjectivity is a continental European phenomenon, as the texts from LOCNESS are from American writers.

5.3 *I feel* and *jeg føler*

Table 17 below shows the usage of *I feel*/*jeg føler* in LOCNESS and NOR ESSAYS:

Table 17: Frequency of *I feel*/*jeg føler* in LOCNESS and NOR ESSAYS

	LOCNESS		NOR ESSAYS	
	Raw frequency in 100 texts	Estimated frequency per 10.000 words	Raw frequency in 100 texts	Estimated frequency per 10.000 words
I feel/*jeg føler*	25	2.85	1	0.09
Total instances of *I*/*jeg*	399	45.60	563	54.89

Interestingly, the LOCNESS writers use the recurrent word combination *I feel* significantly more often than NOR ESSAYS writers[18], who do not use *I føler* at all in their texts, with one exception. This might indicate that *jeg føler* is too emotionally loaded for Norwegian writers, and hence, they rather use recurrent word-combinations they are more comfortable with, like *jeg tror* where they have a higher frequency compared to LOCNESS writers (see section 5.2).

Examples 5.3 and 5.4 below illustrate that unlike in ENG NEWS, where the two instances of *I feel* are used to either distance personal feelings from the text or to express embarrassment (see section 4.3), LOCNESS contains instances where *I feel* is used to express humble subjectivity when giving opinions and statements on topics:

> (5.3)   *I feel that there are both values and consequences to the integration of schools and if the program is going to be successful, it needs support from venues other than the school systems themselves*   (LOCNESS, 8).

---

[18] The difference between LOCNESS and NOR ESSAYS was found to be 99.99 % certain (p < 0.0001).

(5.4)   *An invention of the 20th century which <u>I feel</u> has significantly changed people's lives is the introduction of Bank-cash machines or Automatic teller machines* (LOCNESS, 64).

The writer is in example 5.3 giving her opinion on how the integration of schools is going to be successful, and *I feel* in this context gives room for other people's opinions. If for instance *I know,* or *I am sure* was used instead in this context, the writer would seem very assertive in her claims, but by using the concurrent word combination *I feel*, she expresses humble subjectivity by sharing her views in a respectful matter, which again shows that she is open to other people's suggestions. Example 5.4 is perhaps more of a statement than an opinion, at least when compared to the first example. This might have to do with the fact that the writer uses the word "significantly" to put weight on the given opinion on 20th-century inventions. Still, the opinion/statement is expressed with humble subjectivity, as the writer invites the reader to have a different opinion, by saying *I feel*.

5.4 *I know* and *jeg vet*

In table 18, the difference in frequency of *I know*/*jeg vet* between LOCNESS and NOR ESSAYS is presented:

Table 18: Frequency of *I know*/jeg vet in LOCNESS and NOR ESSAYS

	LOCNESS		NOR ESSAYS	
	Raw frequency in 100 texts	**Estimated frequency per 10.000 words**	**Raw frequency in 100 texts**	**Estimated frequency per 10.000 words**
I know*/*jeg vet	**14**	**1.60**	**7**	**0.68**
Total instances of *I*/ *jeg*	**399**	**45.60**	**563**	**54.89**

It can be seen from table 18 that there is a distinct difference in frequency between the L1 novice subcorpora, as there are twice as many instances of *I know* in LOCNESS than *jeg vet* from NOR ESSAYS. However, since the number of occurrences is quite limited in both subcorpora, the difference between LOCNESS and NOR ESSAYS was not found to be statistically significant.[19]

Interestingly, both LOCNESS and NOR ESSAYS contain instances where the writer is using the recurrent word combination *I know*/*jeg vet* to express assertiveness. Example 5.5 and 5.6 from LOCNESS illustrate this:

> (5.5)   *Too much government relief, over the years, has taken away people's pride in earning a living and being self-sufficient, and it has encouraged them to quit their low paying jobs and rely on the Welfare System for food rent, and medical assistance - all at the expense of hard-working taxpayers. I know this from my own personal experience*   (LOCNESS, 22).

---

[19] The difference between LOCNESS and NOR ESSAYS was not found to be significant.

> (5.6)   *I know the rules are very important to follow because I am living proof. I have been working at an amusement park called Frankie's Fun Park for about a year and a half and I have seen a good bit of people get hurt, because they didn't follow the rules*   (LOCNESS, 99).

Looking at example 5.5, the writer clearly states how the government has given too many advantages to unemployed people at the expense of the taxpayers. Then, he uses *I know* to defend that statement as he is, according to himself, experienced with how the welfare system works, and therefore possesses the knowledge to say what is 'right'. In the second example 5.6, the writer uses the same tactic to justify the confidence in the statement. Because she has experience working as an employee in an amusement park, she possesses the necessary expertise needed to tell the reader how following the rules is an important aspect in go-kart racing, which was the case she argued for.

In NOR ESSAYS, assertiveness through use of *jeg vet* is visible, as in the following example:

> (5.7)   *Det er menneskene generelt som er for griske, eller jeg burde vel si egoistiske. Jeg vet ialefall at Norge sin befolkning er på "syte toppen" i verden, til tross for at Norge er kåret av FN til verdens beste land og bo i!* [Translation: *It is people generally that are too greedy, or I should say selfish. I know in any case that Norway's population is on the "whimper top" in the world, although Norway is named by UN to the world's best country to live in!*] (NOR ESSAYS, 94).

The writer is in example 5.7 very outright in his views, arguing that people in Norway complain a lot without backing it up with any statistics or sources. It needs to be addressed here that while he has every right to state this, the immaturity of this piece of writing is quite revealing, as there are a lot of grammar mistakes, including the exclamation mark, which gives the idea that his strong emotions about human behavior perhaps trumps rationality in this case. Here, the age difference between the novice writers might play a role, as we know that the NOR ESSAYS writers are at high-school level, while LOCNESS and NICLE writers are at university level (see section 3.3.3.2).

5.5 *I would say* and *jeg vil si*

The frequency of the recurrent word-combination *I would say* from LOCNESS and *jeg vil si* from NOR ESSAYS is presented in table 19 below:

Table 19: Frequency of *I would say*/*jeg vil si* in LOCNESS and NOR ESSAYS

	LOCNESS		NOR ESSAYS	
	Raw frequency in 100 texts	**Estimated frequency per 10.000 words**	**Raw frequency in 100 texts**	**Estimated frequency per 10.000 words**
I would say*/*jeg vil si	**0**	**0.00**	**1**	**0.10**
Total instances of *I*/ *jeg*	**399**	**45.60**	**563**	**54.89**

The 100 texts from LOCNESS do not contain any instances of *I would say*, while one instance of *jeg vil si* is found in the 100 texts from NOR ESSASYS. [20] Like the L1 expert writers (see section 4.5), the novices steer away from the use of *I would say*/*jeg vil si*. Example 5.8 is the only one found in NOR ESSAYS:

(5.8)   *Jeg vil ikke ta fra han det at han er en mann med en utrolig sans for markedsføring, og et uvanlig godt hode, men jeg vil si at han ikke er alene om det*. [Translation: *I will not take from him that he is a man with an amazing sense for marketing, and an unusually good head, but I will say that he is not alone of that*]   (NOR ESSAYS, 72).

In example 5.8, *jeg vil si* is used by the writer to express an opinion. If *jeg tror* replaced *jeg vil si* in this sentence, the meaning would remain the same. This is relevant as Hasselgård

---

[20] The difference between LOCNESS and NOR ESSAYS was not found to be significant.

and Johansson (2011, p. 50) found that *I think* could sometimes replace *I would say* in English. Thus, *jeg tror* can in some cases replace *jeg vil si* in Norwegian.

5.6 Discussion of the findings from the novice texts in relation to the expert texts and previous research

The results obtained from the CA of clauses with *I/jeg* as subject in mental, verbal, and other processes (see section 4.1 and 5.1) are summarized in table 20:

Table 20: Frequency of *I/jeg* as subject in mental, verbal, and other processes in the four L1 subcorpora

Frequency per 10.000 words				
	ENG NEWS	**NOR NEWS**	**LOCNESS**	**NOR ESSAYS**
Mental:	26.88	9.83	20.07	29.28
Verbal:	2.13	3.78	1.49	5.99
Other:	25.60	9.83	6.29	10.58
Total:	54.61	23.44	27.86	45.85

The findings from the CA approach reveal an interesting pattern. Both expert L1 subcorpora uses approximately the same amount of *I* as subject in mental and other processes in their texts. ENG NEWS writers have a frequency of 26.88 per 10.000 words in mental processes, and 25.60 per 10.000 words in other processes. NOR NEWS writers has a frequency of 9.83 in mental and other processes. Both novice L1 subcorpora use mental processes approximately three times as often as other processes. In LOCNESS, the frequency in mental processes is 20.07 per 10.000 words, and 6.29 in other processes. NOR ESSAYS has a frequency of 29.28 per 10.000 words in mental processes, and 10.58 per 10.000 words in other processes. These findings suggest that the expert writers use more variety in the use of *I* as subject compared to the L1 novice writers. This is most likely due to expert writers being more experienced in argumentative writing.

Interestingly, the two Norwegian L1 subcorpora have the highest frequency of *jeg* as subject in clauses with verbal processes. This is also the only category where the Norwegian experts surpass the English experts in frequency, although the difference is minor and not statistically significant. However, when looking at the difference between the L1 novices, Norwegian writers clearly prefer to use *jeg* as subject in verbal processes more than English writers. The fact that both Norwegian L1 experts and novices have a higher frequency of *I* as subject in clauses with verbal processes, might indicate that there are cultural differences between the two languages, and that there may be a transfer effect.

Another interesting observation from table 20 is that English expert writers seem to use *I* as subject more often than English novice writers.[21] This outcome is contrary to that of Hasselgård (2009) who found that *I* as subject did not occur a single time in the ICE-GB press editorials, and hence suggested that Norwegian advanced learners use features from conversational English. That might still be true, but the results from the CA in the present study have shown that L1 novices and experts both use *I* as subject in argumentative texts, meaning that there are certainly other influencing factors playing a role in NICLE writers' usage. Whether transfer from Norwegian L1 or developmental factors have an influence on NICLE writers' usage will be investigated after the results from NICLE have been presented in chapter 6.

---

[21] The difference between ENG NEWS and LOCNESS was found to be 99,99 % certain (p < 0,0001).

The results from the CA of the four recurrent word-combinations are presented in table 21 below:

Table 21: Frequency of the four recurrent word-combinations in the four L1 subcorpora

Frequency per 10.000 words				
	ENG NEWS	**NOR NEWS**	**LOCNESS**	**NOR ESSAYS**
I think/jeg tror	1.00	0.11	2.86	4.48
I feel/jeg føler	0.25	0.22	2.85	0.09
I know/jeg vet	0.50	0.11	1.60	0.68
I would say/jeg vil si	0.00	0.10	0.00	0.10
Total instances of *I/jeg*	41.42	19.31	45.60	54.89

As can be seen from table 21, novice L1 writers use the recurrent word-combination *I think* more often than expert writers. Apparently, L1 novice writers are either more humble in subjectivity, or they lack speaker authority. It is most likely a mixture of both. Granger (1998) argued that readers from continental Europe would operate with humble subjectivity due to cultural differences, but this conclusion is challenged by these findings, as LOCNESS text contains a lot of examples of humble subjectivity through use of *I think,* and *I feel*. A possible explanation for the high usage of *I think,* and *I feel* among LOCNESS writers, and *jeg tror* among NOR ESSAYS writers, could be that the writers are younger and less knowledgeable on the topics they discuss compared to the experts, and hence, they tend to be more humble in subjectivity and less authoritative in their texts. However, in the usage *I know/jeg vet*, the novices also have a higher frequency compared to experts, and examples from section 5.4 illustrated how the combination is sometimes used to express certainty. Thus, while novices tend to be more humble in subjectivity compared to experts, they also express certainty

through use of *I know/jeg vet* to a larger degree compared to the experts. A possible explanation might again be connected to differences in age and knowledge on the topics discussed. While an expert might link their opinions to similar cases, the novice writers more easily refer to themselves and their own experiences. Both L1 novices and experts steer away from the recurrent word-combination *I would say/jeg vil si*.

5.6.1 Implications for the investigation of the Norwegian advanced learners

Previous studies have shown that NICLE writers overuse *I* as subject compared to native speakers (Hasselgård, 2009; Fossan, 2011; Paquot et al., 2013). The general hypothesis underlying the present study is that among NICLE writers, overuse of *I* as subject in mental and verbal processes will be found due to transfer from Norwegian L1 and developmental factors. Because of the many differences between the L1 subcorpora, it is difficult to predict in what areas the NICLE writers might resemble the norm of expert English, and in what areas they might be influenced by transfer from their L1 or developmental factors.

Ädel and Erman (2012) suggest that *I* framed constructions are being overused by non-native speakers of English. The underlying hypothesis of this present study is that overuse of the four recurrent word-combinations among NICLE writers will be found due to transfer from Norwegian L1 and developmental factors. Based on the findings from the corpus analyses in presented in this chapter, it is difficult to predict whether NICLE will resemble ENG NEWS in the usage of the recurrent-word combinations, or to predict in what combinations transfer from Norwegian L1 and developmental factors play a role. However, since the L1 novice writers have a higher frequency compared to the L1 experts in most cases (see table 21), it could be assumed that the NICLE writers will follow their pattern, since they are novice writers as well.

6.   Norwegian advanced learners' L2 texts

In this chapter, the novice L2 English texts from NICLE will be compared to texts from the English L1 experts. The aim here is to check whether the NICLE writers resemble the experts or show overuse of *I* as subject in mental and verbal processes, and the four recurrent word-combinations. Discussion of the results in relation to previous studies will take place in chapter 7 where factors that might influence the NICLE writers are investigated.

6.1 Clauses with mental and verbal processes

The following contrastive analysis will show the differences in usage of *I* as subject in mental and verbal processes between NICLE and ENG NEWS writers. Table 22 compares the frequency of *I* as subject in mental, verbal, and other processes:

Table 22: Frequency of *I* as subject in NICLE and ENG NEWS

Function	NICLE		ENG NEWS	
	Raw frequency in 30 texts	**Estimated frequency per 10.000 words**	**Raw frequency in 30 texts**	**Estimated frequency per 10.000 words**
I in mental processes	129	65.02	63	26.88
I in verbal processes	14	7.05	5	2.13
I in other processes	34	17.13	60	25.60
Total instances of *I*	177	89.22	128	54.61

Calculations of the raw frequency data in table 22 reveal that Norwegian advanced learners of English overuse *I* as subject compared to English expert writers.[22] In clauses with mental processes, NICLE have significant overuse compared to ENG NEWS.[23] In clauses with verbal

---

[22] The difference between NICLE and ENG NEWS was found to be 99.99 % certain ($p < 0.0001$).

[23] The difference between NICLE and ENG NEWS was found to be 99.99 % certain ($p < 0.0001$).

processes, NICLE also have significantly higher overuse than ENG NEWS.[24] Lastly, in clauses with other processes, ENG NEWS seems to have overuse compared to NICLE, however, the difference is not statistically significant.[25]

6.2 *I think*

Table 23 below compares the usage of *I think* between NICLE and ENG NEWS:

Table 23: Frequency of *I think* in NICLE and ENG NEWS

	NICLE		ENG NEWS	
	Raw frequency in 100 texts	**Estimated frequency per 10.000 words**	**Raw frequency in 100 texts**	**Estimated frequency per 10.000 words**
I think	72	10.69	8	0.88
Total instances of *I*	568	84.35	329	41.42

It is apparent from this table that NICLE writers overuse the recurrent word-combination *I think* significantly when compared to ENG NEWS writers.[26] The estimated frequency per 10.000 words further reveals that NICLE writers use *I think* approximately ten times as often as ENG NEWS writers, which can be described as excessive overuse. NICLE writers share characteristics with the L1 novice writers in that they also use *I think* to express humble subjectivity in their texts. Examples 6.1 and 6.2 illustrate this:

---

[24] The difference between NICLE and ENG NEWS was found to be 95 % certain (p < 0.05).

[25] The difference between NICLE and ENG NEWS was not found to be significant.

[26] The difference between NICLE and ENG NEWS was found to be 99.99 % certain (p < 0.0001).

(6.1) *I think the reason for the success of the television in the modern society is it's ability to both entertain and inform* (NICLE, 29).

(6.2) *In Norway, I think that we have better prisons then for examples the US* (NICLE, 21).

In example 6.1, the writer is communicating his/her opinion on the success of television in the modern era to the reader. By using *I think*, the writer seems to be open-minded, and open to other people's thoughts on why the television industry has become so successful. The writer is in example 6.2 expressing his thoughts on how he thinks Norway has a better prison system compared to the US. The use of *I think* tones down his assertiveness around the given opinion, as the reader is informed that this is the writer's own subjective opinion on the Norwegian prison system.

6.3 *I feel*

Table 24 compares the frequency of *I feel in* NICLE and ENG NEWS:

Table 24: Frequency of *I feel* in NICLE and ENG NEWS

	NICLE		ENG NEWS	
	Raw frequency in 100 texts	**Estimated frequency per 10.000 words**	**Raw frequency in 100 texts**	**Estimated frequency per 10.000 words**
I feel	13	1.90	2	0.25
Total instances of *I*	568	84.35	329	41.42

Table 24 shows that NICLE writers overuse the recurrent word combination *I feel* significantly compared to ENG NEWS writers.[27] Example 6.3 and 6.4 below show that NICLE writers resemble LOCNESS writers (see section 5.3) in how they use *I feel* in their texts:

> (6.3)   *If schools were better at supporting and making it easier for students to use their practical skills in everyday school I feel everyone would have a more equal chance of being successful and no one is labelled an outcast*   (NICLE, 9).

> (6.4)   *Now when that is said I feel that it is very useful to have these practice periods. Even though it is for a very short period of our education, a student can learn a lot about life in a school and team work*   (NICLE, 99).

Like the LOCNESS writers, the Norwegian advanced learners of English tend to express humble subjectivity in their texts when using the recurrent word-combination *I feel*. In example 6.3, the writer is careful not to be too assertive, and whether this is the writer's intention is hard to say, but it does tell the reader that his opinion is not necessarily correct,

---

[27] The difference between NICLE and ENG NEWS was found to be 99.9 % certain (p< 0.001).

and the reader is invited to agree or disagree. In example 6.4, the writer starts by expressing her humble opinion on the usefulness of practice periods in teacher education. She then elaborates on that idea in the next sentence, which might be seen as a statement. However, the reader has already been informed that this is her own feelings towards the topic of teacher education, and hence, her opinion and statement is based on humble subjectivity.

6.4 *I know*

We now turn to the comparison of *I know* between NICLE and ENG NEWS writers. Table 25 shows the difference in frequency:

Table 25: Frequency of *I know* in ENG NEWS and NICLE

	NICLE		ENG NEWS	
	Raw frequency in 100 texts	**Estimated frequency per 10.000 words**	**Raw frequency in 100 texts**	**Estimated frequency per 10.000 words**
I know	16	2.37	4	0.50
Total instances *I*	568	84.35	329	41.42

It can be seen from table 25 that the NICLE writers have a higher frequency of the recurrent word combination *I know* compared to ENG NEWS writers, and the difference is significant.[28] The NICLE writers use the recurrent word combination often to assure the reader that he/she understands the core, or the moral of the subject, as can be seen from example 6.5 below:

---

[28] The difference between ENG NEWS and NICLE was found to be 99 % certain (p < 0.01).

(6.5) *The lives of these mobsters are portrayed in such a way that it's almost impossible to think that their actions are wrong. I know that they are criminals but still their way of life with easy money, fast cars and gorgeous women sort of takes the focus off the fact that they commit serious crimes* (NICLE, 4).

The writer is, in this case, arguing that the way criminals are portrayed in specific movies tones down the fact that they are criminals, but at the same time, he assures the reader that he is aware that what they are doing is wrong.

The NICLE resembles the L1 novice subcorpora (see section 5.4) in that assertiveness is found in the usage of *I know*, as shown below in example 6.6:

(6.6) *It is hard to state what needs to be done with the situation, but I know for a fact that if something is not done it will become harder to recruit people to the practical college degrees in the future* (NICLE, 95).

6.5 *I would say*

In table 26, NICLE is compared to ENG NEWS in the usage of the recurrent word-combination *I would say*:

Table 26: Frequency of *I would say* in NICLE and ENG NEWS

	NICLE		ENG NEWS	
	Raw frequency in 100 texts	**Estimated frequency per 10.000 words**	**Raw frequency in 100 texts**	**Estimated frequency per 10.000 words**
I would say	13	1.93	0	0.00
Total instances of *I*	568	84.35	329	41.42

Table 26 shows that NICLE writers overuse *I would say* compared to ENG NEWS writers. [29]

Hasselgård and Johansson (2011, p. 50) pointed out how NICLE writers use *I would say* in conclusions, and example 6.7 and 6.8 taken from NICLE provide support to their claim.

> (6.7) *As a conclusion <u>I would say</u> that crime does not pay, although it may seem so sometimes* (NICLE, 4).

> (6.8) *To summarise <u>I would say</u> that I agree with what the topic says to a certain degree* (NICLE, 86).

Hasselgård and Johansson (2011, p. 52) argued that low assertiveness in conclusions through the use of *I would say* might have to do with lack of speaker authority. In example 6.7, the writer is perhaps exercising a lower degree of assertiveness than what is desirable in a conclusion of an argumentative text. The writer first claims that crime does not pay, but then

---

[29] The difference between NICLE and ENG NEWS was found to be 99 % certain (p< 0.01).

says that it sometimes seems so without elaborating. It would perhaps be better to focus exclusively on why crime does not pay in the conclusion, so the reader would better understand the writers' position on the topic of crime. In example 6.8, the writer is also not very assertive in his/her conclusion, but an explanation is given to the reader as to why he/she cannot come to a clear conclusion. This makes it easier to understand the writer's position, and hence, the conclusion cannot be seen as lacking speaker authority.

In example 6.9 and 6.10, the writer is using *I would say* to express an opinion:

> (6.9) *I would say that the increase of science technology and industrialisation are based on both knowledge and imagination* (NICLE, 73).

> (6.10) *When it comes to two different aspects of punishment, life- and death sentences, I would say that if we sentence a person to death, he or she would feel the physical pain, - what about the mental pain?* (NICLE, 79).

In these examples, *I would say* could have been replaced by other recurrent word-combinations, such as *I think*, as both combinations function as tentative when used to give an opinion.

7. Factors influencing the Norwegian advanced learners

In this chapter, we will investigate whether transfer from Norwegian L1 and developmental factors influence the NICLE writers in their usage of *I* as subject in mental and verbal processes, and in the four recurrent word-combinations. Thus far, NICLE has only been compared to ENG NEWS, so it is necessary at this point to investigate the similarities and differences between NICLE and NOR NEWS as well as LOCNESS and NOR ESSAYS. The hypothesis for this study (see section 1.1) is that overuse among NICLE writers is found due to either (a) transfer from Norwegian, (b) developmental factors, (c) other factors. A mixture of the three factors is also a possibility. Other factors refer to aspects such as writing in a foreign language or features of spoken language. To develop a full picture of NICLE writers overuse, additional studies will be needed that focuses on this particular category.

7.1 Clauses with mental and verbal processes

Figure 2 compares all five subcorpora in the use of *I*/jeg in clauses with mental processes and illustrates the difference in frequency:

Figure 2: Frequency of *I/jeg* in clauses with mental processes compared in the five subcorpora
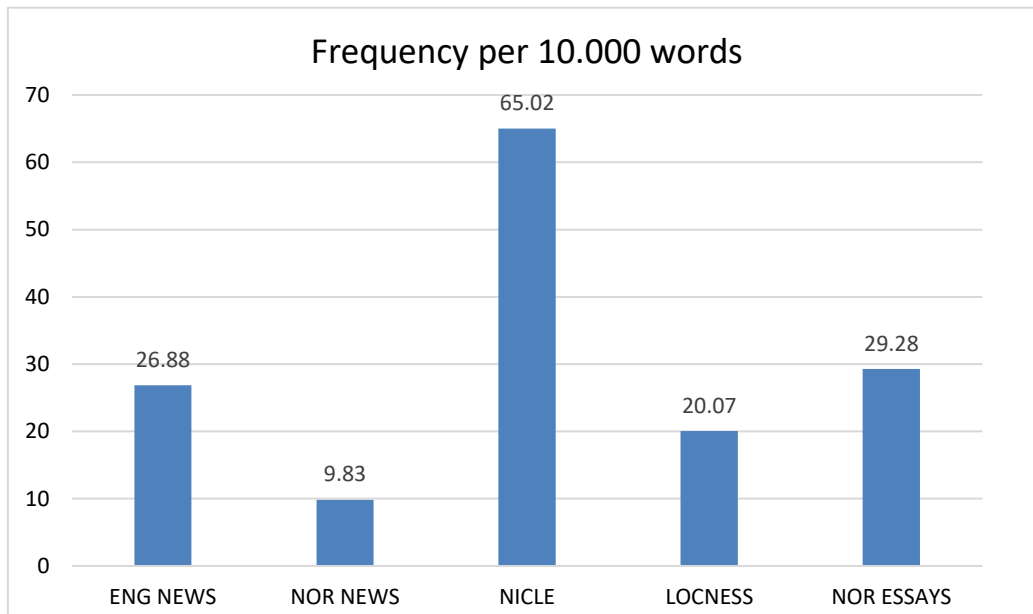
This figure shows that the NICLE writers are not influenced by transfer from Norwegian, as NOR NEWS has by far the lowest frequency, while NICLE has by far the highest frequency.[30] Developmental factors cannot explain the overuse either, as both LOCNESS and NOR ESSAYS differ significantly when compared to NICLE.[31] [32] It is plausible to suggest that the excessive overuse of *I* in mental processes is due to the challenges of writing in a foreign language as well as other factors, like influence from spoken language. Thus, these findings contradict the hypothesis that overuse among NICLE writers can be explained by transfer from Norwegian and developmental factors.

---

[30] The difference between NICLE and NOR NEWS was found to be 99.99 % certain (p< 0.0001).

[31] The difference between NICLE and LOCNESS was found to be 99.99 % certain (p< 0.0001).

[32] The difference between NICLE and NOR ESSAYS was found to be 99.99 % certain (p< 0.0001).

In figure 3, the five subcorpora have been compared in the frequency of *I* as subject in clauses with verbal processes:

Figure 3: Frequency of *I/jeg* in clauses with verbal processes compared in the five subcorpora
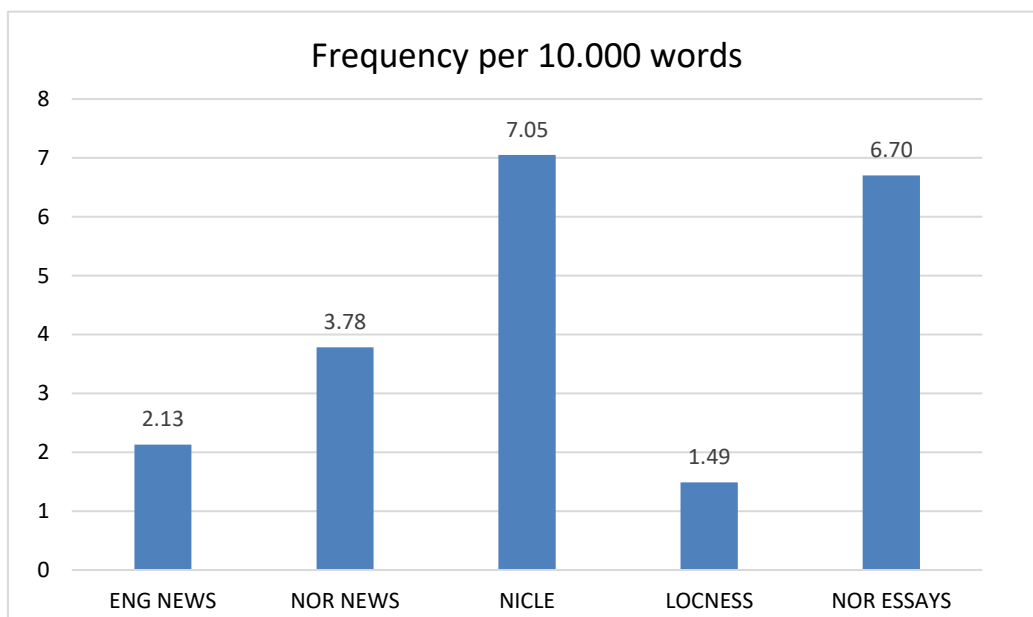
The first thing to notice in this figure is that the three subcorpora comprising texts written by individuals with Norwegian as their L1 have the highest frequency. Traces of transfer can be detected, as both NOR NEWS and NICLE resemble each other to some extent, and the difference in frequency is not significant.[33] Developmental factors seem to play a bigger role, as NICLE is very similar to NOR ESSAYS in frequency [34], indicating that NICLE writers pass on their Norwegian L1 novice competence over to L2 English.

---

[33] The difference between NICLE and NOR NEWS was not found to be significant.

[34] The difference between NICLE and NOR ESSAYS was not found to be significant.

The results found in NICLE corroborate the findings of a great deal of previous research that suggested that NICLE writers overuse *I* as subject compared to native English speakers (Fossan, 2011; Hasselgård, 2009, 2012; Paquot et al. 2013). Although these studies did not focus specifically on clauses with *I* as subject in mental and verbal processes, Hasselgård (2009) found that in NICLE, the majority of self-reference clauses were found with mental and verbal processes.

The frequency of *I* as subject found in the present study and Fossan (2011) are quite similar. We know that Fossan used 100 NICLE texts to calculate the frequency per 10.000 words, while the present study used 30 NICLE texts. The fact that both these studies ended up with a frequency close to 90 per 10.000 words in the usage of *I* as subject heavily indicates that the accurate frequency is around that level. However, the frequency found in Paquot et al. (2013) is lower than both Fossan (2011) and the present study, challenging the claim that 90 per 10.000 is an accurate estimation of actual frequency in NICLE. Paquot et al. (2013) used 189 NICLE texts for comparison, which is approximately twice the amount used in Fossan, and six times the amount used in this project.

In LOCNESS, the findings in frequency differ in all the studies, although not to a large degree. Fossan (2011), who used 100 texts for comparison reports the highest frequency, followed by the present study who used 30 texts. The lowest frequency was again found in Paquot et al. (2013), but it is important to note that in LOCNESS, they used 88 texts for comparison. While it is interesting to find out the exact frequency of *I* in NICLE and LOCNESS, the essential fact to note here is that Norwegian advanced learners do indeed overuse *I* as subject compared to L1 native speakers. It was shown in section 6.1 that the difference between ENG NEWS and NICLE was significant. This is also the case when NICLE is compared to LOCNESS. [35]

The results from clauses with mental processes contradict Fossan (2011) who suggested that transfer from Norwegian might influence Norwegian L2 learners in their usage of 1st person pronouns. It was shown in section 7.1 that the NICLE had a significantly higher frequency of *I* compared to NOR NEWS and NOR ESSAYS, which suggests that transfer is not an influencing factor. In clauses with verbal processes, however, the results are correlating well with Fossan's (2011) suggestion of transfer influence, as the difference in frequency between NICLE, NOR NEWS and NOR ESSAYS is not significant.

---

[35] The difference between NICLE and LOCNESS was found to be 99.99 % certain.

7.2 The recurrent word-combinations

This section will analyze the factors that influence the NICLE writers' use of the four different recurrent word-combinations; *I think*, *I feel*, *I know*, and *I would say*. Before we start by looking at each combination, it is useful to introduce a table that sums up the findings from the five subcorpora:

Table 27: Frequency of the recurrent word-combinations *I think*, *I feel*, *I know*, and *I would say*

Subcorpora:	ENG NEWS		NOR NEWS		NICLE		LOCNESS		NOR ESSAYS	
Size in words:	**79388**		**88010**		**67332**		**87430**		**102499**	
Frequency	**Total instances**	**Per 10.000 words**	**Total instances**	**Per 10.000 words**	**Total instances**	**Per 10.000 words**	**Total instances**	**Per 10.000 words**	**Total instances**	**Per 10.000 words**
I think/jeg tror	8	1.00	1	0.11	72	10.69	25	2.85	46	4.48
I feel/jeg føler	2	0.25	2	0.22	13	1.90	25	2.85	1	0.09
I know/jeg vet	4	0.50	1	0.11	16	2.37	14	1.60	7	0.68
I would say/jeg vil si	0	0.00	1	0.10	13	1.93	0	0.00	1	0.10
Total instances of *I/jeg*	329	41.42	170	19.31	568	84.35	399	45.60	563	54.89

7.2.1 *I think* and *jeg tror*

Figure 4 below summarizes the findings from all the five subcorpora and illustrates the excessive overuse of *I think* from NICLE writers:

Figure 4: Frequency of *I think*/jeg tror per 10.000 words in the five subcorpora
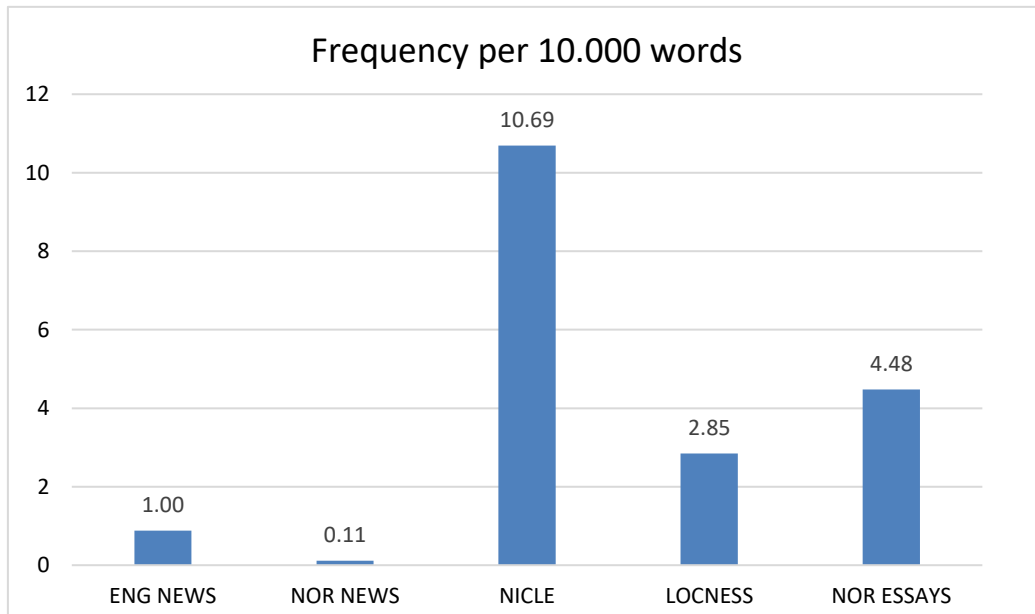
The difference between the L1 expert subcorpora is significant (see section 4.2) with more instances of *I think* in ENG NEWS, and NOR NEWS is differs significantly from NICLE.[36] This indicates that the overuse from NICLE writers is not influenced by transfer from Norwegian L1. The characteristics of the L1 novice writers seem to affect the NICLE writers, as LOCNESS contains more instances of *I think* than ENG NEWS, and NOR ESSAYS contains more instances of *jeg tror* than NOR NEWS. This indicates that developmental factors are influencing the NICLE writers, as they also have a higher frequency of *I think* compared to the L1 expert writers. The fact that writers from all three novice subcorpora mainly use *I think* to communicate humble subjectivity is evidence that supports the influence of developmental factors. However, when the frequency from the 100 NICLE texts are compared to the frequency of the 100 LOCNESS and NOR ESSAYS texts, we find that

---

[36] The difference between NOR NEWS and NICLE was found to be 99.99% certain (p< 0.0001).

NICLE writers overuse *I think* significantly in both cases.[37] [38] This means that developmental factors do not fully explain the excessive overuse, although it is most likely a contributing factor. It could conceivably be hypothesized that the excessive overuse of *I think* among NICLE writers has to do with the fact that they write in a foreign language. Further research should be undertaken in the future to investigate this hypothesis further.

As described in section 2.4.2, Ringbom (1998, p. 41) found that L2 advanced learners of English overuse *I think* compared to LOCNESS writers. The frequency was highest in the Swedish component of ICLE (16 per 10.000 words), followed by the French component of ICLE (10 per 10.000 words). LOCNESS on the other hand had a frequency of 3 per 10.000 words, which was lower than all the L2 learner groups included in the study. The frequency of *I think* among NICLE writers was found to be 10.69 in the present study. It was assumed that the NICLE writers would resemble the Swedish advanced learners, but these numbers seem to reveal that SWICLE writers overuse *I think* compared to NICLE writers. Interestingly, LOCNESS has the lowest frequency of *I think* by far, compared to the L2 advanced learner corpora, but as we have seen in section 5.2, LOCNESS has a significantly higher usage of *I think* compared to English L1 experts. This suggests that overuse of *I think* is a general phenomenon among L2 advanced learners of English, although the overuse varies among speakers with different L1s.

Baumgarten and House (2010) found that L2 speakers of English have a higher frequency of *I think* in relation the total usage of *I* in spoken discourse (53.7 percent and 25.5 percent) compared to L1 native speakers (17.5 percent) (see section 2.4.2). The results obtained from NICLE in the present project (see section 6.2) showed that the writers have a lower frequency of *I think* in relation to the total instances of *I* (12.8 percent) when compared to the L1 and L2 groups from Baumgarten and House's (2010) study. These findings, however, need to be interpreted with great care, as spoken discourse, in general, is less formal than written discourse, apart from speech holding and other scripted forms of speech. We know that the participants in Baumgarten and House's study had a general talk with the interviewer, so there were no signs of a formal setting. In addition, the fact that the expert L1 subcorpora have few

---

[37] The difference between NICLE and NOR ESSAYS was found to be 99.99 % certain (p< 0.0001).

[38] The difference between NICLE and LOCNESS was found to be 99.99 % certain (p< 0.0001).

instances of *I think* compared to the L1 novice subcorpora in the present project (see section 4.2 and 5.2), indicate that the recurrent word-combination is connected to informality, and expert writers are in the vast majority of cases trying to be as formal as possible. On that basis, it is no surprise that the percentage of *I think* in relation to total instances of *I* is lower in the written material from NICLE than in the three spoken groups. A quick examination of the 100 texts from ENG NEWS reveal that they use *I think* in 2.43 percent of the cases where *I* as subject is present, which again shows that NICLE writers have a higher percentage of *I think* in relation to the total instances of *I* when compared to L1 native speakers of English in written discourse.

Hasselgård (2009) suggested that NICLE writers overuse *think* and in self-reference clauses, and Paquot et al's. (2013) findings revealed overuse of the recurrent word-combination *I think*. Hasselgård did not report the actual numbers in her study, but Paquot et al. (2013) included a statistic of *I think* usage in NICLE and LOCNESS. The table below compares their findings to the findings from the present study. It also includes the instances found in ENG NEWS:

Table 28: Frequency of *I think* in Paquot, Hasselgård, & Ebeling (2013) and the present study

|	Paquot, Hasselgård, & Ebeling (2013)		The present study	
NICLE corpus size:	**128.093 words over 189 texts**		**67.332 words over 100 texts**	
LOCNESS corpus size:	**100.702 words over 88 texts**		**87.430 words over 100 texts**	
	Raw frequency	**Per 10.000 words**	**Raw frequency**	**Per 10.000 words**
NICLE	**108**	**10.80**	**72**	**10.69**
LOCNESS	**16**	**1.60**	**25**	**2.85**
ENG NEWS	**Corpus not included**		**8**	**0.88**

The results from the present study reflect those of Paquot et al. (2013). Norwegian advanced learners do seem to overuse *I think* compared to LOCNESS writers. The frequency difference in LOCNESS between Paquot et al. and the present study is surprisingly large per 10.000

words, but since there are few instances of *I think* in the first place, every occurrence found in a text will influence the statistic. Therefore, all findings in this study need to be interpreted and treated with great care, as the numbers alone do not tell the exact truth, but rather serves as an indicator.

7.2.2 *I feel* **and** *jeg føler*

In figure 5, the findings from all five subcorpora are presented, illustrating the differences in frequency of *I feel/jeg føler* per 10.000 words:

Figure 5: Frequency of *I feel/jeg føler* per 10.000 words in the five subcorpora
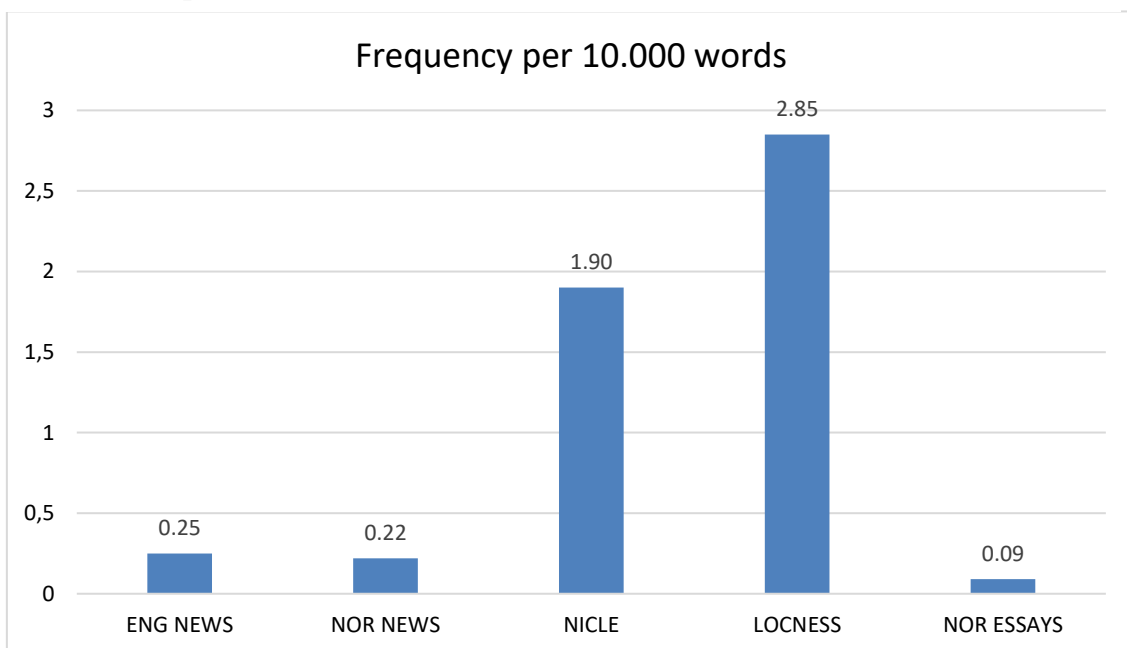
The frequency of *I feel/jeg føler* is quite similar in the L1 expert subcorpora, and NICLE have a significantly higher usage than NOR NEWS [39], which means that transfer from L1 does not seem to influence the NICLE writers. Interestingly, the LOCNESS subcorpus has a higher

---

[39] The difference between NOR NEWS and NICLE was found to be 99.9 % certain (p < 0.001).

frequency of *I feel* than in NICLE, but the difference was not found to be significant.[40] Since NICLE and LOCNESS resemble each other in usage, developmental factors seem to play a role in usage among NICLE writers. Quite surprisingly, *jeg føler* is rarely used among Norwegian L1 writers, as both experts and novices steer away from that particular recurrent word-combination. We know already that Norwegian L1 experts rarely use *I* at all in their texts, so the low frequency from NOR NEWS is expected, but the usage from the NOR ESSAYS writers was expected to be higher. Their reluctance to using *jeg føler* could perhaps partly be explained by looking at the frequency of *jeg tror*, which is often used in the same context. As we have seen, the NOR ESSAYS writers use the recurrent word-combination *jeg tror* more often than LOCNESS writers use *I think* (see section 5.2 and 7.2.1). However, that alone does not explain why the difference in *I feel / jeg føler* is so evident. Further studies should be conducted that cover a wider variety of recurrent word-combinations together with *I* as subject. That way, a pattern could perhaps be detected that show how a particular group uses a set of combinations at the cost of others.

The findings from figure 5 contradict Ädel and Erman (2012) who suggested that non-native speakers overuse *I* framed constructions compared to native speakers. Although NICLE has a higher frequency of *I feel* compared to ENG NEWS, the opposite is true when NICLE is compared to LOCNESS.

---

[40] The difference between LOCNESS and NICLE was not found to be significant.

7.2.3 *I know* and *jeg vet*

In figure 6, the findings from the five subcorpora are presented, illustrating the differences in frequency of *I know*/*jeg vet*:

Figure 6: Frequency of *I know*/*jeg vet* per 10.000 words in the five subcorpora
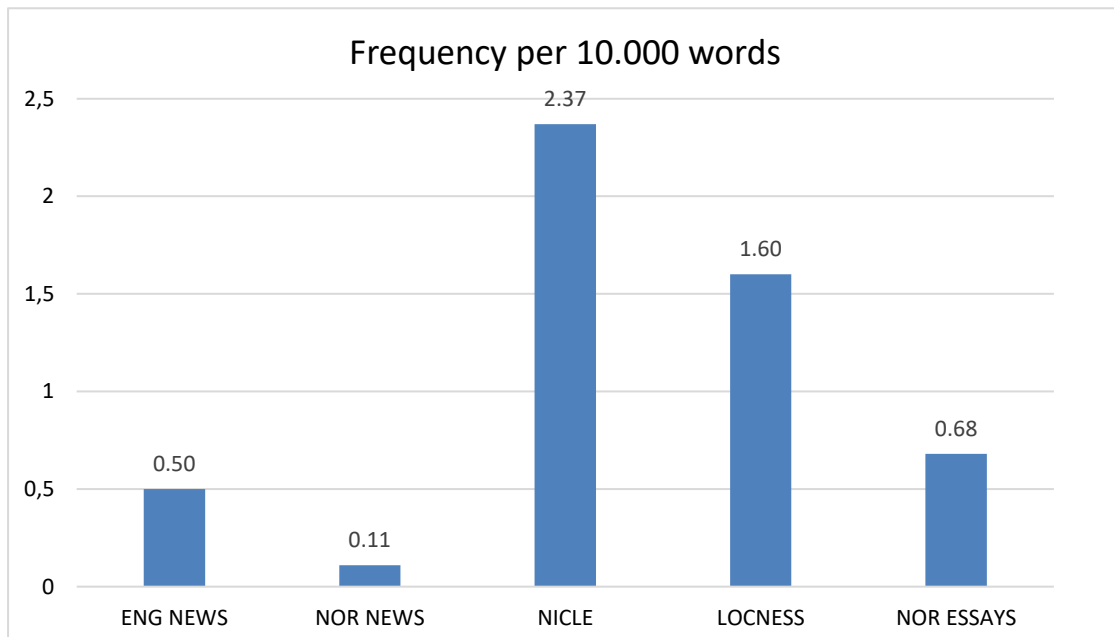
Transfer can be ruled out as an influencing factor as there is no significant difference between the L1 English and Norwegian expert texts, while there is a significant difference between NOR NEWS and NICLE.[41] Both LOCNESS and NOR ESSAYS are closer to NICLE than the two expert L1 subcorpora, which indicate that developmental factors play a role in NICLE writers' usage. However, a closer examination of the L1 novice subcorpora reveals that NICLE is significantly different from NOR ESSAYS [42], but not from to LOCNESS, although the frequency in NICLE is higher. This indicates that NICLE writers are mostly influenced by English L1 novice writers.

[41] The difference between NOR NEWS and NICLE was found to be 99.99 % certain (p < 0.0001)

[42] The difference between NICLE and NOR ESSAYS was found to be 99 % certain (p < 0.01)

The findings found in figure 6 correlate with Ädel and Erman (2012) who suggested that overuse of *I* framed constructions among non-native speakers compared to native speakers are taking place. However, since there is no significant difference between the frequency found in NICLE and LOCNESS, one cannot safely conclude that overuse is taking place among NICLE writers compared to native speakers.

7.2.4 *I would say* and *jeg vil si*

In figure 7, the findings from all five subcorpora are presented, illustrating the differences in frequency of *I would say*/*jeg vil si* per 10.000 words:

Figure 7: Frequency of *I would say*/*jeg vil si* per 10.000 words in the five subcorpora
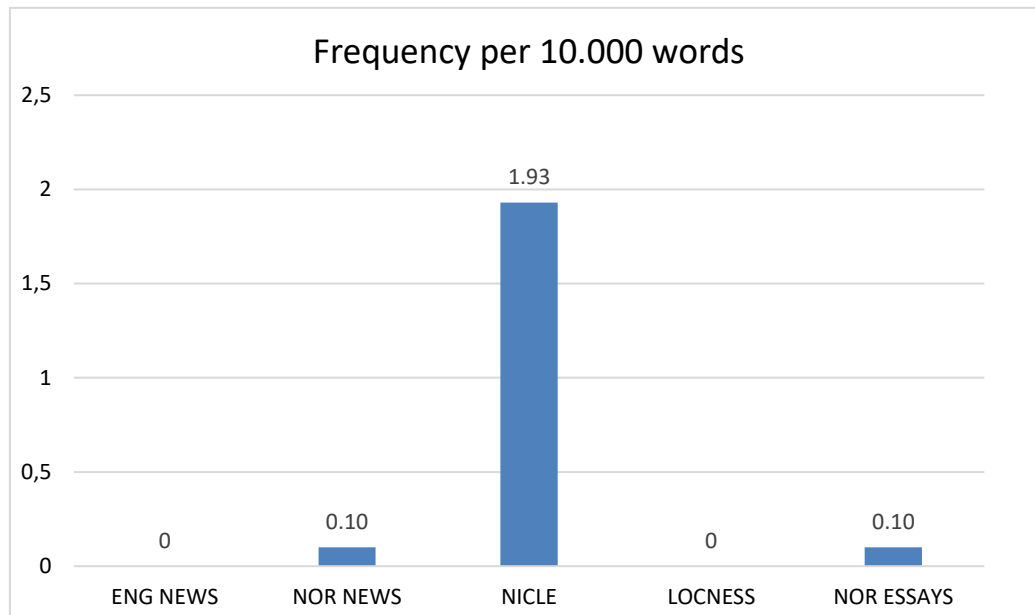
As can be seen from figure 7, only the NICLE writers tend to use the recurrent word-combination *I would say* in their texts. Only one instance of *jeg vil si* was found in the Norwegian L1 subcorpora, while no instances of *I would say* was found in the English L1 subcorpora. This means that both transfer from Norwegian and developmental factors looks to have little impact on the NICLE writers. The results from figure 7 contradict Hasselgård and Johansson's (2011) suggestion that overuse of *I would say* might be due to transfer from Norwegian and developmental factors, as NICLE have a significantly higher usage of *I would say* compared to the other subcorpora. [43] However, the results are correlating well with Ädel and Erman (2012) who argued that non-native speakers overuse *I* framed constructions.

---

[43] The difference between NICLE and the L1 subcorpora was found to be 99.99 % certain in all cases (p < 0.0001).

8. Conclusion

In this study, three research questions have been investigated. These will once again be listed here in section 8.1-8.3 where the findings are summarized. Section 8.4 will discuss the limitations of this thesis and give suggestions for further research.

8.1 Research question 1

a) **Do Norwegian advanced learners of English overuse *I* as subject in clauses with mental and verbal processes in their texts?**

This research question was investigated through examining differences in frequency between NICLE writers and English L1 expert writers. The results presented in chapter 6, section 6.1, strongly suggest that NICLE writers significantly overuse *I* as subject in mental processes compared to ENG NEWS. In clauses with verbal processes, the results also point in the direction of overuse from the NICLE writers, as the difference was significant here as well. However, these findings need to be interpreted with great care, as the material used in this study (30 texts from each subcorpus) is too limited to draw a safe conclusion. Although no other studies to date have investigated the frequency of *I* in clauses with mental and verbal processes among NICLE writers, the results found here correlate well with (Fossan, 2011; Hasselgård, 2009, 2012; Paquot et al., 2013) whose findings suggest general overuse of *I* as subject.

b) **Do Norwegian advanced learners of English overuse the recurrent word-combinations *I think*, *I feel*, *I know*, and *I would say* in their texts?**

As was noted in section 3.2 and 3.3.4, 100 texts from each subcorpus were used to provide more thorough results in the investigation of the recurrent word-combinations. Although the material used to investigate these combinations is larger in size, the sample size is still relatively small, and caution must be applied, as these findings might not be fully representative.

The findings suggest that NICLE writers overuse *I think* significantly when compared to ENG NEWS writers, and this is backed up by Ringbom (1998) who found that ICLE writers overuse *I think* compared to LOCNESS writers. When compared to the spoken material from L1 and L2 speakers in Baumgarten and House (2010), the NICLE writers had the lowest percentage

of *I think* in relation to the total instances of *I* as subject. However, when compared to ENG NEWS and LOCNESS, the NICLE writers had the highest frequency percentage wise. The NICLE writers also had significant overuse of the recurrent word-combinations *I feel*, *I know*, and *I would say* when compared to ENG NEWS writers.

8.2 Research question 2

Is there any evidence that transfer from Norwegian influences the Norwegian advanced learners of English, and, if so, how?

Chapter 4 in this present project examined expert L1 texts and it was revealed that NOR NEWS writers steer away from the use of *I* in their texts. In clauses with mental processes, the difference between NOR NEWS writers and NICLE writers was significant, and no traces of transfer were to be found. In clauses with verbal processes, however, the difference between NOR NEWS and NICLE was not significant, so transfer might have had an impact on usage here. Transfer does not seem to affect the NICLE writers' usage of the four recurrent word combinations, namely *I think*, *I feel*, *I know*, and *I would say*.

8.3 Research question 3

Is there any evidence that developmental factors influence the Norwegian advanced learners of English, and, if so, how?

In clauses with mental processes, developmental factors do not seem to affect the NICLE writers, as both NOR ESSAYS and LOCNESS differ significantly in frequency. The NOR ESSAYS writers look to influence the NICLE writers in clauses with verbal processes, as both subcorpora towers over the others in frequency, which suggests that developmental factors play a role here.

I think usage from NICLE writers seems to be influenced by developmental factors as NICLE has the highest frequency, and both novice L1 subcorpora have a higher frequency compared to their L1 expert subcorpora. However, the frequency in NICLE is significantly higher than LOCNESS and NOR ESSAYS, which means that developmental factors alone cannot explain the overuse compared to English expert texts, but it is most likely a contributing factor. Usage of *I feel* provided interesting results, as only NICLE and LOCNESS writers used the recurrent

word-combination actively in their texts. This suggests that NICLE writers are influenced by developmental factors, in the form of shared characteristics with L1 English writers. The NICLE writers' usage of *I know* seems to be influenced by developmental factors, as NICLE have a higher frequency compared to the other subcorpora, and the usage from novice L1 writers are higher than the usage from L1 expert writers. The NICLE writers overuse of the recurrent word-combination *I would say* do not seem to be influenced by developmental factors, as the combination is not found in LOCNESS, and in NOR ESSAYS, only one occurrence of *jeg vil si* is found.

8.4 Strengths and limitations of the present study

The present study has hopefully expanded the knowledge about NICLE writers' usage of *I* as subject through the use of the Integrated Contrastive Model. The implementation of L1 texts from both novice and expert writers has given a more nuanced picture on NICLE writers overuse of *I* as subject in their texts. This is because they have provided the opportunity to look for influence of transfer and developmental factors.

The material used in the present study has been somewhat limited. Therefore, all the conclusions have been treated with caution, as the findings might not be accurate.

A comment on the quantitative approach seems called for at this stage, as it brings with it both positives and negatives. In the investigation, 30 and 100 texts from each subcorpus have been used for comparison, but individual differences between the writers have not been accounted for. All the calculations between the subcorpora have been performed in a way that treats each subcorpus like a unit. This shows general tendencies, which is an advantage, but it does not account for individual differences that exist within each subcorpus. Instead, this study refers to the average "NICLE writer" and "LOCNESS writer" etcetera, which might not be accurate at all.

8.5 Suggestions for further research

Since this study was to a large degree quantitative, more research should be undertaken that focuses on qualitative aspects. Conducting interviews with a number of students might help explain the reasoning behind the excessive overuse of *I* as subject compared to native speakers. It would also be interesting to see other studies that investigate the usage of *I* as subject among learners of different L1 backgrounds through use of the Integrated Contrastive Model. In that way, we can detect whether transfer and developmental factors appear to influence the learners from different L1 backgrounds in similar or different ways.

Bibliography

Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. Amsterdam: John Benjamins.

Ädel, A. and Erman, B. (2012). Recurrent word combinations in academic writing by native speakers and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, *31*(2), 81-92.

Aijmer, K. (1997). I think – an English modal particle. In T. Swan & O.J. Westvik (Eds.), *Modality in Germanic Languages: Historical and comparative perspectives* (pp. 1- 47). Berlin/New York: Mouton de Gruyter.

Altenberg, B. (1998). On the Phraseology of Spoken English. The evidence of Recurrent Word-combinations. In A. P. Cowie (Ed.), *Phraseology, Theory, Analysis and Applications* (pp. 101-122). Oxford: Oxford University Press.

Baumgarten, N and House, J. (2010). I think and I don't know in English as a lingua franca and native English discourse. *Journal of Pragmatics*, *42*, 1184-1200.

Bazerman, C. (1994). Systems of genres and the enactment of social intentions. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 79-101). London: Taylor and Francis.

Benson, C. (2002). "Transfer/Cross-linguistic Influence". *ELT Journal*, *56*(1), 68-70.

Cutting, J. (2000). Written errors of international students and English native speaker students. In G. Blue., J. Milton & J. Saville (Eds.), *Assessing English for Academic Purposes* (pp. 97-113). Frankfurt: Peter Lang.

Dagbladet. (n.a.). Retrieved from https://www.dagbladet.no/

Fossan, H. (2011). *The writer and the Reader in Norwegian Advanced Learners' Written English: A corpus-based study of writer/reader visibility features in texts by Norwegian learners of English and native speakers of English* (Master's thesis). University of Oslo.

Gilquin, G. & Paquot, M. (2008). Too chatty. Learner academic writing and register variation. *English Text Construction*, 1:1, 41-61.

Gilquin, G. (2000/2001). The Integrated Contrastive Model. Spicing up your data. *Languages in Contrast*, *3*(1), 95-123.

Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer (Ed.), *Languages in Contrast. Text-based cross-linguistic studies* (pp. 37-51). Lund: Lund University Press.

Granger, S. (1997). On identifying the syntactic and discourse features of participle clauses in academic English: native and non-native writers compared. In J. Aarts., I. de Mönnink & H. Wekker (Eds.), *Studies in English Language and Teaching* (pp. 185-198). Amsterdam: Rodopi.

Granger, S. (1998). The computer learner corpus: a versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on Computer* (pp. 3-18). London: Longman.

Granger, S. (2004). Computer Learner Corpus Research: Current Status and Future Prospects. In U. Connor & T. Upton (Eds.), *Applied Corpus Linguistics. A Multidimensional Perspective* (pp. 123-145). Amsterdam: Rodopi.

Granger, G., Dagneaux, E., Meunier, F. & Paquot, M. (2009*). International Corpus of Learner English v2*. Louvain-la-Neuve: Presses universitaires de Louvain.

Hasselgård, H. & Johansson, S. (2011). Learner corpora and contrastive interlanguage analysis. In F. Meunier., S. De Cock., G. Gilquin & M. Paquot (Eds.), *A taste for Corpora. In honour of Sylviane Granger* (pp. 33-61). Amsterdam: John Benjamins Publishing Company.

Hasselgård, H. (2009). Thematic choice and expressions of stance in English argumentative texts by Norwegian learners. In K. Aijmer (Ed.), *Corpora and Language Teaching* (pp. 121-139). Amsterdam: John Benjamins Publishing Company.

Hasselgård, H. (2012). Den norske måten å skrive engelsk på. *Bedre skole, 33*(1), 64-67.

Hellekjær, G. O. (2005). *The acid test: does upper secondary EFL instruction effectively prepare Norwegian students for the reading of English textbooks at colleges and universities?* (Doctoral dissertation). University of Oslo.

Holmes, J. & Nesi, H. (2009). Verbal and Mental Processes in Academic Disciplines. In M. Charles., D. Pecorari & S. Hunston (Eds.), *Academic Writing: At the Interface of Corpus and Discourse* (pp. 58-72). New York: Continuum International Publishing Group.

Hyland, K. & Milton, J. (1997). Qualification and Certainty in L1 and L2 Students' Writing. *Journal of Second Language Writing*, *6*(2), 183-205.

Lado, R. (1957). *Linguistics across cultures: applied linguistics for language teachers*. Michigan, USA: Michigan University Press.

Leech, (1998). Preface. In S. Granger (Ed.), *Learner English on Computer* (pp. xiv-xx). New York: Longman.

Lorenz, G. (1999). Learning to cohere: Causal links in native vs. non-native argumentative writing. In B. Bublitz., U. Lenk., & E. Ventola (Eds.), *Coherence in Spoken and Written Discourse. How to Create it and how to Describe it* (pp. 55-75). Amsterdam: John Benjamins.

McCrostie, J. (2008). Writer visibility in EFL learner academic writing: A corpus-based study. *ICAME Journal*, 97-114.

Meyer, C. F. (2002). *English Corpus Linguistics. An introduction.* New York: Cambridge University Press.

O'Donnell, M., Zappavigna, M & Whiteclaw, C. (2009). A survey of process type classification over difficult cases. In C. Jones & E. Ventola (Eds.), *New Developments in the Study of Ideational Meaning: Form Language to Multimodality* (pp. 47-64). London: Continuum Publishers.

Odlin, T. (1989). *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.

Paquot, M. (2010) *Academic Vocabulary in Learner Writing. From Extraction to Analysis*. New York: Continuum.

Paquot, M., Hasselgård, H., & Ebeling, S. O. (2013). Writer/reader visibility in learner writing across genres: A comparison of the French and Norwegian components of the ICLE and VESPA learner corpora. In S. Granger., G. Gilquin & F. Meunier (Eds.), *Twenty Years of Learner Corpus Research: Looking back, Moving ahead. Corpora and Language in Use* (pp. 377-387). Presses universitaires de Louvain.

Petch-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on Computer* (pp. 107-118). New York, USA: Longman.

Rayson, P. (n.d.). *Log-likelihood calculator*. Retrieved November 7, 2017, from http://ucrel.lancs.ac.uk/llwizard.html

Recski, L. J. (2004). Expressing Standpoints in EFL Written Discourse. *Revista Virtual de Estudos da Linguagem – ReVEL V. 2, n. 3*.

Ringbom, H. (1998). Vocabulary frequencies in advanced learner English: a cross-linguistic approach. In S. Granger (Ed.), *Learner English on Computer* (pp. 41-53). New York: Longman.

Rørvik, S. (2013). *Texture in learner language* (Doctoral dissertation). University of Oslo.

The Guardian (n.d.). Retrieved from https://www.theguardian.com

Thompson, G. (2014). *Introducing Functional Grammar* (3rd ed.). Croydon, GB: Routledge.

Université catholique de Louvain. (2017). *Centre for English Corpus Linguistics*: LOCNESS. Retrieved November 16, 2017, from https://uclouvain.be/en/research-institutes/ilc/cecl/locness.html

University College London. (n.d.). *Cautious Language and Hedging*. Retrieved from http://www.ucl.ac.uk/ioe-writing-centre/develop-academic-voice/caution-hedging

Vass, H. (2004). *Socio-cognitive aspects of hedging in two legal discourse genres*. Retrieved from http://www.redalyc.org/html/2870/287026304007/