Discourse markers in written learner English

A corpus-based study of the discourse markers *so, like, actually, anyway, well, you know* and *I mean* in written Norwegian learner language

Michaela Sandholtet

A thesis presented to the Department of Literature, Area

Studies and European Languages

UNIVERSITY OF OSLO

May 2018

II

Discourse markers in written learner English

A corpus-based study of the discourse markers *so*, *like*, *actually*, *anyway*, *well*, *you know* and *I mean* in written Norwegian learner language

Michaela Sandholtet

MA thesis in English Linguistics

ENG4790 – Master's Thesis in English:
Secondary Teacher Training

Supervisor: Kristin Bech

2018

Discourse markers in written learner English: A corpus-based study of the discourse markers *so*, *like*, *actually*, *anyway*, *well*, *you know* and *I mean* in written Norwegian learner language

Michaela Sandholtet

IV

Abstract

This thesis presents an investigation of discourse markers in written Norwegian learner language. Previous studies indicate that learners of English tend to embrace a style of writing which is influenced by oral language. The aim of this thesis is to find out whether advanced Norwegian learners of English overuse discourse markers in their writing compared to English native speakers, and to find out how Norwegian learners of English use discourse markers compared to English native speakers in their writing. This study is corpus-based, and the Norwegian component of the International Corpus of Learner English (ICLE-NO) and the Louvain Corpus of Native English Essays (LOCNESS) have been used to perform a quantitative and qualitative study of the discourse markers *so*, *like*, *actually*, *anyway*, *well*, *you know* and *I mean*. This study shows that the Norwegian learners of English in ICLE-NO overuse discourse markers in their writing compared to native speakers in LOCNESS. The analysis also shows that the Norwegian learners use discourse markers with an interpersonal function more often than the native speakers. This coincides with previous research which has found that Norwegian learners of English tend to show reader/writer visibility to a greater extent than both native speakers of English and other learner groups. There seem to be several reasons for this overuse and use of discourse markers in Norwegian learner writing, such as differences in writing cultures, register unawareness due to insufficient teaching and lack of sufficient training in academic writing.

Acknowledgements

This semester has surely been exciting! I have been writing my thesis, I have been working as a teacher and on top of that, I got married. My time as a student has come to an end (for now) and I would like to thank those who have kept me going all these years, and those who have helped me to make my dream possible: to become a teacher of English and Norwegian.

First of all, I would like to thank my supervisor Kristin Bech for guiding me through this project and cheering me up from the start. You made me feel confident and relaxed before starting this project, which has helped me to avoid a lot of extra stress and pressure this hectic semester. I value and appreciate all the time you have spent on helping me with my thesis.

To Anders, my husband and best friend, who always tells me that "everything is going to be okay" and "you can do this". Thank you for always supporting me and always helping me when I feel stressed. Thank you for being interested in my work and thank you for all our interesting conversations about language. Without you, I would never have had the slightest chance to finish my studies!

To Jeanette, my mother and my role model. You have always inspired me to work hard and to do my best. You have never told me who I should become or what I should do with my life, and that has made me confident in making my own decisions and to follow my dreams.

To my cousin Beatrice. Thank you for always listening to me, and thank you for putting up with my nonsense, and sometimes complaints about being a student. You are amazing!

Thank you.

Oslo, 19 May 2018

Michaela Sandholtet

VIII

Table of Contents

List of tables

List of figures

List of abbreviations

CIA – Contrastive Interlanguage Analysis

EFL – English foreign language

FL – Foreign language

IV – Interlanguage variety

L1 – First language

L2 – Second language

RLV – Reference language variety

RQ – Research question

SFL – Systemic Functional Linguistics

WS – WordSmith tools

Corpora mentioned

BNC – The British National Corpus

ICLE – The International Corpus of Learner English

ICLE-NO – The Norwegian component of the International Corpus of Learner English

ICLE-SE – The Swedish component of the International Corpus of Learner English

ICLE-US – The American component of the Louvain Corpus of Native English Essays

LOCNESS – The Louvain Corpus of Native English Essays

1 Introduction

The field of second language research is devoted to the research of learner performance: those who are in the process of acquiring and learning a second language. Since the compilation of digital corpora, research within this field has flourished. Digital corpora give second language researchers (and other researchers) access to a vast amount of language, which makes it possible to perform quantitative and qualitative studies on a larger scale than before. This opportunity has yielded many interesting research projects. One finding made is the tendency among learners of English to overuse features of spoken language in writing compared to native speakers of English. This has even been observed in texts written by advanced learners of English, i.e. learners who use English in higher education and have studied and used English for many years. Previous studies such as Gilquin and Paquot (2008), Altenberg (1997), Aijmer (2002), Ädel (2008), Hasselgård (2009, 2016) and Fossan (2011) have all found an overuse of several different features that are associated with the oral register in learner writing. This style of writing is considered more informal and personal, and is not considered typical of the academic genre in English. Therefore, researchers are discussing whether learners of English in general are unaware of register differences, or whether there are other possible reasons for this overuse. The previous studies presented above have all sparked an interest in the investigation of the use of oral features in Norwegian learner language, since there is to date limited research on the use of spoken-like features in Norwegian learner language.

1.1 Aim and scope

The aim of this study is to find out whether advanced learners of English overuse oral features in their texts compared to native speakers of English, and to investigate how Norwegian learners use these features in their writing. Thereby, I hope to add to the discussion of whether learners of English are in fact more influenced by oral language in their writing than native speakers are. The oral feature I have chosen to investigate is discourse markers, due to the fact that there is general agreement that these are associated with and used in the oral register. Also, there is limited research on discourse markers in Norwegian learner writing. The definition of discourse markers will be further presented in Chapter 3. To perform this study, I have chosen to do a contrastive interlanguage analysis using two corpora: the Norwegian part of the International Corpus of Learner English (ICLE-NO) and the Louvain Corpus of Native

English Essays (LOCNESS). The method and the corpora will be further outlined and discussed in Chapters 4 and 5. The study is both quantitative and qualitative: the discourse markers will be investigated in terms of their frequency in the two corpora, their position, and their function in the sentence. In addition to a quantitative approach, a qualitative approach has been chosen to get a fuller understanding of how these markers are used in writing by learners of English compared to native speakers. If an overuse is revealed in the quantitative analysis, the functional analysis will hopefully prove useful to discuss why learners of English overuse discourse markers in their academic texts.

This study is based on a pre-study (Johnsson[1] 2017), where the discourse markers *so* and *well* were investigated in texts written by Norwegian learners. In this pre-study, I found that advanced Norwegian learners of English in ICLE-NO overuse *so* and *well* in their academic writing compared to the English native speakers in LOCNESS. This study was performed under certain restrictions such as length and a limited amount of time. Even though I found some interesting results, the study was limited because I only had the opportunity to investigate two discourse markers. Therefore, I wanted to perform a more nuanced study that included a few more discourse markers to hopefully yield a more substantial result. I have chosen to expand my pre-study by adding the discourse markers *like*, *actually*, *anyway*, *you know* and *I mean* to this study. *So* and *well* are also part of the investigation. Even though I have analyzed the material for *so* and *well* in the pre-study, I chose to analyze the material again since the present study focuses further on the different functions of the discourse markers. Therefore, some instances may have been assigned a different function in this study than in the pre-study.

1.1.1 Research questions

Based on previous research and the aim of this paper, I have defined three research questions which are presented below:

> **RQ1:** Do Norwegian learners of English overuse discourse markers in their writing compared to native speakers of English?

> **RQ2:** If they overuse discourse markers, how do Norwegian learners of English use discourse markers in their writing compared to native speakers of English?

---

[1] Johnsson was my surname before I changed to Sandholtet.

RQ3: If the answer to RQ1 is 'yes', what are possible reasons for this overuse of discourse markers in Norwegian learner writing?

Based on previous research performed on learners from different first language backgrounds, it would be natural to suggest that also Norwegian learners of English overuse oral-like features in their writing. The question is whether they use discourse markers in their writing, and if so, to what extent. My hypothesis is that the learners in ICLE-NO in fact overuse discourse markers compared to native speakers. If the quantitative analysis confirms my suspicions, the qualitative functional analysis may help to answer RQ3, and reveal some possible reasons for this overuse.

1.2 Thesis outline

This study consists of a total of seven chapters. Chapter 1 presents some background information and the aim and scope of the paper, and also outlines the research questions that guide the study. In Chapter 2, some selected important previous studies that have observed spoken features in learner language are presented. Chapter 2 also contains a section that presents possible reasons for overuse of spoken-like features in learner language. Chapter 3 takes a closer look at the spoken feature investigated in this study: discourse markers. Firstly, discourse markers as a group is defined, and thereafter, all discourse markers in this study are outlined in terms of their characteristics and functions. Chapter 4 gives a presentation of corpus methods in second language research and learner corpora, and gives a short introduction to Contrastive Interlanguage Analysis (CIA). Chapter 5 presents the material in this study: ICLE-NO and LOCNESS. They are both outlined and also compared to each other in terms of representativeness and authenticity. The framework of classification of the material is also included in Chapter 5. Chapter 6 presents the results from the quantitative and qualitative analyses, followed by a summary and discussion of the findings. In chapter 7, the study is summed up, along with concluding remarks and an overview of pedagogical implications. Lastly, Chapter 7 presents some limitations of the study and suggestions of further research.

2   Previous studies

The following chapter introduces some selected previous studies that reveal the spoken-like nature of learner writing from different L1 backgrounds, while section 2.2 narrows the focus to Norwegian learners of English. Thereafter, section 2.3 presents some potential reasons for the overuse of speech features in learner writing.

2.1 Previous research on spoken-like features in learner writing

This section presents previous research dealing with overuse of certain spoken-like features in advanced learner writing. These projects have sparked my interest in investigating oral features in Norwegian learner language. A selection of important research will be introduced, namely Altenberg (1997), Aijmer (2002) and Ädel (2008), as well as one of the main inspirations that encouraged the development of this project, Gilquin and Paquot's (2008) study of learner academic writing and register variation. All the studies presented in this section indicate that learners of English in general seem to lack sufficient knowledge of how to write academic texts in English.

2.1.1 Gilquin and Paquot 2008

In their study, Gilquin and Paquot (2008) investigate various spoken-like features in writing produced by learners from several different L1 backgrounds, and argue that learners of English use certain items that are associated with speech in their writing (2008, 45). Their analysis shows that there are certain characteristics which are more commonly used in spoken discourse and less prevailing in academic writing that are overused by learners of English:

- Certain expressions of possibility, such as *maybe*, and underuse of other commonly used expressions in native production such as *apparently* and *presumably*.

- Items expressing certainty, such as *really*, *of course* and *certainly*.

- Expressions associated with a high degree of writer visibility. Learners show personal stance in their texts, in form of using personal pronouns and personal structures such as *I think that* or *it seems to me*. Moreover, they are more visible when they introduce new topics or ideas which they show using constructions such as *I would like* and *I am going to talk about*.

- Items in initial and final position: sentence initial *and* and sentence final *though*.

Gilquin and Paquot (2008) conclude that these features can be generalized to all academic interlanguages[2] of English (2008, 57), and that this overuse of spoken-like features in writing can "account for learners' 'chatty' style" (2008, 57).

2.1.2 Altenberg 1997

In his study, Altenberg (1997) explores vocabulary, noun phrase complexity and involvement and detachment in argumentative writing by Swedish learners of English in the Swedish component of the ICLE corpus (ICLE-SE). His findings show a general tendency for Swedish learners of English to be influenced by informal language in their argumentative writing (1997, 130). Swedish learners tend to use lexical items which are classified as informal and they use simpler noun phrase constructions compared to native speakers, which are more common to use in speech than in academic writing (1997, 126). Altenberg's (1997) study also shows that Swedish learners underuse passive constructions, which are more common in academic writing, and overuse words and phrases expressing personal involvement, such as *well, you see, I think*, tag questions, first person pronouns, disjuncts and questions, compared to the native speakers in LOCNESS (1997, 129). Altenberg's (1997) findings suggest that Swedish learners and English native speakers choose a different approach when writing argumentative texts: the English students are not as present in their argumentative writing and they take a more objective stance, while Swedish learners of English are more personally involved and interactive in their argumentative writing (1997, 130). He concludes that "[t]he difference between the Swedish learners and the native speakers is so striking that it is justified to talk about two entirely different approaches to argumentative writing" (1997, 130).

2.1.3 Aijmer 2002

Aijmer (2002) investigates modal auxiliaries, modal adverbs and the combination of both in the interlanguages of Swedish learners of English and compares this learner group to French and German learners. Her analysis shows that there is an extensive overuse of modal auxiliaries and adverbs by Swedish, French and German learners. Modal auxiliaries and modal adverbs are markers of stance, and the use of some of these modal expressions is more likely to be associated with speech, which in turn creates a chatty or spoken-like style in texts written by learners of English (2002, 73). Even though it is necessary to perform further

---

[2] The language of a second- or foreign language learner.

studies on several other learner groups to generalize these findings, Aijmer (2002) points out that these findings "[…] are of interest, both in what they reveal about modality in learner writing, and in the research avenues they open up" (2002, 72).

2.1.4 Ädel 2008

Ädel (2008) addresses the overuse of reader/writer visibility in her comparative study of metadiscourse in American English, British English and advanced Swedish learner English. She distinguishes between 'personal' and 'interpersonal' metadiscourse, where personal metadiscourse is when the writer makes explicit reference to him- or herself or the reader while impersonal metadiscourse is when the writer organizes the text without explicit reference to him-or herself or the reader (2008, 51). In Ädel's (2008) study, advanced Swedish learners of English use both personal and impersonal metadiscourse more frequently in their argumentative writing compared to American and British native speakers. Ädel (2008) concludes that Swedish learners of English are most visible in their writing, while the British writers are least visible (2008, 60).

2.2 Previous research on spoken-like features in Norwegian learner writing

Previous research such as Gilquin and Paquot (2008), Altenberg (1997), Aijmer (2002) and Ädel (2008) suggests that those who are in the process of acquiring English on an advanced level overuse certain spoken-like features in their writing. This would also most certainly include Norwegian learners of English. This section presents previous studies on the overuse of speech features in Norwegian interlanguage. Furthermore, the pre-study for this project, Johnsson (2017), will be introduced.

2.2.1 Hasselgård 2009

Hasselgård (2009) examines whether Norwegian learners of English transfer certain structures from the Norwegian language and Norwegian style of writing, and thus investigates whether Norwegian learners have the ability to adapt when they write in certain genres in English. She looks at different patterns in initial position and finds that Norwegian learners overuse several of them. One of those patterns concerns writer visibility and subjective stance, where Norwegian learners overuse expressions such as *I think, I believe, I guess* and *I suppose* (2009, 133). Not only do Norwegian learners refer to themselves in their English writing, they

also do this to a somewhat higher degree compared to other learners, for example Swedish learners of English (2009, 133). Hasselgård's (2009) study also shows that Norwegian learners, like Swedish learners (c.f Aijmer 2002), overuse other markers of stance such as modality and adverbs/adverbials.

2.2.2 Fossan 2011

In her master's thesis, Fossan (2011) investigates reader/writer visibility in Norwegian learner language. Similar to Ädel's (2008) study on Swedish learners, Fossan finds that also Norwegian learners are more present in their academic writing compared to English native speakers (2011, 153). Fossan (2011) also finds that Norwegian learners are distinctly more visible in their writing compared to other learner groups of English (2011, 153).

2.2.3 Hasselgård 2016

Hasselgård's (2016) study focuses on the use of metadiscourse in Norwegian interlanguage. She compares Norwegian learners to novice writers of English, but also to expert writers in two disciplines: linguistics and business. Similar to Ädel's (2008) study of metadiscourse in Swedish learner written English, Hasselgård (2016) concludes, as suspected, that Norwegian learners who write in both disciplines use both personal and interpersonal metadiscourse more frequently in their English writing compared to novice L1 writers and expert writers (2016, 127). The biggest difference between the groups in the study is found in the interpersonal category. Norwegian learners use both personal and impersonal metadiscourse more frequently than any other group in the study. However, Norwegian learners seem to favor personal over impersonal metadiscourse (2016, 127).

2.2.4 Pre-study: Johnsson 2017

The pre-study for this project by Johnsson (2017) investigates the use of discourse markers in written production by advanced Norwegian learners of English. Discourse markers are associated with speech production, and therefore, this pre-study aims at adding to the discussion of whether learners of English in general lack the ability to adapt their language to different register and genres. The analysis shows an overuse of the two discourse markers studied, *so* and *well*, which indicates that Norwegian learners use spoken-like features in their writing. The study also shows that both discourse markers are used with an interpersonal function more frequently by Norwegian learners than by native speakers: the learners use

these discourse markers to show their presence in the text. These findings resonate with the conclusions made by Hasselgård (2009, 2016) and Fossan (2011).

2.3 Possible reasons for overuse of spoken-like features in learner writing

This section gives an account of possible reasons for the overuse of spoken-like features in learner writing, and tries to explain why learners as a group have a hard time to adapt their language to the academic written genre in English.

2.3.1 Influence of speech

One possible explanation for the spoken-like nature of learner writing may be the influence of the English spoken language the learners hear around them, through movies, television, series, YouTube and other channels. If learners are heavily influenced by these channels, they may resort to this type of informal spoken language when they do not know how to approach the writing task. It may be a learner strategy in order to feel that they master the task in hand; the learner choose words that they feel safe with and this in turn creates the informal tone (Hasselgren 1994, 243). Even though the English spoken language may have an impact on what choices learners make when writing, there are some problems with this explanation. Not all learner groups are equally influenced by the English language in their everyday lives; some groups rather learn English through instruction at school. Additionally, Gilquin and Paquot (2008) find this explanation less likely since the ICLE corpus was collected in the 1990s and the learners then were not as influenced by English media as some learner groups are today.

2.3.2 Transfer from the native language

It is natural to resort to the explanation that the oral nature of learner texts is influenced by the learners' native language. However, as Gilquin and Paquot (2008) suggest, the oral nature of written L2 production seems to be a common problem for all learners of English (2008, 42), and is thus not associated with a specific learner group. Even though this may be true, Gilquin and Paquot (2008) also report a particular overuse of imperative structures associated with speech (*let's/let us*) by French learners, which seems to be due to the fact that French learners use imperatives more frequently in their native writing (2008, 54). In addition, French learners seem to overuse structures which are more common in informal English written

genre. These French "translational equivalents are deeply entrenched in French speakers' mental lexicon" (Paquot 2013, 410), and therefore "anchored to important communicative or metatextual functions" (Paquot 2013, 411). Thus, French learners may be influenced by this style when they write in English.

Another example of possible transfer from the native language is reported in the findings of Hasselgård (2009, 137). Extrapositioning and the use of subjective stance markers seems to play part in the structural choices Norwegian L2 writers make in English writing. Moreover, as Aijmer (2002) points out, the overuse of modal expressions in English writing by Swedish learners can be due to transfer from Swedish. Contrary to English, epistemic modality in Swedish is usually expressed with either an adverb or an adverb and a modal verb. Consequently, the Swedish learners may use unnecessary complements to the modal auxiliary, which is neither needed nor preferred in English (Aijmer 2002, 72). These findings suggest that transfer from the native language may be part of the reason why learners overuse oral features in their writing.

2.3.3 Register unawareness

Another possible reason for the learners' overuse of speech in their English written discourse could be that they are not aware of certain differences between the spoken and written register, and differences between different written genres in English; they lack sufficient communicative competence. One reason for this possible unawareness may be insufficient training in writing different genres, but it may also be faulty or poor teaching (Altenberg 1997, 130), or the actual teaching process itself. Gilquin and Paquot (2008) mention one example of linking adverbs, where some English textbooks do not distinguish different linking phrases from each other (such as *therefore, so, hence* and *because of this*) in terms of formality/informality, but rather gives the impression that these words and phrases are synonymous, when they are in fact used in different genres in English (2008, 55). The instruction in textbooks may thus impact the learners' choice of linking adverbs in English, which could result in an inappropriate use of these adverbs. Although register unawareness is one possible reason for the overuse of spoken-like features in written discourse, "it remains to be seen, however, whether lack of register awareness is a typical feature of EFL learner writing or whether it is a more general characteristic of novice writing" (Paquot 2010, 152).

2.3.4 The learners' own development

One factor we must consider is the fact that the learners in the ICLE corpus are novice writers. To illustrate this, Gilquin and Paquot (2008) compared the learner results to a native novice writer group and a native expert writer group. The comparison showed that also native novice writers overuse features of speech in their writing, but to a lesser degree than learners of English (Gilquin and Paquot 2008, 56). This is also supported by Hasselgård (2016), who found that the novice writer L1 group in her study used metadiscourse more frequently compared to the expert writer group (2016, 124). This shows that "an oral tone in writing is not limited to foreign learners, but is actually very much part of the process of becoming an expert writer" (Gilquin and Paquot 2008, 57).

2.4 Considerations and further research of spoken-like features in learner writing

Although Gilquin and Paquot's (2008) study provides a valuable overview of the overuse of certain spoken-like features in learner academic writing, there are some limitations which need to be addressed. The limitations concern the comparison of different text types and the level of writing proficiency. Gilquin and Paquot (2008) use the spoken and written academic parts of the BNC (British National Corpus), which consist of book samples and articles from several different disciplines, and spoken discourse from various genres (2008, 44). The learner corpus used in the study is the ICLE corpus, which consists of argumentative texts and essays written by learners with a proficiency level of higher intermediate to advanced level. Even though these writers are advanced learners of English, they cannot be considered experts; writers of books and journal articles. In addition, even though argumentative writing could be considered academic, it is a text type which differs from the genre of books and articles in terms of language use. In one part of their study they compare the learner data in ICLE to novice writing in LOCNESS. However, it is not clear if they have compared all words and phrases in the study or if they have only selected a few for comparison. To yield more comparable results concerning learners' and native speakers' use of spoken-like features in written discourse, we would preferably want to compare the argumentative writing of learners to novice native speakers. Therefore, the LOCNESS corpus, which contains argumentative essays written by novice native writers, was chosen for this study as a more comparable corpus to ICLE-NO.

3 Discourse markers and previous frameworks of analysis

This chapter provides an overview of the speech feature investigated in this study: discourse markers. Section 3.1 presents the different functions which we can assign discourse markers and section 3.2 offers a short summary of how discourse markers are interpreted and defined in this study. In addition, it includes a presentation of the features of discourse markers (semantic, syntactic, functional and stylistic features) that are relevant for this study. Due to the diversity of the discourse marker group, it is necessary to outline the different features of each discourse marker based on the group's common features. Therefore, we take a closer look at the functional and syntactic features of each selected discourse marker: *so, like, actually, anyway, well, you know* and *I mean*. I have retrieved all examples from the spoken part of the British National Corpus (BNC).

3.1 Metafunctions

Systemic Functional Linguistics (SFL), founded and developed by Michael Halliday, is one of many approaches to language. SFL holds that language is not only a large system of linguistic elements that are part of larger units, but that these elements also have a purpose, and they are uttered or written to express something. Therefore, language is functional and semantic. Halliday has introduced three metafunctions of language: the ideational, textual and interpersonal functions. These functional categories, "provide an interpretation of grammatical structure in terms of the overall meaning potential of the language" (Halliday and Mattheissen 2004, 52). When we assign a function to an item in the sentence, we "show what part the item is playing in any actual structure" (Halliday and Mattheissen 2004, 52). Items which are considered to have a textual function organize language and create cohesion, while items with an interpersonal function are there to "form patterns of exchange involving two or more interactants […]" (Halliday and Mattheissen 2004, 589). The ideational function is concerned with human experience and how we express this experience in our language (Halliday and Mattheissen 2004, 29).

3.2 Discourse markers

Discourse markers are words or phrases such as *so, like, oh, you know, um, I mean, well*, which are a natural part of conversations and interactions. All discourse markers have different grammatical properties, which makes it difficult to characterize this group of words as a word class (Sandal 2016, 7). However, we can establish some common features and functional similarities of these words when they operate as discourse markers in an utterance.

There is general agreement (Biber, Johansson, Leech, Conrad and Finegan (1999), Müller (2005), Buysse (2012), Sandal (2016)) that discourse markers belong to the spoken register, thus the use of discourse markers is usually associated with informal language. The words themselves are said to have little or vague meaning (Müller 2005, 6; Sandal 2016, 9), but, when they are used, they add some kind of extra meaning to the utterance (Müller 2005, 1). The meaning which the utterance express is not dependent on the discourse marker, which means that the marker can be omitted without changing the essential meaning. Even though discourse markers are voluntary, they help the speaker to organize the speech, and thus they "have the general metainteractional (or procedural) function to comment on or signal how an upcoming utterance fits into the developing discourse" (Aijmer 2002, 265), and/or help the speaker to indicate a relationship between the speaker, hearer and the message (Biber et al. 1999, 1086). Thus, they have a semantic function in the sentence, which can be ideational, textual or interpersonal. Table 1 summarizes some of the functions and uses of discourse markers.

Table 1: Summary of functions and uses of discourse markers

- Initiate discourse
- Mark a boundary in discourse (change topic)
- Preface a response or reaction
- Aid the speaker in holding the floor
- Bracket the discourse either cataphorically or anaphorically
- Mark foregrounded or backgrounded information
- Effect an interaction or sharing between speaker and hearer

Source: **Müller (2005, 9)**

Discourse markers are characterized as multifunctional, since they are able to serve different functions in an utterance at the same time, and also because they facilitate "the hearer's task of understanding the speaker's utterances" (Müller 2005, 8) while as previously mentioned, adding extra pragmatic meaning to the utterance. Syntactically, discourse markers are usually

placed in initial position in a sentence, but depending on the function of the marker, they can be placed in all positions, also in medial and final position (Müller 2005, 5).

3.2.1 *So*

So is an adverb and connector, but *so* is also used as a discourse marker. When *so* functions as an adverb or conjunction, it cannot be omitted from the sentence without changing the meaning. Examples (1) and (2) from the BNC illustrate these non-discourse markers uses of *so*:

> (1) […] this wasn't possible then because **so** many women had been called up […].
> <div align="right">(BNC D8Y 63)</div>

> (2) […] like a saucepan with a a kettle that fitted on top **so** that you could boil your vegetables […]. (BNC D8Y 271)

Both these utterances show that when we use *so* as an adverb (here as an adverb of degree) or connector (here showing purpose), *so* cannot be omitted without changing the meaning of the utterances. Compare (1) and (2) with example (3):

> (3) **So** if anybody does patchwork knitting or makes blankets or anything for charity and they'd like to give me a ring any time, I could give you the pattern.
> <div align="right">(BNC D90 23)</div>

Example (3) shows that when *so* is used as a discourse marker (here to mark result), *so* can be omitted without changing the meaning of the utterance. This utterance can be perfectly understood without the use of *so*; *so* is rather used here to help the listener to interpret the message.

The general features of discourse markers presented in section 3.2 resonate with the features of the discourse marker *so*; it is associated with informal language use and most preferably used in speech, it is usually placed in initial position and as example (3) shows, it is optional in the sentence but helps to add extra meaning to the utterance.

Functions of *so*

One of the most common ways of describing *so* is that it marks result or consequence (Schiffrin 1987, 201). Müller (2005, 68) characterizes this function of *so* as textual, while Schiffrin (1987) and Buysse (2012) characterize *so* as ideational, since it helps the hearer to understand how two utterances or clauses relate to each other. Müller (2005) argues that while *so* is ideational, it functions at a textual level at the same time because it "indicates particular

relationships between propositional contents expressed in the narrative or discussion" (Müller 2005, 74). Therefore, I have chosen to label resultative *so* as textual when analyzing the functions of *so*.

The characterization of *so* as a discourse marker when marking result has been criticized since *so* in this context seems to have core meaning. Müller (2005) argues that the result or consequence is already implied in the message because we are able to understand the result based on our previous knowledge (2005, 72). This means that *so* is used by the speaker voluntarily to emphasize the result. Therefore, the message would still be understandable to the hearer even if we removed *so* from the utterance. This is illustrated in example (4):

(4) A new germ enters the body. Now there aren't enough 'soldier' cells
to beat the germ, **so** it multiplies. (BNC A01 34-35)

Example (4) shows that *so* is voluntarily used by the speaker to emphasize the result, and it can be replaced with an alternative expression, such as *and consequently*, without changing the meaning of the utterance.

So can serve other textual functions in an utterance. First of all, Schiffrin (1987) finds that one main function of *so* is to direct the topic back to the main idea of the conversation (1987, 193). This function of *so* can also be found in Müller's (2005, 68) and Buysse's (2012, 1767) studies, along with several other textual functions, such as summarizing, rewording, introducing an example or elaboration on the topic. Additionally, both Müller (2005) and Buysse (2012) find that *so* can be used by the speaker to introduce a new sequence in the discourse. *So* can be used by the speaker to either introduce a new topic or refer to a previous utterance or idea within the same turn (Buysse 2012, 1773). In her material, Müller (2005, 81) finds the function of *so* as a boundary marker, in this case between instructions and narrative.

The interpersonal functions of *so* have in common that they in some way are directed towards the hearer (Müller 2005, 82), to signal some type of interaction, action or relationship between speaker and hearer. *So* has an interpersonal function when the speaker uses *so* to indicate that he or she is going to continue speaking (Buysse 2012, 1770). Moreover, both Buysee (2012, 1769) and Müller (2005, 84) find that *so* can be used as a signal that the hearer can take over the turn. Buysse (2012) also suggests that *so* can be used to draw a conclusion. Some researchers do not separate the resultative *so* from the conclusive *so*; however, if we paraphrase conclusive *so* we would get "from state of affairs X I conclude the following: Y" (2012, 1768), while a resultative *so* could be paraphrased "state of affairs Y is the result/consequence of the state of affairs X" (2012, 1768). This shows that the resultative and

the conclusive *so* should be distinguished from each other. One important interpersonal function of *so* is that it introduces and marks speech acts: questions, requests and opinions. This function clearly shows the interactional nature of the discourse marker *so*. The textual and interpersonal functions of *so* are summarized in Table 2.

Table 2: Summary of discourse marker functions of *so* in previous research

- Mark result or consequence - Lead back to the main thread - Preface a summary - Preface an example - Mark transition - Reword/mark self-correction - Preface a new sequence - Preface a new section - Put an opinion into different words - Hold the floor - Induce action of hearer - Preface a conclusion - Preface speech acts: questions, requests and opinion.

Sources: **Schiffrin (1987), Müller (2005) and Buysse (2012)**

3.2.2 *Like*

There are many non-discourse marker functions of *like*, and some of them are presented below:

> (5) You've got to **like** the smell. (BNC FM3 225)
> (6) […] give them things **like** coffee and things **like** that […]. (BNC D8Y 396)
> (7) I mean w-- **like** I said early on […]. (BNC FYK 349)
> (8) […] by people who are of **like** mind […]. (BNC KB0 3681)

These examples illustrate some of the non-discourse functions of *like*: *like* as a verb (5), *like* as a preposition (6), *like* as a conjunction (7) and *like* as an adjective (8).

Like has a discourse marker function when it is used as an optional element in an utterance to express some kind of extra meaning or function and to organize speech. *Like* can occur in all positions in the utterance, but it normally occurs in initial or medial position. The discourse marker *like* has several different functions, one of them being a marker of "looseness" in speech (Andersen 1998, 152), illustrated in example (9):

> (9) I just normally buy **like** water bombs […]. (BNC KSW 771)

The speaker in (9) reduces his or her "commitment to the literal truth of his/her utterance" (Müller 2005, 210), which creates this looseness towards the message. Andersen (1998, 153) argues that the discourse function use of *like* can be interpreted as a marker of looseness whenever it is used in an utterance. In contrast to Andersen (1998), Müller (2005) finds that when *like* is used as a premodifier in a noun phrase (or before a verb phrase, adjective or adverb), it can be used by speakers, not only to distance themselves from the utterance, but also to put focus on the lexical item (2005, 220). The lexical item in the utterance may have some importance for the message implied in the utterance. Even though we can characterize *like* as being a marker of looseness and to mark lexical focus, it has the ability to serve several other functions in an utterance.

Functions of *like*

Müller (2005) characterizes all functions of *like* that she found in her study as having only a textual function since *like* does not "play a role in the interaction between speaker and hearer" (Müller 2005, 225). Both Müller (2005, 210) and Schourup (1985, 38) state that *like* is used by the speaker to mark an approximate number or quantity. This in turn supports the notion of *like* being a looseness marker, since *like* in this context "can be seen as a device available to speakers to provide for a loose fit between their chosen words and the conceptual material their words are meant to reflect" (Schourup 1985, 42). Furthermore, *like* can be used by the speaker to introduce an example, which makes *like* in this context semantically equivalent to 'for example' (Schourup 1985, 48). One other common use of *like* is *like* as a hesitator when it is used with other markers or words indicating hesitation (Müller 2005, 208). The speaker then uses *like* while searching for the right words or expression. Müller (2005) also finds that *like* can be used to introduce explanations: to make the information given more under-standable, or to repeat what has been said before or to reformulate the information given (2005, 219).

One major function of *like* is to introduce direct speech (Schourup 1985, 43; Müller 2005, 226), as illustrated in (10):

(10) someone else came round to her house she was **like** you know get off my yard.
(BNC G4W 212)

This function of *like* has not been characterized as a discourse marker in this present study since in this context, *like* is preceded by a verb which makes it syntactically bound to the

utterance and therefore cannot be removed without leaving the utterance incomplete. The functions of the discourse marker *like* in previous research are summarized in Table 3.

Table 3: Summary of discourse marker functions of *like* in previous research

- Looseness marker - Mark lexical focus - Mark approximate number or quantity - Introduce an example - Hesitator - Introduce an explanation

Sources: **Schourup (1985), Andersen (2000) and Müller (2005)**

3.2.3 *Actually*

The word *actually* is an adverb, but it has developed into a discourse marker as well (Aijmer 2002, 251). To distinguish between *actually* as an adverb and discourse marker, Aijmer (2002, 257-259) chooses to define *actually* as a discourse marker based on position. When *actually* occurs clause finally (11), utterance finally (12), utterance initially (13) or in a post head position (14), it has a discourse marker function:

> (11) Er one of my worst experiences **actually** was going to school […]. (BNC D90 280)
> (12) I wouldn't know **actually**. (BNC D91 78)
> (13) **Actually** some friends of mine were quite confused about […]. (BNC D97 68)
> (14) […] he's in court **actually** in the Birmingham area […]. (BNC JSN 146)

All these examples also show that when *actually* functions as a discourse marker, it is syntactically optional, and as previously mentioned, this is the most important distinguishing feature of discourse markers. These examples also show that *actually* has the ability to occur in all positions in the utterance.

How we interpret the meaning of *actually* depends on its use. When *actually* is used as a discourse marker, it expresses some kind of attitude toward an unexpected event (Aijmer 2002, 274), thus, it is usually referred to as an expectation marker. *Actually* is most frequently used in speech, but it is also commonly used in writing where the writers express their opinion on the topic (Aijmer 2002, 259), such as in argumentative writing.

Functions of *actually*

One of the main textual functions of *actually* is as marker of contrast and clarification. When *actually* is used in this way, it helps the speaker to create a contrast between a previous utterance and the current utterance, and it can be used for several purposes in the utterance,

such as to object, reformulate an utterance or to deny something (Aijmer 2002, 266). In this context, *actually* can be paraphrased as either 'but *actually*' (contrast) or 'no *actually*' (clarification) (Aijmer 2002, 265). Examples (15) and (16) illustrate these uses:

> (15) **Actually** just just quickly er I noticed on that list of your <pause> questionnaires that we got back a couple […]. (BNC D97 1807)

> (16) No, no **actually** I don't. (BNC FXX 164)

In (15), the speaker is marking a contrast between a previous utterance and the current: it seems as if the speaker has got new information about the questionnaires in the conversation. In (16), the speaker seems to regret the previous utterance and thereby clarifies his or her point of view by using *actually*. The contrastive *actually* can also be used by the speaker "to distance himself from the factuality of an earlier assertion […] and to express contrast with it (Aijmer 2002, 266).

    Actually can also be used in an utterance to emphasize the speaker's personal opinion by explaining or justifying something (Aijmer 2002, 269). It can also be used to introduce an elaboration. Example (17) illustrates these uses of *actually*:

> (17) Well, I mean **actually**, we wouldn't say that to him if he stuck something up in his front garden […]. (BNC KRL 422)

In example (17), *actually* is both used to emphasize the speaker's personal opinion that may be in contrast of what the other speaker has expressed, and at the same time elaborate on the topic of discussion.

    Even though *actually* is used to create a contrast, clarify or elaborate on something and express a personal opinion, *actually* "appear[s] to introduce repairs to the common ground" (Smith and Jucker 2000, 208). This suggests that *actually* does not only have a textual function, but also an interpersonal function: marking politeness in an utterance (Aijmer 2002, 272). When *actually* is used, it seems as if the speaker is trying to express their own opinion or thought in a politer and softer way, as shown in (18):

> (18) […] Yeah, I think they're about four sizes too big **actually**. (BNC KSV 5234)

When *actually* has an interpersonal function, it is usually placed in final position in the utterance (Aijmer 2002, 272). Table 4 (see page 19) summarizes the discourse marker functions of *actually* in previous research.

Table 4: Summary of discourse marker functions of *actually* in previous research

- Mark contrast - Preface a clarification - Emphasize speaker opinion - Preface an elaboration - Mark politeness

Sources: **Aijmer (2002) and Smith and Jucker (2000)**

3.2.4 *Anyway*

The non-discourse marker use of *anyway* is when it functions as an adverb, which can be divided into two sub-types, one equivalent to *besides* and one comparable to *nonetheless* (Ferrara 1997, 347). Compare examples (19) and (20):

(19) […] these were the only colours available **anyway**. (BNC D8Y 327)
(20) We bought the storage boxes **anyway**. (BNC D97 523)

In (19), the semantic meaning of *anyway* can be replaced with *besides* (besides, these were the only colours available), while in (20), *anyway* has the same meaning as *nonetheless* would have (nonetheless, we bought the storage boxes). If we remove *anyway* in example (19) and (20), the semantic meaning of the sentence would be altered. Example (21) illustrates *anyway* in a discourse marker context:

(21) **Anyway**, back to the point. (BNC D97 789)

Here, the speaker uses *anyway* to signal to the conversation partner(s) that the topic has got off track, and that the speaker wants to resume the earlier topic. However, in this context, *anyway* is optional and can be omitted without changing the meaning of the utterance. Ferrara (1997, 350) argues that the discourse marker *anyway* only occurs in initial position.

Functions of *anyway*

Anyway is used by the speaker to organize his or her speech. Therefore, it seems as if this marker only has a textual function. Ferrara (1997, 358) distinguishes between two different cases of *anyway* that are "triggered" by either the speaker or the hearer/listener: teller-triggered cases and listener-trigged cases. This means that *anyway* can be brought into the conversation based on what the speaker has uttered before, or by the hearer's saying or expression. Even if *anyway* is triggered by the speaker or the hearer, it is mainly used by the speaker to move the conversation forward in some way. The speaker can use *anyway* to lead

the conversation back to the main thread, either to manage self-digression or to regain control from the hearer (Ferrara 1997, 373). It can also be used to introduce a new topic, or to fill a pause, and when *anyway* collocates with verbs such as *think* and *believe* it is used by the speaker to introduce his or her own mental state at the time of the event (Ferrara 1997, 360), as illustrated in example (22):

(22) […] but **anyway** I think it was a superb night […]. (BNC J3T 230)

Table 5 summarizes the discourse marker functions of *anyway* in previous research.

Table 5: Summary of discourse marker functions of *anyway* in previous research

- Manage self-digression - Regain control from the hearer - Introduce a new topic - Pause filler - Introduce the speakers mental state

Source: **Ferrara (1997)**

3.2.5 *Well*

Except the use of *well* as a noun, the non-discourse marker functions of *well* are presented in (23), (24) and (25):

(23) The furniture was **well** designed […]. (BNC D8Y 316)
(24) And this style lent itself very **well** to uniform hats and caps. (BNC D8Y 412)
(25) Can I just way something else **as well**? (BNC D91 207)

In (23), *well* is an adverb, in (24) *well* is an adjective and in (25), *well* is part of an expression similar to 'in addition' (Müller 2005, 108). Example (26) shows that the word *well* also has a discourse marker function, since the meaning of the utterance would not change if we removed *well*:

(26) […] and you will find that your muscles will soon cooperate. **Well** I think we have to stop there for a little while because it's nine o'clock […].
(BNC D8Y 427-428)

Here, *well* is used by the speaker to mark transition in the discourse, to signal that the conversation or topic at hand has come to an end. *Well* has the ability to occur in all positions: initial, medial and final position. The discourse marker *well* has both a textual and interpersonal function.

Functions of *well*

Well's main function is to organize speech and mark transitions; thus it has a textual function. Depending on which context we find this discourse marker in, it can be used by the speaker to manage the discourse somehow: to conclude, to explain, to clarify, to justify, to reformulate and to introduce a new topic (Aijmer 2011, 236). It can also be used as a pause filler while searching for the right word or phrase or in a quotation (Müller 2005, 107).

Well can also have an interpersonal function, and is "described as a discourse marker signalling that what is said is not in line with expectations" (Aijmer 2011, 236). This is shown when *well* is used in the discourse to express some kind of disagreement with the previous utterance and also when the speaker is expressing an opinion. Müller (2005, 122) also mentions that *well* is used interpersonally when it prefaces an answer to a question, as displayed in (27):

> (27) Do you not got to the school's for suggestions?
> **Well** yes of course. (BNC D91 99-100)

Table 6 summarizes the discourse marker functions of *well* in previous research.

Table 6: Summary of the discourse marker functions of *well* in previous research

- Preface a conclusion
- Preface an explanation
- Preface a clarification
- Preface a justification
- Introduce a new topic
- Search for the right word/phrase
- Express an opinion
- Signal disagreement
- Preface an answer to a question

Sources: **Müller (2005) and Aijmer (2011)**

3.2.6 *You know*

The discourse marker *you know* is a common feature of conversations. *You know* only functions as a discourse marker when it is syntactically optional (Müller 2005, 157). Compare (28) and (29):

> (28) Do **you know** why you lost the Eastern Arts drama? (BNC D91 131)
> (29) […] my little fingers were like rolling pins **you know** and they were long […].
> (BNC D90 36)

If we remove *you know* from the utterance in (28), it would leave the utterance syntactically incomplete. If we do the same in (29), the sentence would still be syntactically complete and understandable. *You know* can occur in all syntactic positions in the utterance.

Functions of *you know*

The discourse marker *you know* has a large number of functions, both textual and interpersonal. Müller (2005, 147) mentions that this marker has been described to have up to 30 different functions. According to Müller (2005, 157), when *you know* has a textual function it usually takes part in the discourse as a pause filler while the speaker is searching for the right word or content, or to mark repairs. Furthermore, it can be used by the speaker to introduce an explanation, to mark that something is not so precise and when the speaker wants to introduce a quote (Müller 2005, 157). When *you know* has an interpersonal function, it tries to appeal to the hearer somehow, whether it is for understanding, acknowledgement or to mark reference to shared knowledge (Müller 2005, 157), or to monitor the hearer's understanding of the utterance (Fox Tree and Schrock 2002, 739). Fox Tree and Schrock (2002, 737) mention that *you know* can also be used to mark politeness: "[by] saying *you know* and leaving ideas less filled out, speakers can distance themselves from potentially face-threatening remarks and invite addressees' interpretations […]" (2002, 737). Table 7 summarizes the discourse marker functions of *you know* in previous research.

Table 7: Summary of discourse marker functions of *you know* in previous research

- Mark a search for the right word or content
- Mark false start and repair
- Mark approximation
- Introduce an explanation
- Introduce a quote
- Appeal for understanding
- Mark reference to shared knowledge
- "Imagine the scene"
- "See the implication"
- Acknowledge that the speaker is right
- Mark politeness

Sources: **Müller (2005) and Fox Tree and Schrock (2002)**

3.2.7 *I mean*

Like *you know*, *I mean* is also common in talk and may be even more common in talk where the speakers have the possibility to express their own opinion about the topic (Fox Tree and

Schrock 2002, 741). It only has a discourse marker function when it is syntactically optional. Compare examples (30) and (31):

> (30) And what **I mean** by that is […]. (BNC FUG 404)
> (31) **I mean** I know an awful lot of people […]. (BNC D91 183)

Example (31) shows the discourse marker function of *I mean*. In this context we can omit *I mean*. *I mean* can occur in all positions in the utterance (Fox Tree and Schrock 2002, 741).

Functions of *I mean*

I mean "focuses on the speaker's own adjustments in the production of his/her own talk" (Schiffrin 1987, 309). This means that *I mean* mainly has a textual function where it usually prefaces upcoming discourse such as explanations, clarifications, misinterpreted meanings, expansions of previous utterance and also to express the speaker's tone towards the message (Schiffrin 1987, 298) as illustrated in (32):

> (32) […] Community Service Volunteer placements involve things like looking after very severely handicapped people who are erm in higher education or something. […] **I mean** really severely handicapped so they really need […].
> (BNC HDY 744-746)

Example (32) shows that the speaker uses *I mean* to enhance the tone, in this case the seriousness, of the previous message. Even though *I mean* is mainly used to make transitions in the discourse, it can also have an interpersonal function when it is used by the speaker to instruct the hearer to continue attending to the prior utterance made (Schiffrin 1987, 310).

Table 8 summarizes the discourse marker functions of *I mean* in previous research.

Table 8: Summary of the discourse marker functions of *I mean* in previous research

- Mark upcoming modification
- Preface an explanation
- Preface a clarification
- Preface misinterpreted meaning
- Preface an expansion
- Express speaker tone
- Instruct the hearer to continue attending to the prior utterance

Source: **Schiffrin (1987)**

4 Method

This study aims at contributing to the discussion of whether the written language of Norwegian learners of English is influenced by oral language to a higher degree than the written language of native speakers of English and it also aims to describe how learners use discourse markers in their academic writing. To be able to compare these two groups, the International Corpus of Learner English (ICLE) and The Louvain Corpus of Native English Essays (LOCNESS) corpora will be the providers of data. These corpora will be described and discussed in Chapter 5. The method used in this study is the Contrastive Interlanguage Analysis (CIA) method. In the following sections in this chapter, corpora, learner corpora and the CIA method will be defined and discussed.

4.1 What is a corpus?

How do we define a corpus? Could any sample of texts be considered a corpus? The definitions below capture the essence of what a corpus is:

"A helluva lot of words, stored on a computer." (Leech, 1992, 106)

"A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research." (Sinclair 2005, 16)

"A collection of written or spoken material in machine-readable form, assembled for the purpose of linguistic research." (English Oxford Living Dictionaries)

"[…] the notion of "corpus" refers to a machine-readable collection of (spoken of written) texts that were produced in a natural communicative setting, and the collection of texts is compiled with the intention (1) to be representative and balanced with respect to a particular variety or register or genre and (2) to be analyzed linguistically." (Gries 2009, 7)

Based on these explanations and definitions, certain common features emerge: A corpus a) is (usually) a massive collection of texts that represents authentic language, b) which is consciously put together based on certain principles, c) which is stored in a digital format, d) and used for linguistic reserach purposes. Therefore, as Sinclair (2005) puts it: "The World Wide Web is not a corpus, […], an archive is not a corpus, […], a collection of citations is not a corpus, […], a text is not a corpus." (Sinclair 2005, 16).

4.1.1 Authenticity and representativeness

"The corpus builder should retain, as target notions, representativeness and balance. While these are not precisely definable and attainable goals, they must be used to guide the design of a corpus and the selection of its components" (Sinclair 2005, 10).

What Sinclair (2005, 10) suggests here is that balance and representativeness are important considerations for building a valuable corpus which is possible and desirable for researchers to use. Even though there are many variables to take into consideration in the corpus design, balance and representativeness should be guiding any corpus builder. How well the corpus sample represents the total population of interest is important for assessing the validity of the corpus. Representativeness is always a consideration when making use of corpus methods.

We have to consider both size and balance to assess representativeness. When a corpus is constructed, the designer has to consider how many samples are needed to make the corpus representative of the population of interest (size), whether the samples should consist of full texts or extracts, and the size of the samples (Nelson 2010, 57). However, there is no absolute answer to how large a corpus should be; it is the area of study and the purpose that should guide the corpus builder to the appropriate size (Nelson 2010, 57). Apart from these guidelines, the question of size seems to be a question which has no right answer. Balance is concerned with the proportion between different properties of the texts in the corpus. This concerns aspects such as register (written and spoken texts), as well as genre and production variables (gender, age, social class etc.).

The composition of the corpus in terms of balance and representativeness is crucially important for the possibility of generalizing any findings made on the basis of corpus research. The corpus is representative if the findings can be generalized (Clancy 2010, 86). Since balance and representativeness are important considerations when constructing a corpus, we as corpus users also have to take these notions into account in order to evaluate the validity of the corpus and the possible shortcomings of the material in the corpus (Johansson 2011, 119).

When assessing the validity of a corpus, both representativeness and authenticity have to be considered. Authenticity concerns the production of the language the corpus holds. The material in a corpus should be naturally occurring language which has been produced in an authentic communicative context. Sinclair (1996) defines naturally occurring language or

authentic data as "[…] material gathered from the genuine communications of people going about their normal business" (1996[3]). This suggests that language that has not been produced in a natural environment could not be considered possible material for a corpus. This will be further discussed in section 4.2.2.

The representativeness and authenticity of the two corpora used in this study will be evaluated in section 5.1.2 and 5.2.1.

4.1.2 Other considerations and limitations

Total accountability concerns the principle that we have to include all data relevant for our study, even if some instances are difficult to classify (McEnery & Hardie 2012, 252). The question is, to what extent do we get all examples of the phenomena/construction we searched for and to what extent are the results of our search relevant? Ball (1994) warns against uncritical use of corpora and mentions one of the most serious pitfalls while using corpora, "the recall problem" (1994, 295). The recall problem concerns the balance between recall and precision: how do we know that we get all the examples of the specific construction we searched for, and to what extent are all the results we get relevant for our study? (Ball 1994, 295). This means that if we widen our search, we would get many instances that are not relevant for our study. However, if we narrow our search we cannot be sure that we get all the examples of the item we want to study, since, for example, words may be misspelt. This is even more important to consider when searching for words or phrases in a learner corpus. We need to be aware of this in order to assure the validity of our results.

The development of corpus linguistics has expanded our understanding of language and created platforms which enable linguistic research to become much more accessible. We are able to access vast amount of data and find evidence for our research questions, and we have the possibility to analyze language more quantitatively and not only study language in isolation (Johansson 2011, 116). In spite of this, we cannot solely rely on corpus methods when we study language: it is sometimes necessary to analyze language without the aid of an electronic corpus.

---

[3] http://www.ilc.cnr.it/EAGLES96/corpustyp/node12.html

4.2 Corpora and second language research

"As far as I see, there is hardly a subdiscipline of linguistics, whether theoretical or applied, that cannot be enriched by the use of corpora" (Johansson 2011, 123).

This statement highlights the importance of the development of corpora, and like other disciplines of linguistics, second language research has also been enriched by the emergence of corpora, and most importantly, learner corpora which started to surface during the 1980s (Granger 2015, 7). With massive data available, researchers had the possibility to access new knowledge about learner language and interlanguages and thus supply insight to the field of second language research.

4.2.1 Learner corpora

Like any other corpus, a learner corpus is also a collection of texts which is consciously put together based on certain principles, stored in a digital format and used for linguistic purpose and research. The main difference between any other corpus and a learner corpus is that the material compiled consists of written or/and oral texts produced by learners of English. Another important difference concerns certain principles upon which the corpus is built. Interlanguage is different from native language in the way that it is influenced by other linguistic, situational and psychological features and "failure of control for these factors greatly limits the reliability of findings in learner language research." (Granger 2008, 263). Therefore, Granger (2008) suggests a corpus design, illustrated in Figure 1, which makes it easier to control these variables.
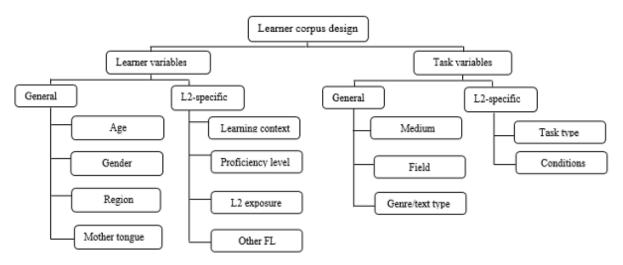

Figure 1. Learner corpus design as suggested by Granger (2008, 264) for attaining valid research results

Figure 1 (see page 27) illustrates the different variables learner corpus builders have to provide information about in the corpus design. If the corpus user has the information about the learners and the context in which the text or speech was produced, he or she will be able to attain more reliable and generalizable results. Figure 1 shows that the corpus designer should provide information about both general and L2-specific variables. The general variables should be part of any corpus design (Granger 2008, 264) (age, gender, region, mother tongue, medium, field and genre/text type), while the L2-specific learner variables should be included in a learner corpus design in order to provide the user with specific information about the learners (learning context, proficiency level, exposure to the target language, knowledge about other foreign languages, task type and conditions). The L2-specific task variables explain what kind of task the learners have performed when they produced the material for the corpus, such as argumentative writing, interviews and conversations, and the conditions explains under what circumstances the material was produced like time restrictions, use of reference tools and topic (Granger 2008, 265).

4.2.2 Learner material

As previously mentioned, the material in a corpus should represent natural language use, authentic material from people "going about their normal business". This creates an issue concerning learner data since learners usually do not use the target language as a way of communicating in their daily lives (Granger 2008, 261), but rather use the target language in specific situations such as communicating abroad, writing essays in school or communicating with other people who do not speak the native language. However, when texts or speech are compiled for the specific purpose of corpus building, there are certain degrees of naturalness concerning the tasks that the learners engage in, which range from activities which are exclusively elicitation exercises (reading out loud), to activities where learners produce the target language on their own (Granger 2008, 261), such as casual conversations or essay writing. In order to refer to a collection of learner speech or texts as a 'corpus', the tasks should elicit language that the learners have produced on their own (Granger 2008, 261). The learner data authenticity in the ICLE corpus will be discussed in section 5.1.2.

4.2.3 The learners in learner corpora

Another concern is which data we should consider to be learner data:

"The language learners whose language is covered by learner corpora are to be understood as foreign language learners, i.e. speakers who learn a language which is neither their first language nor an institutionalized additional language in the country where they live." (Granger 2008, 260).

This definition may seem straightforward; however, Granger (2008, 260) mentions that this definition may not be as applicable to the English language as to other languages, since English is a widespread language which might be used for daily communication by non-native speakers even though it is not an official language in the country. This would include the use of English as a lingua franca, when non-native speakers communicate in English with people with a different native language (Seidlhofer 2004, 211). If we wish to accept Granger's definition, it would eliminate these groups and those groups who define English as an official second or additional language (Seidlhofer 2004, 224). This suggests that the definition of what English learner data consists of is rather complex.

As Figure 1 (see page 27) illustrates, L2-specific variables, such as L2 exposure, proficiency and learning context, are important factors to consider when designing a learner corpus. Since these variables are somewhat dependent on the status of English in the country where the learners come from, we should discuss the status of English when we define and describe the learners in the corpus we are researching. In addition, our research purpose and focus would most likely depend upon what type of status English has for the learners (Seidlhofer 2004, 224).

4.3 Contrastive Interlanguage Analysis

The emergence of learner corpora created new possibilities for researching language. This generated a need for new methods, in order to retrieve knowledge about learner language. One approach for investigating learner data is the Contrastive Interlanguage Analysis (CIA) method, which provides knowledge on the differences between learner and native speaker performance. With this method, the researcher is able to compare learner production to data produced by native speakers of a particular language of interest. It is also possible to compare different interlanguages of the same language, which can be of interest if the researcher wants

to retrieve information about how generalizable certain interlanguage features are across different learner groups (Granger 2009, 18).

With the CIA method, we now have the possibility to study other types of linguistic phenomena than plain errors in interlanguages. We are able to study overuse and underuse of certain linguistic features connected to lexis and discourse, and therefore, this method is suitable for comparing advanced interlanguage to native speaker production. Consequently, CIA studies have provided the field of second language research with new insight on advanced interlanguage (Granger 2015, 11).

CIA has been subject to criticism, and it especially concerns the comparison between learner language and native language, where the method has been accused of "comparative fallacy": "comparing learner language to a native speaker norm and thus failing to analyze interlanguage in its own right" (Granger 2009, 18). Although this is valid criticism, the method has proven very important for uncovering features of learner language which were not known or studied before, and one can argue that when we study interlanguage of any sort, we study this interlanguage with the notion of a target language (Granger 2009, 19). Even though this criticism does not weigh up for all the possibilities that CIA provides for second language research, Granger (2015) points out that this debate is a good reminder that interlanguages should be studied in their own right, i.e. without the comparison to a native speaker norm (2015, 14).

Another criticism is the use of the term 'native speaker language' used in CIA. Granger (2015) introduces an alternative term: 'Reference Language Varieties' (RLV), which can be understood as a more neutral term which entails the possibility of several different varieties of the same native language rather than the thought that there is only one standard norm (2015, 17). Granger (2015) also proposes the term 'Interlanguage Variety' (IV), where the addition of 'variety' brings into focus the fact that interlanguages are highly variable (2015, 17). The terms 'native language', 'learner language' and 'interlanguage' are used in this paper. However, even if these are the terms used, this paper recognizes the fact that native languages have different varieties and also that interlanguages are variable.

5   Material

In this chapter, the two corpora used in this study, LOCNESS and ICLE-NO will be outlined in terms of content, followed by a discussion of the corpora's authenticity, representativeness and comparability. Furthermore, this chapter explains how the data was extracted from the corpora and gives a presentation of the framework used for classifying the material.

5.1 ICLE and ICLE-NO

In the ICLE corpus we find essays written by learners of English with a proficiency level of higher intermediate to advanced level. The corpus consists of several subcorpora in which groups of learners share the same native language. This corpus project, initiated by Professor Sylviane Granger of the Université catholique de Louvain, was the first of its kind (Johansson 2008, 115). ICLE provides the possibility to compare different types of interlanguages to a native language, but it also offers the possibility to compare the interlanguage of learners from different first language backgrounds. All the different subcorpora have to follow specific collection guidelines to ensure comparability between the different subcorpora.

The Norwegian subcorpus of ICLE is referred to as ICLE-NO. This subcorpus consists of roughly 212,000 words, and most of the texts collected are written by Norwegian students in their first year who attend English courses at the university (Johansson 2008, 116). The ICLE-NO follows the same corpus collection guidelines as the other subcorpora in ICLE.

5.1.1 The learners in ICLE-NO

The learners in ICLE-NO can be characterized as advanced learners of English, even if they are novice writers. Although English does not have the official status of a second language in Norway, English is taught already from first grade and is one of the core subjects throughout the students' entire education. This means that Norwegian students have been exposed to the English language for a long period of time both through education and also through other channels such as the internet, television and movies. However, we have to remember that ICLE-NO was collected in the 1990s which means that the input from media was less extensive compared to the input learners get today. Even so, the Norwegian learners of English in the ICLE-NO corpus is a suitable group to compare to native English speakers when trying to answer this study's research questions since they are considered advanced learners.

5.1.2 Authenticity and representativeness

The material in ICLE-NO consists of texts produced for the specific purpose of corpus building. One can argue that this is less authentic material since the learners have been asked to write these texts for this specific purpose, and that they have not been writing while they were "going about their normal business". However, the material in ICLE-NO consists of texts written by learners who produce English on their own, thus the material can be characterized as being natural to a high degree. In terms of learner production, this may be the most authentic production we can collect.

The corpus collection guidelines are designed to create valid and representative data. The corpus builders have to request students to fill in a learner profile and they have to collect the right type of material (essays: argumentative or literary (no more than 25% of the corpus can consist of literary texts) (Corpus Collection Guidelines). These guidelines have to be followed by the corpus builders to ensure valid and representative data which can be used to draw general conclusions about the specific group we want to study.

Even though the material in the corpus can be defined as authentic and representative, we always have to consider the limitation of the corpus size: we cannot be certain that the sample is generalizable to the entire population. However, when the material is characterized as authentic and representative, we can make general assumptions about the population and it certainly can provide insight on the topic.

5.2 LOCNESS

The Louvain Corpus of Native English Essays is a corpus that contains material written by native speakers of English that are novice writers. The corpus holds argumentative and literary essays written by American and British University students from all over Britain and the United States, and also argumentative essays written by British A-level students. The essays in LOCNESS were produced under different circumstances. Some essays were produced in an exam situation while some were produced during a longer period of time. Some essays were written with the assistance of reference tools, while others were written without this type of aid. Nine students speak another language at home apart from English (LOCNESS description). The rest of the texts are written by students who only have English as their native language.

5.2.1 Authenticity and representativeness

The material in LOCNESS may be referred to as 'naturally occurring data' since the texts were collected from students 'going about their normal business' at the university. In other words, the material in LOCNESS can be characterized as authentic material. The entire LOCNESS corpus contains 324,304 words of native speaker production, and the texts that are represented consist of full text samples. All texts samples have been thoroughly described in the meta data according to different variables such as total number of words, essay topic, situational features, additional native language of writer and reference tools. This controlled form of corpus design plays a part in creating valid and representative material. As previously mentioned, we always have to take into account that the sample may not be generalizable to the entire population, but if the material is authentic and representative we can at least make general assumptions about the entire population.

5.3 Comparability

LOCNESS was compiled to function as a reference corpus to ICLE (Hasselgård and Johansson 2011, 38), and as in many other research projects, the LOCNESS corpus has been used as a reference corpus to ICLE in this study. Several considerations have to be taken into account when we choose a suitable native reference corpus, such as register, text type, age and proficiency of the contributors. In this case, both ICLE-NO and LOCNESS hold argumentative and literary essays, the students are about the same age and they are novice writers, which means that LOCNESS is more favorable to use compared to general native corpora (Granger 2015, 17). Even though the LOCNESS corpus is the preferred use of reference corpus to ICLE-NO, it does not provide as much information about its writers and situational features as the ICLE-NO corpus does and the texts in LOCNESS are more diverse in terms of content and its writers (some writers are defined as more advanced) (Hasselgård and Johansson 2011, 38). We should take these factors into consideration when we compare the ICLE-NO to LOCNESS. We also have to remember that the reference native speaker corpus only gives us a tool for measuring the standard of learner performance. However, the reference corpus, in this case LOCNESS, may not be a standard the learner should strive for: "[t]he LOCNESS is a reference corpus, not a norm for EFL learners" (Granger 2015, 18).

5.4 Extraction of the material

The material used in this study has been retrieved using the Concord function in WordSmith Tools 6 (Scott 2012). The material from LOCNESS contains 324,043 words and the material from ICLE-NO contains 212,005 words. Both these numbers were retrieved using the WordList function in WordSmith Tools 6 (Scott 2012). Since I have used WS to extract the material, I have not been able to control or sort the material, thus all texts from LOCNESS and ICLE-NO have been included in the study. The search strings used were *so, like, anyway, well, you know, I mean* and *actually.* The output of the search strings was manually sorted and all instances that were not defined as a discourse marker according to the features presented in 3.2 were discarded. Thereafter, the relative frequency of the discourse markers was calculated. Lastly, *so*, *like*, *actually*, *anyway*, *well*, *you know* and *I mean* were classified according to their functional features in the sentence. Since this project is based on a pre-study (Johnsson 2017), the material in the pre-study for the discourse marker *so* and *well* has also been used in this project.

5.5 Framework of classification

The framework of classification for this study is created on the basis of general previous research on discourse markers, and most importantly built on previous research of *so*, *like*, *actually*, *anyway*, *well*, *you know* and *I mean*. First of all, I have distinguished all instances of the words/phrases *so*, *like*, *actually*, *anyway*, *well*, *you know* and *I mean* from non-discourse marker uses. This classification is based on the features presented in section 3.2. The most important factor for determining if a word or phrase is a discourse marker or not, has been if this word or phrase is syntactically optional in the sentence. Thereafter, all instances of discourse markers have been categorized in terms of their syntactic position in the sentence. Lastly, all discourse markers have been assigned one or more pragmatic function. Some discourse markers are multifunctional; they function both at a textual and an interpersonal level. However, all discourse markers organize the discourse in some way (thus they have a textual function) and therefore, if the marker both has a textual and interpersonal function, I have assigned the marker an interactional function. Not all of the functions of the selected discourse markers presented in sections 3.2.1–3.2.7 were found in the material from ICLE-NO and LOCNESS, and therefore, the framework of classification of this study (see Table 9, page 35) does not include all functions. Moreover, a few other functions than what is

presented in sections 3.2.1–3.2.7 were found in my material. These have been added to the classification framework. The framework for classifying the discourse markers' syntactic position and function for this study is presented in Table 9.

Table 9: Framework of classification: position and semantic function

Syntactic position	
	Initial position
	- Clause initially
	Medial position
	- Pre head position
	- Postverbal position
	Final position
	- Clause finally
Interpersonal functions	Preface a request
	Preface an opinion
	Preface an answer to a question
	Mark politeness/common ground
	Mark reference to shared knowledge
	Instruct the hearer to continue attending to the prior utterance
	Acknowledge that the speaker is right
Textual functions	
	Preface a clarification
	Preface an elaboration
	Preface a conclusion
	Preface an explanation
	Preface an expansion
	Preface an example
	Preface a justification
	Mark contrast
	Mark result or consequence
	Mark lexical focus
	Mark transition
	Emphasize speaker opinion
	Manage self-digression
	Continue the discussion
	Continue an opinion
	Express speaker tone
	Lead back to the main thread
	Search for the right word/phrase

6 Results and analysis

The following chapter presents the results of this study and the analysis of the selected discourse markers *so*, *like*, *actually*, *anyway*, *well*, *you know* and *I mean*. The chapter is divided into two parts. Section 6.1 presents the results of the quantitative analysis and section 6.2 presents the qualitative analysis of the study. Section 6.3 provides a discussion of the most important and interesting findings in the material.

6.1 Quantitative analysis of discourse markers in Norwegian learner writing compared to native writing

This section presents the results of the quantitative analysis of discourse markers in ICLE-NO and LOCNESS. The quantitative analysis provides an overview of the frequency of the selected discourse markers, the tendency of their position and what their main functions are. Therefore, section 6.1 is divided into three separate parts: frequency, position and function. The quantitative analysis reveals the differences between the Norwegian learners in ICLE-NO and the native novice writers in LOCNESS in terms of these categories, and adds to the discussion of whether learners of English are aware of register and genre differences when writing in English. The results and findings presented in this section will be further investigated in the qualitative analysis and thereafter discussed.

6.1.1 Frequency

The frequency analysis of the selected discourse markers in ICLE-NO and LOCNESS is presented in Table 10 (see page 37). The results presented in Table 10 tell us that the selected discourse markers are used both by the learners in ICLE-NO and the native novice writers in LOCNESS. However, they are more frequently used by the learners in ICLE-NO. These results suggest that the selected discourse markers are overrepresented in the ICLE-NO corpus compared to the LOCNESS corpus.

6.1.2 Position

Table 11 (see page 37) presents the total number of instances of discourse markers in each syntactic position found in the material and their percentages of the total number of instances.

Table 10: Raw frequency and relative frequency per 10,000 words of *so, like, actually, anyway, well, you know* **and** *I mean* **in ICLE-NO and LOCNESS**

	ICLE-NO		LOCNESS	
	Raw frequency	Relative frequency	Raw frequency	Relative frequency
So	205	9.67	248	7.65
Well	39	1.84	15	0.46
Actually	10	0.47	4	0.12
Anyway	9	0.42	0	0
I mean	7	0.33	2	0.06
You know	3	0.14	0	0
Like	1	0.04	0	0.00
Total	274	12.91	269	8.29

Source: **Data from the ICLE-NO and LOCNESS**

Table 11: Raw frequencies and percentages of the position of *so, like, actually, anyway, well, you know* **and** *I mean* **in ICLE-NO and LOCNESS**

	ICLE-NO		LOCNESS	
	Raw frequency	%	Raw frequency	%
Initial position Clause initially	270	98.54	269	100
Medial position Pre head position	1	0.36	0	0
Final position Clause finally	3	1.10	0	0
Total	274	100	269	100

Source: **Data from the ICLE-NO and LOCNESS**

Table 11 shows that almost all the selected discourse markers in ICLE-NO occur in initial position (clause initially). A total number of three instances in ICLE-NO (1.10%) occur in final position (clause finally), and these are the discourse markers *actually* (1), *I mean* (1) and *you know* (1). One instance occurs in medial position: *like.* In LOCNESS, all of the selected discourse markers found occur in initial position. The syntactic position of *anyway*, *like* and

you know cannot be compared between the two corpora since there are no occurrences of these markers in LOCNESS.

6.1.3 Functions

All discourse markers in this study have been categorized as having a textual or interpersonal function. Table 12 presents the total number of discourse markers that have been assigned either a textual or interpersonal function in ICLE-NO and LOCNESS.

Table 12: Raw frequencies of the total number of interpersonal and textual functions in ICLE-NO and LOCNESS

	ICLE-NO	LOCNESS
Interpersonal functions	105	56
Textual functions	169	213
Total number of instances	274	269

Source: **Data from the ICLE-NO and LOCNESS**

Figure 2 further illustrates the distribution between the main functional categories in each corpus, and also the differences in terms of distribution between ICLE-NO and LOCNESS.
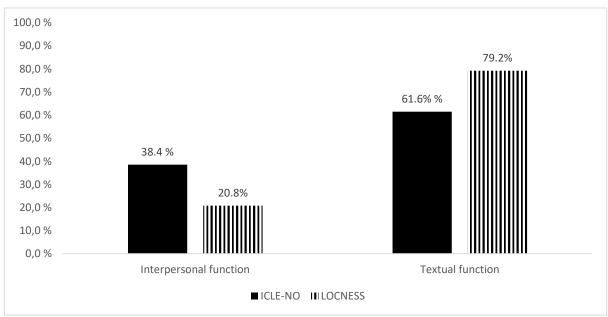

Figure 2. Illustration of the distribution between the textual and interpersonal functions compared between ICLE-NO and LOCNESS

Figure 2 shows that the textual functions of the selected discourse markers are more frequently used compared to the interpersonal function in both ICLE-NO (61.6% compared to 38.4%) and LOCNESS (79.2% compared to 20.8%). However, there are differences in terms of the distribution of the main functional categories between the two corpora. The Norwegian learners in ICLE-NO use the discourse markers with an interpersonal function more often than the writers in LOCNESS (38.4% in ICLE-NO compared to 20.8% in LOCNESS).

6.2 Qualitative analysis of discourse markers in Norwegian learner writing compared to native writing

This section presents the qualitative analysis of the discourse markers in the study. The quantitative analysis revealed differences in terms of frequency, position and main function between ICLE-NO and LOCNESS. The qualitative analysis further describes each discourse marker's functions in written discourse in ICLE-NO and LOCNESS, and at the same time highlights the differences in terms of use between the two corpora. All instances of each discourse marker have been thoroughly analyzed and all discourse marker functions found in the material are presented along with examples to illustrate their function in the sentence. Even though the main focus in this qualitative part is on the different functions of each marker, the frequency and position will also be commented on. This section aims to provide further insight on how learners of English use discourse markers in writing. The findings of the qualitative analysis will be discussed in section 6.4.

6.2.1 *So*

As shown in Table 10 (see page 37), *so* is the most frequent discourse marker in both ICLE-NO and LOCNESS compared to the other markers investigated in this study. Even though *so* is a common discourse marker in both corpora, *so* is more frequent in ICLE-NO, where the marker *so* has a relative frequency of 9.67 compared to 7.65 in LOCNESS. All instances of *so* in both ICLE-NO and LOCNESS are positioned clause initially. This corresponds with the fact that discourse markers are usually placed in an initial position. According to Müller (2005), *so* can occur at the end of an utterance to imply a result that can be understood by the hearer even if the speaker does not explicitly state the result (2005, 84). However, this may be a function of *so* that is restricted to oral discourse and may be the reason why there were no instances in the material of clause final *so*. *So* had many different functions in the material. The functions found are presented in Table 13 (see page 40).

Table 13: Raw frequency and percentage of the functions of *so* in ICLE-NO and LOCNESS

	ICLE-NO		LOCNESS	
	Raw frequency	%	Raw frequency	%
Interpersonal functions				
Preface a request	4	1.95	1	0.40
Preface an opinion	29	14.15	39	15.73
Preface a question	39	19.03	5	2.02
Total	72	35.13	45	18.15
Textual functions				
Mark result or consequence	44	21.46	78	31.45
Preface a conclusion	49	23.90	92	37.10
Mark transition	12	5.85	23	9.27
Preface an example	18	8.78	9	3.63
Lead back to the main thread	10	4.88	1	0.40
Total	133	64.87	203	81.85
Total	205	100.00	248	100.00

Source: **Data from ICLE-NO and LOCNESS**

Figure 3 further illustrates the distribution between the interpersonal and textual functions of *so* in each corpus, and also the differences in terms of distribution of the functions between ICLE-NO and LOCNESS.
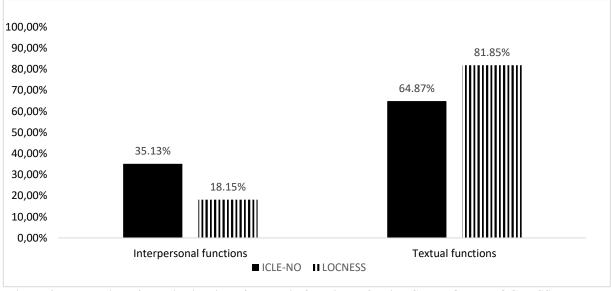

Figure 3. Illustration of the distribution of the main functions of *so* in ICLE-NO and LOCNESS.
*Source***: Data from ICLE-NO and LOCNESS**

Both textual and interpersonal functions were found in the corpus material. As Table 13 (see page 40) and Figure 3 (see page 40) show, the textual function of *so* is more common than the interpersonal function in both corpora. If we compare the two corpora, the interpersonal function of *so* is more frequently used in ICLE-NO than in LOCNESS.

As previously mentioned in section 3.2.1, *so* is usually described as a marker of result. Therefore, it may not be surprising that this is the most common use of the marker *so* in both ICLE-NO and LOCNESS. One other function that is recurrent in both corpora is *so* as a preface to a conclusion. In section 3.2.1, it was stated that some researchers do not separate the resultative *so* from the conclusive *so*. Examples (33) and (34) show the resultative *so* and the conclusive *so* and illustrate the difference between the two functions:

> (33) For example, she is a member of the Delta sorority and they collect elephants, **so** I've bought her different elephant pins to enhance her collection.
> <div align="right">(ICLE-US[4]-SCU-0007.2)</div>

> (34) A person will always feel guilty for what he has done and if he gets caught it is even worse for him. **So** crime does not pay. (ICLE-NO-AC-0023.1)

In example (33) the writer uses *so* to mark result; the fact that 'I've bought her different elephant pins' is the result of the fact that 'she is a member of the Delta sorority and they collect elephants'. In example (34) the writer uses *so* to introduce a conclusion: based on the fact that '[a] person will always feel guilty for what he has done', I conclude the following: that 'crime does not pay'.

In some instances in both ICLE-NO and LOCNESS, *so* is only used to mark transition in the text:

> (35) But Florida State and Notre Dame were the two teams getting all the hype and recognition to be playing for the national championship and that's because they had played one of the greatest games in college football history during the regular season. **So** on New Year's Day […]. (ICLE-US-SCU-0002.3)

Here, the writer uses *so* to continue the narrative in the text, and it is therefore used here as a marker of transition. Moreover, there are two other textual functions in the material:

> (36) As well, the fact that so many people (especially in the US) have television sets means that everybody (well, at least everybody who watches) receives the same inflow of information & ideas. **So**, for example, people in Spain can be informed about how people in California or Japan speak & act […]. (ICLE-US-MICH-0040.1)

---

[4] LOCNESS

> (37) In today's modern society we live after the principe that all men are equal." […]. This is true, to some extent. […]. **So** to the question of how true the statement […] is. (ICLE-NO-HE-0007.1)

In (36) *so* is used to preface an example, while in (37), the writer uses *so* to lead the discourse back to the main thread of the text, or here, the main idea.

As Table 13 (see page 40) shows, three different interpersonal functions were found in the material from ICLE-NO and LOCNESS. These different interpersonal functions are illustrated in the examples below:

> (38) Crime does not pay? Huh. It has paid for my entire apartment and education. I have wooed so many girls by taking them to the most fancy places just because I ripped off a car the day before **so** don't come here and lecture me about moral.
> (ICLE-NO-UO-0048.1)

> (39) Each of these characters find their personal strength in the defiance of naturalism. **So** I say, yes, naturalism is a prominent idea of ethnic American literature. (ICLE-US-PRB-0005.1)

> (40) **So**, what is a good job? (ICLE-NO-UO-0045.1)

> (41) **So** why does humanity still refuse to pay heed? (ICLE-NO-HO-0041.1)

In (38), *so* is used by the writer to preface a request which is directly addressed to the reader of the text, while in (39), *so* is used to preface an opinion that the writer expresses on the basis of the previous statement. The last interpersonal function of *so*, which is illustrated in both (40) and (41) is when *so* prefaces questions. It is important to note that these three uses of *so* are not only interpersonal, but have a textual function as well. They are either resultative or conclusive, since *so* here also refers back to the previous part of the discourse (Müller 2005, 82). This shows the multifunctional nature of *so*. Since these statements are directed towards the reader, they have been classified as interpersonal rather than textual.

In conclusion, the differences between the two corpora in terms of the textual functions of *so* are quite small. All functions found in the ICLE-NO corpus were also found in LOCNESS. However, there are some differences in terms of how frequent the different functions are. Table 13 (see page 40) shows that the novice native writers in LOCNESS use *so* more often with a resultative and conclusive function compared to the learners in ICLE-NO. Also, there is a slight difference between the corpora in terms of *so* marking transition in the text, where it occurs in 9.27% of the instances in LOCNESS and in 5.85% of the instances in ICLE-NO. The functions that are more frequent in ICLE-NO compared to LOCNESS are

so when prefacing an example or *so* leading back to the main thread of the text.

In terms of the interpersonal functions of *so*, the same functions were found in both ICLE-NO and LOCNESS. The difference lies in the proportion of the functions between the two corpora. There is only a small difference between the corpora in terms of *so* prefacing requests and opinions, while there is a greater difference between the corpora in terms of *so* prefacing questions. In LOCNESS 2.10% (5 occurrences) of the instances of *so* were prefacing questions, while in ICLE-NO, 19.02% (39 occurrences) of the instances had this function. The most prominent difference between the corpora is the use of interpersonal functions, where the learner writers use *so* with an interpersonal function more often than the novice native writers in LOCNESS.

6.2.2 *Like*

There was only one instance of the discourse marker *like* in the material. This instance was found in ICLE-NO:

(42) When I mention untrue violence, I mean **like** movies. (ICLE-NO-AC-0009.1)

Example (42) shows that the writer uses the marker *like* to mark lexical focus. By doing this, the speaker is directing the reader's focus towards the word *movie*, and thereby implying that this word is important for the statement made. The fact that there was only one instance found in the material suggests that both the learners in ICLE-NO and native novice speakers in LOCNESS may be aware of the informal association of the discourse marker *like*.

6.2.3 *Actually*

Actually occurs ten times in ICLE-NO and four times in LOCNESS. There is only one instance where *actually* is placed clause finally in ICLE-NO. All other instances of *actually* in both ICLE-NO and LOCNESS occur clause initially. Nine out of ten instances of *actually* in ICLE-NO are textual and one interpersonal, while all occurrences in LOCNESS are textual. The different functions of *actually* found in the material are presented in Table 14 (see page 44).

Table 14: Raw frequencies of the functions of *actually* in ICLE-NO and LOCNESS

	ICLE-NO	LOCNESS
Interpersonal functions		
Mark politeness/common ground	1	0
Textual functions		
Mark contrast	4	4
Emphasize speaker's opinion	5	0
Total number of instances	10	4

Source: **Data from ICLE-NO and LOCNESS**

Only one of the instances of *actually* in the material has an interpersonal function:

> (43) I think we have even more place for dreaming and imagination than before **actually**, only that the imagination is on another level. (ICLE-NO-HE-0005.1)

As Aijmer (2002) points out, when *actually* occurs clause finally (as illustrated in example (43)), it usually focuses on the social relationship between the speaker and hearer (2002, 258). In (43), *actually* is used by the writer to decrease the assertiveness of the writer's thoughts, and thereby the writer is trying to establish common ground with the reader.

Table 14 tells us that a total of eight instances of *actually* mark contrast, four in ICLE-NO and four in LOCNESS. This function is illustrated in examples (44), (45) and (46):

> (44) Yoda, that little green guy from Star Wars? **Actually** no, not yoda, but yoga.
> (ICLE-US-MRQ 0005.1)

> (45) Even though the pharmaceutical industry argues that medical pricing boards would raise prices and eliminate competition between companies, **actually** the opposite seems to be true. (ICLE US MRQ 0010.1)

> (46) Because of the theoretic exams, we are very concerned about learning the theoretic stuff. We are less concerned about learning the teaching methods because we want be asked for it at an exam. Or **actually**, we have had questions about how to teach at some written exams […]. (ICLE-NO-HO-0001.1)

These examples show that the writers use *actually* as a preface to a new clause to mark a contrast with a previous statement. In these examples where *actually* marks contrast, it clearly also indicates that this new upcoming statement is unexpected.

All instances of *actually* in LOCNESS mark contrast. There is one other textual

44

function represented in ICLE-NO: *actually* is used to emphasize the speaker's opinion. This function is illustrated in examples (47), (48) and (49):

> (47) Not all people like the taste of food in the morning. **Actually** a lot of people hate to eat right after they have gotten out of bed. (ICLE-NO-OS-0026.1)

> (48) We all hear about the dangers of global warming. Global warming is not dangerous in it self, **actually** we need it to survive! (ICLE-NO-UO-0012.1)

> (49) Furthermore, we have to listen to our children who are our professional dreamers. We need to find the child in ourselves. **Actually**, we should study the young ones closely and look at how they create their words. (ICLE-NO-OS-0004.1)

These examples show the opposite of the function of creating a common ground with the reader by using *actually* to decrease the assertiveness of the expression. In examples (47), (48) and (49) *actually* is used by the writers to enhance and support their statement. In these cases, they mark a further assertiveness of their previous statement. As previously mentioned, it may not be surprising to find *actually* in argumentative writing since *actually* can be used to support the writer's statements or thoughts. This could explain the occurrences of *actually* in ICLE-NO. However, *actually* as an emphasizer was not found in LOCNESS.

6.2.4 *Anyway*

There were nine instances of the discourse marker *anyway* in ICLE-NO, while no instances of *anyway* were found in the material from LOCNESS. All instances of the discourse marker *anyway* were placed in initial position (clause initially), which corresponds to Ferrara's (1997, 350) conclusion; that the discourse marker *anyway* only occurs in initial position. As previously mentioned, *anyway* is a textual marker, and all the instances of *anyway* in the material had a textual function. The different textual functions of *anyway* are presented in Table 15.

Table 15: Raw frequencies of the functions of *anyway* in ICLE-NO and LOCNESS

	ICLE-NO	LOCNESS
Textual functions		
Manage self-digression	2	0
Continue the discussion	6	0
Introduce a new topic	1	0
Total number of instances	9	0

Source: **Data from ICLE-NO and LOCNESS**

As presented earlier in section 3.2.4, using *anyway* to manage self-digression is a common use of this marker. Two instances of this function were also found in the ICLE-NO material, as illustrated in examples (50) and (51):

> (50) The last year we have no practice at all. From this it doesn't sound like it prepare the students for the real world. Compared with the students studying to be nurses, they have 2 months of practice every year, I believe. But **anyway**, the way the practice period for the students studying to be teachers is made […].
> <div align="right">(ICLE-NO-HO-0011.1)</div>

> (51) I think it was last Monday…I came home from school, completely aware of the fact that this day would be a boring day, with a lot to read, a meeting a work and by the way my room looked like a…well, I don't think a messy place would cover the it… **Anyway**, as I opened the mailbox […]. (ICLE-NO-OS-0015.1)

Both these examples show that the writer uses *anyway* as a way to manage the writer's own digression; as a way to return to the main point of the discussion (50) or to return to the main narrative of his or her story (51).

Another function of *anyway* is to introduce a new topic in the discourse. There is one instance in the material which displays this function:

> (52) They took part in a simple life, during their youth. The agriculture was based upon very simpel equipment's, and a great deal of the population lived in poverty. For those who experienced such a time, the revolution of science technology and industrialisation, must have been hard to handle. **Anyway**, from my point of view, I will say there is a great space for both dreaming and imaginations in our lives. (ICLE-NO-UO-0059.1)

In example (52), *anyway* is used as a linking word, to introduce a new topic, or in this case a conclusion. There are six other instances in the ICLE-NO material that show the characteristics of this function of *anyway* as they introduce something new in the discourse. However, these instances seem to introduce a new sequence rather than a new topic, and therefore, the discourse still seems to focus on the same topic. Consequently, I have chosen to name this function 'Continue the discussion' and add this to the textual functions of *anyway*. This function is illustrated in examples (53), (54) and (55) below:

> (53) This was when I discovered that boys in their 20's DRINK, and thus couldnt care less that your face looked like …….. well, something very strange, **anyway**, and that you were on the heavier side of the ricki lake […]. (ICLE-NO-UO-0064.1)

> (54) My practice teacher was a lot younger than the two first ones, and she had only been working as a teacher for four years. I found her more open-minded and not as restrictive as the other practice teachers I've had. She gave us quite a free hand

to do what we wanted to do. This could also be because we were 3rd year student, with at least some experience. **Anyway**, I relate it more to the fact that she was younger and not yet stuck […]. (ICLE-NO-HO-00006.1)

(55) This makes us busy creatures. We always have to be available. We're either on the Internet or on the phone. Quite stressful, and most of the time we don't even take notice. **Anyway**, before we were probably a bit more down to earth.
(ICLE-NO-UO-0080.1)

Discourse markers serve a function in the discourse, and as presented in the examples above, the writers use *anyway* as a linking word to organize the discourse, as also shown in example (52). Even though discourse markers are voluntary in discourse, they still serve a purpose, namely to help guide the reader though the discourse. However, the marker *anyway* in examples (53), (54) and (55) seems to be excessive and futile for the discourse. The question is whether these writers know how *anyway* helps to organize the discourse, and how they should use it. Since there are no instances of *anyway* in LOCNESS, we cannot find out if this function is also present in native novice writing.

6.2.5 *Well*

There are 39 instances of *well* in ICLE-NO and 15 in LOCNESS. Even though *well* can occur in all positions, *well* only occurs clause initially in the material from both corpora. *Well* is mainly a textual marker, and in those instances where *well* has been classified as interpersonal, *well* also has a textual function, and thus it is multifunctional. However, when *well* has been classified as interpersonal, the interpersonal function is the main function of the marker. Table 16 (see page 48) presents the different interpersonal and textual functions of *well* ICLE-NO and LOCNESS. As Table 16 shows, *well* is mostly used interpersonally in both corpora, although there is a difference in proportion between the corpora. The learners in ICLE-NO have a higher frequency of the interpersonal *well* compared to the native speakers in LOCNESS. In 17 out of 39 instances of *well* in ICLE-NO and eight out of 15 instances of *well* in LOCNESS, *well* prefaces an answer to a question:

(56) You are not any less tired after ten minutes more sleep when it is eight o'clock in the morning anyhow, are you? **Well**, if you ask me […]. (ICLE-NO-OS-0036.1)

(57) Does he have problems falling to sleep at night because of bad conscience? **Well**, first you might ask yourself, should he really […]. (ICLE-NO-OS-0038.1)

(58) Why did this all happen? **Well**, it goes back to who has the most power.
(ICLE-US-SCU-0017.2)

(59) How did I get there? **Well**, I applied to this program […]. (ICLE-US-MICH-0037.1)

Table 16: Raw frequencies of the functions of *well* in ICLE-NO and LOCNESS

	ICLE-NO	LOCNESS
Interpersonal functions		
Preface an opinion	11	3
Preface an answer to a question	17	8
Textual functions		
Preface a clarification	5	2
Preface a conclusion	2	2
Searching for the right word/phrase	2	0
Continue an opinion	2	0
Total number of instances	39	15

Source: **Data from ICLE-NO and LOCNESS**

Examples (56), (57), (58) and (59) illustrate occurrences where *well* prefaces an answer to a question, which was the most common use of *well* in the material. However, there are differences between the examples from ICLE-NO, (56) and (57), and the examples from LOCNESS, (58) and (59). In the examples from LOCNESS, *well* is used to answer a question which is asked to make a transition to the next scene in the text, to move the discussion forward. In the examples from ICLE-NO, the questions asked are highly interactional, they are asked directly to the reader and thereafter answered by the writers themselves. This type of reader involvement when answering questions was not found in the LOCNESS material. However, to ask and answer a question to make a transition in the text was also found in the ICLE-NO material.

The second most common function of *well* in both corpora was *well* prefacing an opinion, as illustrated in examples (60) and (61):

(60) It will not be like it is in science fiction movies. **Well**, I don't think so.
(ICLE-NO-HO-0042.1)

(61) Thus when serious criminal cases occur, and the state that they occur in does not have the death penalty, a debate occurs over it necessity. […]. **Well**, I believe that no matter what the circumstances, there is no need for a death penalty […].
(ICLE-US-MRQ-0016.1)

In these examples, the writer explicitly expresses his or her opinion about the topic. Even though this was the second most common function of *well* in the material, it only occurred eleven times in ICLE-NO and three times in LOCNESS.

There are relatively small differences between the two corpora in terms of the textual functions. There are four instances in LOCNESS where *well* has solely a textual function, while there are eleven instances in ICLE-NO. However, there are two textual functions of *well* in ICLE-NO which were not found in the LOCNESS material.

In both ICLE-NO and LOCNESS, *well* has a textual function when it prefaces a clarification and a conclusion. These functions are illustrated in (62) and (63) from LOCNESS:

(62) As well, the fact that so many people (especially in the US) have television sets means that everybody (**well**, at least everybody who watches) receives the same inflow of information and ideas. (ICLE-US-MICH-0040.1).

(63) In fact some who support the death penalty may only support it so they can gain political support by showing that they will "take no prisoners" and be "tough on law and order". **Well**, let's be tough on law and order by cracking down on criminals, but no by doing it by committing another crime […].
(ICLE-US-MRQ-0016.1)

There are two other functions in ICLE-NO which were not found in LOCNESS. These functions are illustrated in examples (64) and (65):

(64) This will almost most certainly not be the same for everybody, but hopefully we can reach some sort of compromise. If we don't, **well** who knows what the future might bring? (ICLE-NO-HO-0037.1)

(65) This was when I discovered that boys in their 20's DRINK, and thus couldnt care less that your face looked like ……. **well**, something very strange […].
(ICLE-NO-UO-0064.1)

In (64) the writer uses *well* to continue his or her opinion about the topic, after interrupting him- or herself. In (65) the writer uses *well* while he or she is searching for the right word to use next. These functions of *well* may not be what we expect to find in written discourse.

6.2.6 *You know*

There are three instances of the discourse marker *you know* in ICLE-NO, while there are none in the LOCNESS corpus. The marker *you know* can occur in all positions, and this is also displayed in ICLE-NO. Of the four instances, two occur clause initially, while one occurs post verbally in medial position and one clause finally. *You know* can function both textually and interpersonally, but in the ICLE-NO material, the three instances were all classified as having an interpersonal function. The interpersonal functions found in the material are presented in Table 17 (see page 50).

Table 17: Raw frequencies of the functions of *you know* in ICLE-NO and LOCNESS

	ICLE-NO	LOCNESS
Interpersonal functions		
Mark reference to shared knowledge	1	0
Acknowledge that the speaker is right	2	0
Total number of instances	3	0

Source: **Data from ICLE-NO and LOCNESS**

These interpersonal functions of *you know* have in common that they in some way help the writer to address the reader. However, they serve different purposes:

> (66) I'm not meaning to be reactionary about anything. Actually, **you know**, I'm not a reactionary kind of a (modern) man. (ICLE-NO-UO-0043.2)

> (67) But we need the courage to blow the whistle every now and then and grant ourselves some breading space. Breading space can go hand in hand with reflection **you know**. Reflection may develop into dreams and imagination. (ICLE-NO-UO-0043.2)

> (68) So as far as sex and girls go, I have been told that when a woman has casual sex, she will expect something more; **you know**, its the classic "a whole week and he still hasnt called" scenario" […]. (ICLE-NO-UO-0065.1)

In (66), with the help of both markers *actually* and *you know*, the writer comes to terms with and enhances the fact that he is not a 'reactionary man'. The writer uses *you know* to ask the reader to agree with him. In example (67), the writer uses *you know* to ask the reader to agree with him or her, while in example (68), the writer uses *you know* to imply that the reader should understand what 'a whole week and he still hasn't called'-scenario is. The functions displayed by *you know* in ICLE-NO are functions that are well-known functions of this discourse marker.

6.2.7 *I mean*

The discourse marker *I mean* occurs in both ICLE-NO and LOCNESS, even though there are only a few instances represented in the material; seven in ICLE-NO and two in LOCNESS. The two instances in LOCNESS are both placed clause initially, while six of the seven instances in ICLE-NO are placed clause initially and one clause finally. In the material, both textual and interpersonal functions of *I mean* were found. These are presented in Table 18 (see page 51).

Table 18: Raw frequencies of the functions of *I mean* in ICLE-NO and LOCNESS

	ICLE-NO	LOCNESS
Interpersonal functions		
Instruct the hearer to continue attending to the prior utterance	1	0
Textual functions		
Preface an explanation	2	0
Preface an expansion	4	1
Express speaker tone	0	1
Total number of instances	7	2

Source: **Data from ICLE-NO and LOCNESS**

Table 18 shows that there is no interpersonal function of *I mean* in the material from LOCNESS, while there is one interpersonal function present in the ICLE-NO material:

> (69) She phoned us later that same day, and calmed my parents by saying – "It is a small world, so don't worry about me". But is it? A small world, **I mean**? Is that what the immigrants wrote back to Europe in 1607? – "It's a small world!".
> (ICLE-NO-UO-0089.1)

In (69), the writer uses *I mean* to make sure that the reader understands that 'it' refers to 'a small world', thereby asking the reader to continue to attend to the prior clause 'It is a small world, so don't you worry about me', to be able to understand the upcoming information.

The other functions of *I mean* in the material are textual. In ICLE-NO there are two instances which preface an explanation:

> (70) I never eat breakfast, and I don't believe it's damaging my health at all. **I mean**, we all know how it's just to close to lunch. (ICLE-NO-OS-0035.1)

> (71) **I mean**, YES, circumcision of women is clearly a very bad thing, as is abusive husbands, obsessive boyfriends, date rape or just plain rape. Im not dumb, I know that theese things happen. (ICLE-NO-UO-0064.1)

This function was not found in the LOCNESS material. One function that was found in LOCNESS and not ICLE-NO was the writer's use of *I mean* to express tone in the message:

> (72) When the police arrived, I went with them into my house and found that everything, **I mean** everything, had been taken. (ICLE-US-IND-0019.1)

In example (72), the writer uses *I mean* to express the tone of the situation. The writer wants to the reader to really understand the seriousness of the situation the writer is portraying. *I mean* as a preface to an expansion were found in both corpora: four instances in ICLE-NO and one in LOCNESS. This function of *I mean* is illustrated in examples (73) and (74):

> (73) Maybe many of these daydreamers actually did something about their dreams – **I mean** if you consider the amount of people immigrating to America, at least some of them must have had a dream of something better. (ICLE-NO-UO-0040.2)

> (74) Television and magazine ads display the beauty products or diets in a manner which we women think that we need them. **I mean** if the model in the commercial can look like that because she uses the product – so can I (yeah right). (ICLE-US-SCU-0004.2)

Both these examples show that the writer uses *I mean* as a link between the previous clause and the next in order to connect the two and expand the previous statement in a new clause.

I mean occurs in both ICLE-NO and LOCNESS, and even though there is a slight difference between the corpora in terms of frequency, *I mean* is still slightly more frequent in ICLE-NO. *I mean* as a preface to an expansion is found in both corpora. In terms of differences, in LOCNESS there are no instances that have an interpersonal function, nor is *I mean* used as a preface to an explanation. In ICLE-NO there are no instances where the writer uses *I mean* to express writer tone.

6.3 Summary

The quantitative analysis (see section 6.1) of frequency, position and function of the selected discourse markers revealed both similarities and differences between the two corpora. Both the learners and the native writers tend to place discourse markers in initial position, which coincides with the notion that discourse markers are usually placed in initial position. All instances of the selected discourse markers were placed in initial position in LOCNESS, while there were only a few instances which occurred in medial and final position in ICLE-NO. The syntactic position of *anyway*, *like* and *you know* cannot be compared between the two corpora since there are no occurrences of these markers in LOCNESS. The results show that there is a small difference in terms of discourse marker position between the two groups.

The frequency analysis revealed a difference between the two corpora in terms of how frequent these markers are in each corpus. Discourse markers are more frequently used in ICLE-NO than in LOCNESS, and thus overrepresented in the learner group compared to the native writer group.

In terms of how these discourse markers are used, the quantitative results (see section 6.1) revealed that the selected discourse markers are more frequently used with a textual function by both groups. However, the Norwegian learners of English in ICLE-NO use discourse markers interpersonally more often than the native novice writers in LOCNESS. The qualitative analysis (see section 6.2) showed that the learners in ICLE-NO use the discourse markers more interactively than the novice writers in LOCNESS, as they both had a higher percentage of interpersonal instances and also because there were some interpersonal functions in ICLE-NO which were not found in the LOCNESS material.

6.4 Discussion

The aim of this study was to investigate whether and to what extent Norwegian learners of English use discourse markers in their writing, and how they use these discourse markers. The aim was to answer these research questions:

RQ1: Do Norwegian learners of English overuse discourse markers in their writing compared to native speakers of English?

RQ2: If they overuse discourse markers, how do Norwegian learners of English use discourse markers in their writing compared to native speakers of English?

RQ3: If the answer to RQ1 is 'yes', what are possible reasons for this overuse of discourse markers in Norwegian learner writing?

The findings presented in the quantitative analysis suggest that both Norwegian learners and native speakers use discourse markers in their writing. However, the Norwegian learners in ICLE-NO use discourse markers more frequently compared to the native speakers in LOCNESS. The qualitative analysis showed that both groups use discourse markers for organizing purposes to a greater extent than using them to appeal to or to include the reader. Even so, both groups used discourse markers with interpersonal functions. This suggests that discourse markers are not only used by writers to organize text, but also as a way for writers to include the reader in the text and argumentation. However, even though both groups use discourse markers interpersonally, there was a higher percentage of the use of interpersonal functions in ICLE-NO compared to LOCNESS.

The findings from the quantitative and qualitative analysis show two things. Firstly, the fact that the Norwegian learners in ICLE-NO use discourse markers in their writing to a

greater extent compared to the native writers in LOCNESS supports the suggestion that learners of English are more likely to use spoken-like features in writing compared to native speakers. This also supports the notion that learners of English use more informal language when writing academic texts than native speakers do. Secondly, as the research presented in section 2.2 suggests, Norwegian learners of English are considerably more visible, personally involved and interactive in their writing compared to English native writers. This is also supported by the results in this study's quantitative and qualitative analysis. Both the textual and the interpersonal uses of the discourse markers in the study include functions such as emphasize speaker opinion, mark reference to shared knowledge with the reader, make requests, preface opinions and questions and mark a common ground with the reader. These functions are examples of how the writer shows writer and/or reader visibility.

The question is why the Norwegian writers in ICLE-NO overuse these markers compared to the native speaker group in LOCNESS. Is it an unconscious choice based on their unawareness of register, transfer from their mother tongue, influence from oral language, or is it due to the fact that they are novice writers and that there is a cultural difference between Norwegian and English writing? There might be several reasons for the use of oral features, in this case discourse markers, in Norwegian learner writing. The discourse markers could be a way for the writers to create a personal tone in their texts, i.e show reader/writer visibility. This may be a possible reason since Norwegian learners are in fact more visible and personal in their texts compared to other learner groups and English native speakers (c.f Hasselgård 2009, 2016 and Fossan 2011). The overuse compared to native English speakers could therefore be due to a difference between writing cultures. As Fossan (2001) points out, overuse of reader/writer visibility can be "caused by transfer of norms from the L1, and perhaps cultural norms regarding the acceptance of a more personal style in formal genres" (Fossan 2011, 154).

Even so, the use of discourse markers as organizers in writing is considered informal and not common in academic writing. It is difficult to pinpoint the cause of this informal tone, but there are two reasons that may be plausible. First of all, we have to remember that the writers in ICLE-NO (and LOCNESS) are considered novice academic writers. They have not yet received sufficient training to master the academic genre compared to expert writers. It is even more difficult to master this genre in another language. Furthermore, the total number of instances of discourse markers found in ICLE-NO is relatively low compared to the total number of words. This might suggest that most of the writers in ICLE-NO are in fact aware

that discourse markers belong to the spoken register. Therefore, the development aspect could be the most likely reason why the learners adopt a more informal style of writing, since some of the writers seem less experienced than others. Secondly, there are several other, more formal linking words that the writers could have used in their writing. There is a possibility that some learner writers have not received sufficient training in terms of the differences between genres when it comes to linking words.

Since it is difficult to pinpoint one reason that is more plausible than another, it would be natural to resort to the answer that the use of discourse markers in learner language is caused by several different factors: cultural differences between Norwegian and English writing, and acceptance of personal involvement in texts, unsatisfactory teaching of the difference between different genres, and that the writers in ICLE-NO are in fact novice writers, which means that they do not yet have sufficient training to master the academic genre.

7   Concluding remarks

The aim of this study was to reveal spoken-like features (discourse markers) in Norwegian learner language and thereafter compare the findings to native speakers of English, in order to add to the discussion whether learners of English apply a more 'chatty' style when writing texts in English. To investigate the study's research questions, a contrastive interlanguage analysis has been carried out using two comparable corpora, ICLE-NO and LOCNESS. The quantitative analysis revealed an overuse of the discourse markers *so, actually, anyway, well, you know* and *I mean* by the Norwegian learners in ICLE-NO compared to the English native speakers in LOCNESS. The qualitative analysis revealed that Norwegian learners of English use discourse markers with an interpersonal function more frequently than English native speakers. The findings from the analysis resonated with previous studies such as Gilquin and Paquot (2008), Fossan (2011), Hasselgård (2009) and Hasselgård (2016). Possible reasons for the overuse of discourse markers and the more frequent use of discourse markers with an interpersonal function was thereafter discussed. There is no absolute answer to why Norwegian learners overuse discourse markers; rather there seem to be multiple reasons, such as insufficient teaching of the difference between English genres, influence of the Norwegian writing culture and that the writers in ICLE-NO are novice writers and have not yet had sufficient writing training within the academic genre.

7.1 Pedagogical implications

This study has shown that Norwegian learners of English overuse discourse markers, which are considered informal and characteristic of spoken language, in their academic writing compared to native speakers. The study has also shown that Norwegian learners of English are more interactive in their writing than English native writers. Even though there seem to be differences between the two writing cultures and norms, it is important to make our students aware of these differences when we teach. As Gilquin and Paquot (2008) mention, some English foreign language textbooks give the impression that linking words and phrases are synonymous (2008, 55), when in fact they are very different from each other in terms of stylistics and in what genre they are most common. This study will hopefully draw attention to how we teach text coherence and stylistics across different genres in English.

7.2 Limitations of the study and suggestions for further research

This study has provided further insights on spoken features in written learner language, and therefore added knowledge to the field of second language research. However, the material in LOCNESS and ICLE-NO is relatively small, and therefore we cannot generalize these findings before performing the same research on a larger set of data. Also, we have to consider individual variation. Some writers in ICLE-NO and LOCNESS are less experienced than others, and an investigation of the individual variation in both corpora, might reveal that some writers use discourse markers more frequently in their texts than others. This means that some writers may have contributed more to the results than others. We also have to take genre into consideration. These texts are written by students in higher education, but the argumentative genre is more open to personal involvement and maybe this leads to a more informal use of language than another genre would.

Even though this study has its limitations, it has still provided interesting findings, and hopefully it has sparked a further interest in investigating spoken features in written learner language. I hope that this study has created further awareness about this topic amongst teachers of English in Norway. It is of the utmost importance that we teachers always change, update and improve our teaching to make it relevant and important for our students. This study and several other studies have established that learners of English tend to use oral language features in their academic writing. For further research it would be interesting to investigate the individual variation of learners of English to find out why they tend to use spoken-like features in writing. If we know more about the learners' writing experience, background, teaching and alike, we would understand better why learners write as they do. In addition, it may prove useful to do research on another academic genre that is less open for personal involvement. It would also be interesting to collect material for an updated and more current corpus, and see if there is any change in learner writing from the 1990s to 2018.

References

Primary sources

The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/; CQP-edition version 4.0; The CQP-edition of BNCweb was developed by Sebastian Hoffmann and Stefan Evert, accessed via http://www.tekstlab.uio.no/bnc/BNCquery.pl?theQuery=search&urlTest=yes. (06.03.2018).

ICLE (The International Corpus of Learner English): https://uclouvain.be/en/research-institutes/ilc/cecl/icle.html

LOCNESS (The Louvain Corpus of Native English Essays): http://www.learnercorpusassociation.org/resources/tools/locness-corpus/

Secondary sources

Aijmer, Karin. 2002. *English Discourse Particles: Evidence from a Corpus*. Amsterdam: John Benjamins Publishing Company.

Aijmer, Karin. 2002. "Modality in advanced Swedish learners' written interlanguage". In *Computer Learner Corpora, Second Language Acquisition and Foreign Language Acquisition*, edited by Sylviane Granger, Joseph Hung and Stephanie Petch-Tyson, 55–76. Philadelphia: John Benjamins.

Aijmer, Karin. 2011. "*Well I'm not sure I think*…The use of *well* by non-native speakers". *International Journal of Corpus Linguistics* 16, (2): 231–254.

Altenberg, Bengt. 1997. "Exploring the Swedish component of the International Corpus of Learner English". In *Proceedings of International Conference on Practical Applications in Language Corpora*, edited by Barbara Lewandowska-Tomaszczyk and Patrick James Melia, 1197–132. Łódź: Łódź University Press.

Andersen, Gisle. 1998. "The pragmatic marker *like* from a relevance-theoretic perspective". In *Discourse Markers: Descriptions and Theory*, edited by Andreas H. Jucker and Yael Ziv, 147–170. Amsterdam: John Benjamins Publishing Company.

Ball, Catherine. 1994. "Automated text analysis. Cautionary tales". *Literary & Linguistic Computing* 9, (4): 295–302.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Buysse, Lieven. 2012. "*So* as a multifunctional discourse marker in native and learner speech". *Journal of Pragmatics* 44, (13): 1767–1782. Accessed March 03, 2018. https://doi.org/10.1016/j.pragma.2012.08.012

Clancy, Brian. 2010. "Building a corpus to represent a variety of a language". In *The Routledge Handbook of Corpus Linguistics*, edited by Anne O'Keeffe and Michael McCarthy, 80–92. London: Routledge

Corpus collection guidelines. Accessed February 10, 2018. https://uclouvain.be/en/research-institutes/ilc/cecl/corpus-collection-guidelines.html

English Oxford Living Dictionaries. "Corpus" Accessed January 16, 2018. https://en.oxforddictionaries.com/definition/corpus

Ferrara, Kathleen Warden. 1997. "Form and function of the discourse marker *anyway*: implications for discourse analysis". *Linguistics* 35, (2): 348–378. Accessed March 09, 2018. https://doi.org/10.1515/ling.1997.35.2.343

Fossan, Heidi. 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 thesis, University of Oslo.

Fox Tree, E. Jean and Josef C. Schrock. 2002. "Basic meanings of *you know* and *I mean*". *Journal of Pragmatics* 34, (6): 727–747. Accessed March 28, 2018. https://doi.org/10.1016/S0378-2166(02)00027-9

Gilquin, Gaëtanelle and Magali Paquot. 2008. "Too chatty: Learner academic writing and register variation". *English Text Construction* 1, (1): 41–61.

Granger, Sylviane. 2008. "Learner corpora". In *Handbook on Corpus Linguistics*, edited by Anke Lüdeling and Merja Kytö, 259–275. Berlin and New York: Walter de Gruyter.

Granger, Sylviane. 2009. "The contribution of learner corpora to second language acquisition and foreign language teaching. A critical evaluation". In *Corpora and Language Teaching*, edited by Karin Aijmer, 13–32. Amsterdam: John Benjamins Publishing Company.

Granger, Sylviane. 2015. "Contrastive interlanguage analysis. A reappraisal". *International Journal of Learner Corpus Research* 1, (1): 7–24.

Greis, Stefan Thomas. 2009. *Quantitative Corpus Linguistics with R: A Practical Introduction*. New York: Routledge.

Halliday, M.A.K and Christian M.I.M Mattheissen. 2004. *An Introduction to Functional Grammar*. 3rd ed. London: Arnold.

Hasselgren, Angela. 1994. "Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary". *International Journal of Applied Linguistics* 4, (2): 237–259.

Hasselgård, Hilde. 2009. "Thematic choice and expressions of stance in English argumentative texts by Norwegian learners". In *Corpora and Language Teaching*, edited by Karin Aijmer, 121–139. Amsterdam: John Benjamins

Hasselgård, Hilde. 2016. "Discourse-organizing metadiscourse in novice academic English". In *Corpus Linguistics on the Move: Exploring and Understanding English through Corpora*, edited by María José López-Couso, Belén Méndez-Naya, Paloma Núñez-Pertejo & Ignacio M. Palacios-Martínez, 106–131. Leiden & Boston: Brill Rodopi.

Hasselgård, Hilde and Stig Johansson. 2011. "Learner corpora and contrastive interlanguage analysis". In *A Taste for Corpora. In honour of Sylviane Granger*, edited by Fanny Meunier, Sylvie De Cock, Gaëtanelle Gilquin and Magali Paquot, 33–62. Amsterdam: John Benjamins Publishing Company

Johansson, Stig. 2008. "Contrastive analysis and learner language: A corpus-based approach". Accessed February 10, 2018. http://www.hf.uio.no/ilos/forskning/grupper/Corpus_Linguistics_Group/papers/contrastive-analysis-and-learner-language_learner-language-part.pdf

Johansson, Stig. 2011. "A multilingual outlook of corpora studies". In *Perspectives on Corpus Linguistics*, edited by Vander Viana, Sonia Zyngier and Geoff Barnbrook, 117–129. Amsterdam: Benjamins.

Johnsson, Michaela. 2017. "Discourse markers in written discourse: Influence of speech in written learner English". Term paper, University of Oslo.

Leech, Geoffrey. 1992. "Corpora and theories of linguistic performance". In *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*, edited by Jan Svartvik, 105–123. Berlin: Mouton de Gruyter.

LOCNESS Description. Accessed February 10, 2018. https://uclouvain.be/en/research-institutes/ilc/cecl/locness.html

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

Müller, Simone. 2005. *Discourse Markers in Native and Non-native English Discourse*. Amsterdam: John Benjamins Publishing Company.

Nelson, Mike. 2010. "Building a written corpus. What are the basics?". In *The Routledge Handbook of Corpus Linguistics*, edited by Anne O'Keeffe and Michael McCarthy, 53–65. London: Routledge

Paquot, Magali. 2010. *Academic Vocabulary in Learner Writing: From Extraction to Analysis*. London: Continuum

Paquot, Magali. 2013. "Lexical bundles and L1 transfer effects". *International Journal of Corpus Linguistics* 18, (3): 391-417.

Sandal, Karoline Lilleås. 2016. ""And like, they said…well, you know": A corpus-based study of the discourse markers 'like', 'well', and 'you know' in spoken Norwegian learner language and British English". Master thesis, University of Oslo.

Schiffrin, Deborah. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.

Schourup, Lawrence. 1985. *Common Discourse Particles in English Conversation*. New York: Garland

Scott, Mike. 2012. WordSmith Tools version 6. Stroud: Lexical Analysis Software.

Seidlhofer, Barbara. 2004. "Research perspectives on teaching English as a lingua franca". *Annual Review of Applied Linguistics* 24: 209–239. Accessed February 02, 2018. https://doi.org/10.1017/S0267190504000145

Sinclair, John. 1996. "Quality". *EAGLES. Preliminary Recommendations on Corpus Typology*. Accessed January 30, 2018. http://www.ilc.cnr.it/EAGLES96/corpustyp/node12.html

Sinclair, John. 2005. "Corpus and text: Basic principles". In *Developing Linguistic Corpora: a Guide to Good Practice*, edited by Martin Wynne, 1–16. Oxford: Oxbow Books. Accessed January 16, 2018. http://ota.ox.ac.uk/documents/creating/dlc/chapter1.htm

Smith, Sara W. and Andreas H. Jucker. 2000. "*Actually* and other markers of an apparent discrepancy between propositional attitudes of conversational partners". In *Pragmatic Markers and Propositional Attitude*, edited by Gisle Andersen and Thorstein Freitheim, 207–237. Amsterdam: John Benjamins Publishing Company.

Ädel, Annelie. 2008. "Metadiscourse across three varieties of English: American, British and advanced-learner English". In *Contrastive Rhetoric*: *Reaching to Intercultural Rhetoric*, edited by Ulla Connor, Ed Nagelhout and William Rozycki, 45–63. Amsterdam: John Benjamins.