

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.*

Heidi Fossan



A thesis presented to
The Department of Literature, Area Studies and European Languages
University of Oslo
November 2011

Supervisors: Hildegunn Dirdal and Professor Hilde Hasselgård

Abstract

In the present study, Norwegian advanced learners of English are found to overuse many features of writer/reader visibility compared to native speakers of English. This is shown by comparing corpora of Norwegian learner writing and native speaker student writing.

Norwegian learners are also found to have a higher level of writer/reader visibility than many other learner groups. An in-depth study of pronouns as markers of writer/reader visibility shows that the high level of writer/reader visibility in Norwegian learner writing may have been caused by transfer from the learners' first language, because it seems like pronouns are more common in the Norwegian language than in the English language. These findings are relevant to the teaching of English as a foreign language in Norway, because students should be made aware of the differences between Norwegian and English. In order to be taken seriously when writing in English, the students may have to reduce the level of personal involvement in their texts.

In addition to quantitative investigations, the present study includes a qualitative investigation of pronouns, particularly with respect to the way they express different functions of the writer and the reader. This investigation shows that pronouns are mainly used to express the same functions in Norwegian learner writing as in native speaker student writing. However, several functions were found to be significantly overused, such as the organizing function of the writer, expressed by the pronoun *I*, and the inclusive (reader-involving) function of the pronoun *we*. As a complement to the quantitative information about writer/reader visibility features, such detailed descriptions of the learners' usage of pronouns may be useful in the teaching of English as a foreign language.

Acknowledgements

I would like to thank my supervisor Hildegunn Dirdal for helpful advice and support in the process of writing this thesis. By thoroughly reading, correcting and giving response to my drafts, she was to me a committed and reliable guide in a completely new landscape. Thanks also to my supervisor Professor Hilde Hasselgård, who offered invaluable guidance on my thesis, particularly during the final stressful weeks. I would also like to thank Hilde Hasselgård and Signe Oksefjell Ebeling for having enough faith in me to let me participate in the ENPC and VESPA projects, and for inviting me to the Learner Corpus Research conference in Louvain. These were very useful and motivating experiences. Hildegunn, Hilde and Signe have all been great sources of encouragement and inspiration.

I am also grateful to my husband, Øivind, who has listened patiently to my ideas and given me encouraging feedback. He has been my constant supporter in this challenging, but rewarding, final semester.

Finally, thanks to my precious university lunch companions, Miriam, Edda and Kari Helen, for filling my time at Blindern with interesting debates as well as complete nonsense, and for adding an absurd sense of humor to the otherwise serious university activities.

Table of contents

ABSTRACT	- 1 -
ACKNOWLEDGEMENTS	- 2 -
TABLE OF CONTENTS	- 3 -
LIST OF TABLES	- 8 -
LIST OF FIGURES	- 10 -
LIST OF MENTIONED CORPORA	- 12 -
LIST OF ABBREVIATIONS	- 13 -
1. INTRODUCTION	- 14 -
1.1. Writer/reader visibility	- 15 -
1.2. Aims and scope	- 18 -
1.3. Research questions	- 19 -
1.4. Outline of the thesis	- 21 -
2. THEORETICAL BACKGROUND AND PREVIOUS STUDIES	- 22 -
2.1. Expected level of writer/reader visibility in argumentative writing	- 23 -
2.2. Previous research on writer/reader visibility in argumentative learner writing	- 24 -
2.2.1. First person singular pronouns	- 25 -
2.2.2. Patterns of subjective stance	- 28 -
2.2.3. First person plural pronouns	- 30 -
2.2.4. Second person pronouns	- 31 -
2.2.5. Patterns introducing other voices	- 31 -
2.2.6. Disjuncts	- 32 -

2.2.7. Emphatic particles	- 33 -
2.2.8. Reference to situation of writing/reading	- 35 -
2.2.9. Sentence types	- 36 -
2.2.10. Summing up writer/reader visibility in argumentative learner writing	- 39 -
2.3. Expected level of writer/reader visibility in academic writing	- 39 -
2.4. Previous research on writer/reader visibility in academic learner writing	- 41 -
2.4.1. Underuse of academic vocabulary in learner writing	- 41 -
2.4.2. Overuse of writer/reader visibility features in academic (and argumentative) writing by Norwegian and French learners	- 42 -
2.4.3. Underuse of personal reference in academic writing by Hong Kong learners	- 44 -
2.4.4. Overuse of personal reference in academic writing by Chinese and Greek learners	- 47 -
2.4.5. Summing up writer/reader visibility in academic learner writing	- 49 -
2.5. Suggested reasons for overuse and underuse of writer/reader visibility features	- 49 -
2.5.1. General language learning issues	- 50 -
2.5.2. L1 transfer	- 51 -
2.5.3. Lack of register awareness	- 51 -
2.5.4. Tasks and setting	- 52 -
2.5.5. Differences in instruction	- 52 -
2.6. Functions of the writer and the reader	- 53 -
2.6.1. <i>I</i> in argumentative writing	- 53 -
2.6.2. <i>I</i> in academic writing	- 54 -
2.6.3. <i>We</i> in argumentative writing	- 56 -
2.6.4. <i>We</i> in academic writing	- 57 -
2.6.5. <i>You</i> in both genres	- 58 -
2.7. What the present study aims to contribute	- 59 -
3. METHOD AND MATERIAL	- 62 -
3.1. The preliminary study of writer/reader visibility	- 62 -
3.1.1. Features studied	- 63 -
3.1.2. The corpus investigations	- 68 -
3.2. The in-depth study of pronouns	- 68 -
3.2.1. Choosing personal pronouns	- 68 -

3.2.2. Studying pronouns quantitatively	- 69 -
3.2.3. Studying pronouns qualitatively	- 70 -
3.3. The main corpora	- 71 -
3.3.1. NICLE	- 71 -
3.3.2. LOCNESS	- 75 -
3.3.3. N-VESPA	- 77 -
3.3.4. BAWE(-ling)	- 79 -
3.4. The sources of Norwegian L1 writing	- 81 -
3.4.1. NOESS	- 81 -
3.4.2. NOBA	- 82 -
3.5. Comparable information about the corpora and collections	- 85 -
4. RESULTS AND ANALYSIS PART 1: THE PRELIMINARY STUDY	- 86 -
4.1. Overall writer/reader visibility in learner writing	- 87 -
4.2. Detailed overview of writer/reader visibility in argumentative writing	- 88 -
4.2.1. First person singular pronouns	- 90 -
4.2.2. Patterns of subjective stance	- 91 -
4.2.3. First person plural pronouns	- 92 -
4.2.4. Second person pronouns	- 93 -
4.2.5. Patterns introducing other voices	- 94 -
4.2.6. Disjuncts	- 95 -
4.2.7. Emphatic particles	- 96 -
4.2.8. Reference to situation of writing/reading	- 97 -
4.2.9. Sentence types	- 98 -
4.3. Detailed overview of writer/reader visibility in academic writing	- 99 -
4.3.1. First person singular pronouns	- 99 -
4.3.2. Patterns of subjective stance	- 100 -
4.3.3. First person plural pronouns	- 101 -
4.3.4. Second person pronouns	- 102 -
4.3.5. Patterns introducing other voices	- 103 -
4.3.6. Disjuncts	- 104 -
4.3.7. Emphatic particles	- 105 -
4.3.8. Reference to situation of writing/reading	- 105 -

4.3.9. Sentence types	- 106 -
4.4. Summing up the preliminary study	- 106 -
5. RESULTS AND ANALYSIS PART 2: THE QUANTITATIVE IN-DEPTH STUDY	- 109 -
5.1. The effect of L1 transfer	- 109 -
5.1.1. Argumentative writing in Norwegian and English	- 110 -
5.1.2. Academic writing in Norwegian and English	- 113 -
5.1.3. Norwegian learners and learners with other L1s	- 115 -
5.1.4. Final remarks on the L1 transfer issue	- 120 -
5.2. The effect of academic experience	- 120 -
5.3. Summing up the quantitative in-depth study	- 124 -
6. RESULTS AND ANALYSIS PART 3: THE QUALITATIVE IN-DEPTH STUDY	- 125 -
6.1. I in argumentative writing	- 126 -
6.1.1. Results	- 126 -
6.1.2. Analysis	- 130 -
6.2. I in academic writing	- 132 -
6.2.1. Results	- 132 -
6.2.2. Analysis	- 136 -
6.3. We in argumentative writing	- 136 -
6.3.1. Results	- 136 -
6.3.2. Analysis	- 140 -
6.4. We in academic writing	- 140 -
6.4.1. Results	- 140 -
6.4.2. Analysis	- 144 -
6.5. You in argumentative writing	- 145 -
6.5.1. Results	- 145 -
6.5.2. Analysis	- 147 -
6.6. You in academic writing	- 147 -
6.6.1. Results	- 147 -

6.6.2. Analysis	- 149 -
6.7. Summing up the qualitative in-depth study	- 150 -
7. CONCLUSION	- 153 -
7.1. Summing up the findings	- 153 -
7.2. Pedagogical implications	- 155 -
7.3. Looking ahead	- 157 -
BIBLIOGRAPHY	- 159 -
CORPORA USED	- 163 -

List of tables

Table 1: Overview of writer/reader visibility features	- 17 -
Table 2: 1 st person sg pronouns, Petch-Tyson (1998).....	- 26 -
Table 3: W/R visibility in French and Norwegian learner writing, Paquot, Hasselgård and Ebeling (forthcoming)	- 43 -
Table 4: Personal reference per 10,000 words in novice and professional writing, Hyland (2002a)	- 45 -
Table 5: 1 st and 2 nd person references per 10,000 words in learner writing and academic writing, based on table 3 in Coffin and Mayor (2004).....	- 47 -
Table 6: 1 st and 2 nd person references per 10,000 words in learner writing (in Coffin and Mayor 2004) and student writing in LOCNESS	- 48 -
Table 7: Functions of 1 st person pronouns in novice and professional writing, based on tables 4 and 5 in Hyland (2002a).....	- 55 -
Table 8: Different uses of <i>we</i> , Herriman (2009)	- 57 -
Table 9: Information about the NOBA texts	- 83 -
Table 10: Comparable information about NICLE, LOCNESS, NOESS, N-VESPA, BAWE and NOBA.....	- 85 -
Table 11: 1 st person sg pronouns in NICLE and LOCNESS	- 90 -
Table 12: Patterns of subjective stance in NICLE and LOCNESS	- 91 -
Table 13: 1 st person pl pronouns in NICLE and LOCNESS	- 92 -
Table 14: 2 nd person pronouns in NICLE and LOCNESS	- 93 -
Table 15: Patterns introducing other voices in NICLE and LOCNESS	- 94 -
Table 16: Disjuncts in NICLE and LOCNESS	- 95 -
Table 17: Emphatic particles in NICLE and LOCNESS	- 96 -
Table 18: Reference to situation of writing/reading in NICLE and LOCNESS	- 97 -
Table 19: Sentence types (questions and exclamations) in NICLE and LOCNESS.....	- 98 -
Table 20: 1 st person sg pronouns in N-VESPA and BAWE-ling.....	- 99 -
Table 21: Patterns of subjective stance in N-VESPA and BAWE-ling	- 100 -
Table 22: 1 st person pl pronouns in N-VESPA and BAWE-ling	- 101 -
Table 23: 2 nd person pronouns in N-VESPA and BAWE-ling	- 102 -
Table 24: Patterns introducing other voices in N-VESPA and BAWE-ling.....	- 103 -
Table 25: Disjuncts in N-VESPA and BAWE-ling	- 104 -

Table 26: Emphatic particles in N-VESPA and BAWE-ling.....	- 105 -
Table 27: Reference to situation of writing/reading in N-VESPA and BAWE-ling.....	- 105 -
Table 28: The significantly overused features in NICLE.....	- 107 -
Table 29: The significantly overused features in N-VESPA	- 108 -
Table 30: The distribution of the different pronoun groups in learner writing and native speaker writing, partly based on table 8.1 in Petch-Tyson (1998).....	- 119 -
Table 31: Functions of <i>I</i> in samples of 100 from NICLE and LOCNESS.....	- 129 -
Table 32: Functions of <i>I</i> – estimated frequencies in NICLE and LOCNESS and overuse in NICLE	- 130 -
Table 33: Functions of <i>I</i> in samples of 200 from N-VESPA and BAWE-ling	- 135 -
Table 34: Functions of <i>I</i> – estimated frequencies in N-VESPA and BAWE-ling and overuse in N-VESPA	- 135 -
Table 35: Functions of <i>we</i> in samples of 100 from NICLE and LOCNESS.....	- 139 -
Table 36: Functions of <i>we</i> – estimated frequencies in NICLE and LOCNESS and overuse in NICLE	- 139 -
Table 37: Functions of <i>we</i> in samples of 200 from N-VESPA and BAWE-ling.....	- 143 -
Table 38: Functions of <i>we</i> – estimated frequencies in N-VESPA and BAWE-ling and overuse in N-VESPA.....	- 144 -
Table 39: Functions of <i>you</i> in samples of 100 from NICLE and LOCNESS	- 146 -
Table 40: Functions of <i>you</i> – estimated frequencies in NICLE and LOCNESS and overuse in NICLE	- 147 -
Table 41: Functions of <i>you</i> in samples of 131 from N-VESPA and BAWE-ling.....	- 148 -
Table 42: Functions of <i>you</i> – estimated frequencies in N-VESPA and BAWE-ling and overuse in N-VESPA.....	- 149 -

List of figures

Figure 1: 1 st person sg pronouns per 10,000 words, based on table 8.1 in Petch-Tyson (1998)	27 -
Figure 2: 1 st person sg pronouns in SWICLE, based on table 7 in Ädel (2008).....	28 -
Figure 3: <i>I think</i> in learner writing, based on table 5.17 in Paquot (2010).....	29 -
Figure 4: 1 st person pl pronouns, based on table 8.1 in Petch-Tyson (1998)	30 -
Figure 5: 2 nd person pronouns, based on table 8.1 in Petch-Tyson (1998).....	31 -
Figure 6: Disjuncts in Swedish learner writing, based on table 7 in Ädel (2008).....	32 -
Figure 7: <i>Maybe</i> in learner writing, based on figure 5.7 in Paquot (2010).....	33 -
Figure 8: <i>Just</i> in different learner groups, based on table 8.1 in Petch-Tyson (1998).....	34 -
Figure 9: <i>Really</i> in different learner groups, based on table 8.1 in Petch-Tyson (1998).....	34 -
Figure 10: <i>Here</i> in different learner groups, based on table 8.1 in Petch-Tyson (1998).....	35 -
Figure 11: <i>Now</i> in different learner groups, based on table 8.1 in Petch-Tyson (1998).....	36 -
Figure 12: Questions in Swedish learner writing, based on table 7 in Ädel (2008)	37 -
Figure 13: Questions in learner writing, based on table 7.2 in Virtanen (1998)	38 -
Figure 14: Questions in different learner groups, based on table 7.2 in Virtanen (1998)	38 -
Figure 15: W/R visibility features in total per 100,000 words, based on Paquot, Hasselgård and Ebeling (forthcoming)	44 -
Figure 16: Personal reference per 10,000 words in novice and professional writing, based on table 3 in Hyland (2002a).....	45 -
Figure 17: Personal reference per 10,000 words in learner (novice) writing, professional writing and native speaker student (novice) writing, partly based on table 3 in Hyland (2002a)	46 -
Figure 18: W/R visibility in texts by male and female NICLE writers.....	74 -
Figure 19: Concordance list for <i>me</i> in LOCNESS	75 -
Figure 20: Overall level of W/R visibility in argumentative writing	88 -
Figure 21: Overall level of W/R visibility in academic writing	88 -
Figure 22: 1 st and 2 nd person pronouns per 10,000 words in L2 English, L1 Norwegian and L1 English (argumentative writing).....	111 -
Figure 23: 1 st person sg pronouns per 10,000 words in L2 English, L1 Norwegian and L1 English (argumentative writing).....	112 -
Figure 24: 1 st person pl pronouns per 10,000 words in L2 English, L1 Norwegian and L1 English (argumentative writing).....	112 -
Figure 25: 2 nd person pronouns per 10,000 words in L2 English, L1 Norwegian and L1 English (argumentative writing).....	112 -
Figure 26: 1 st and 2 nd person pronouns per 10,000 words in L2 English, L1 Norwegian and L1 English (academic writing).....	114 -

Figure 27: 1 st person pronouns (singular and plural) per 10,000 words in L2 English, L1 Norwegian and L1 English (academic writing)	- 114 -
Figure 28: 2 nd person pronouns per 10,000 words in L2 English, L1 Norwegian and L1 English (academic writing).....	- 115 -
Figure 29: 1 st and 2 nd person pronouns in NICLE, ICLE and LOCNESS.....	- 116 -
Figure 30: 1 st person sg pronouns in NICLE, ICLE and LOCNESS.....	- 117 -
Figure 31: 1 st person pl pronouns in NICLE, ICLE and LOCNESS	- 117 -
Figure 32: 2 nd person pronouns in NICLE, ICLE and LOCNESS	- 117 -
Figure 33: 1 st and 2 nd person pronouns in the writing of learners with different L1s, partly based on table 8.1 in Petch-Tyson (1998)	- 119 -
Figure 34: <i>I, we</i> and <i>you</i> , separately, in the writing of bachelor and master students in N-VESPA (per 10,000 words).....	- 121 -
Figure 35: <i>I, we</i> and <i>you</i> , combined, in the writing of bachelor and master students in N-VESPA (per 10,000 words).....	- 122 -
Figure 36: <i>I, we</i> and <i>you</i> , separately, in BAWE-ling at different levels (per 10,000 words)	- 123 -
Figure 37: : <i>I, we</i> and <i>you</i> , combined, in BAWE-ling at different levels (per 10,000 words).....	- 123 -

List of mentioned corpora

BAWE – British Academic Written English corpus

BAWE-ling – British Academic Written English corpus – linguistics discipline

BNC – the British National Corpus

BNC-AC-HUM – the humanities and arts section of the British National Corpus

COCA – Corpus of Contemporary American English

COMMENT – Herriman’s collection of British newspaper articles

ICE-GB – the British part of the International Corpus of English

ICLE - the International Corpus of Learner English

LOCNESS – the Louvain Corpus of Native Essays

NICLE – the Norwegian part of the International Corpus of Learner English

NOBA – Norwegian Bachelor Assignments

NOESS – Norwegian essays

SWICLE – the Swedish part of the International Corpus of Learner English

N-VESPA – the Norwegian part of the Varieties of English for Specific Purposes database

VESPA – the Varieties of English for Specific Purposes database

List of abbreviations

EFL – English as a foreign language

ESL – English as a second language

NNS – non-native speaker

NS – native speaker

Pl - plural

Sg - singular

W/R – writer/reader

1. Introduction

Mastering the English language is becoming increasingly important for students at university level, not only in English courses, but in all academic fields. English is now being referred to as “(...) the major international language for research and publication” (Paquot 2010, 1).

Studies have shown that the proportion of English titles on Norwegian university and college syllabi, range from 50 to 65 %, depending on the field of study (Hellekjær 2005, 14). In addition to reading English texts, many students are required to write assignments and exam papers in English. In the present study, I analyze such student writing, comparing Norwegian advanced learners of English to native speaker students, aiming to identify some of the challenges related to advanced learner writing.

One such challenge has been indicated in many previous studies of learner language, and is related to the style in learner writing. Studies have shown that advanced learners of English tend to write in a more personal style than their native speaker peers, a phenomenon which is often referred to as a high level of writer/reader visibility. The present study aims to describe to what extent this applies to Norwegian learners of English, by looking at the level of writer/reader visibility in Norwegian learner writing compared to native speaker writing, and compared to learner writing by learners with other first languages.

The level of writer/reader visibility is explored by studying the use of numerous features, such as pronouns, disjuncts, questions and exclamations. In addition to studying the frequency of many such features, the present study particularly focuses on the use of 1st and 2nd person pronouns. A quantitative in-depth study of pronouns compares the frequency of different pronouns in Norwegian learner writing and in EFL (English as a foreign language) writing by learners with other L1s. The frequency of pronouns in L1 Norwegian writing is also considered, in order to investigate the possibility that Norwegian EFL learners are influenced by the norms of their L1 when writing in English. Finally, a qualitative in-depth study investigates the use of 1st and 2nd person pronouns to express different functions of the writer and the reader. This will hopefully improve our understanding of Norwegian learners' use of writer/reader visibility features to express different meanings, which may make it easier to apply the findings from the present study in the teaching of English as a foreign language in Norway.

The following section (1.1.) will give an introduction to the study of the writer and the reader in a text. Here, I will explain the term *writer/reader visibility* and offer examples of

features of writer/reader visibility and how they are used in learner writing. In section 1.2., I will describe the aims and scope of the present study, and in section 1.3 I will present the research questions, hypotheses and methods applied to answer the questions. Section 1.4 provides an outline of the thesis.

1.1. Writer/reader visibility

Writers and readers may be given active or passive roles and various functions in the text, and they may have varying levels of visibility in the text. The terms *writer/reader visibility* (Petch-Tyson 1998), *involvement* (Ädel 2008) and *interpersonality* (Smith 1983) are used to describe the textual phenomenon where parts of a text draw attention towards the writer and/or the reader. Such parts of a text include features such as single words, word patterns and sentence types. Wales' dictionary of stylistics offers some examples of such features: "Obvious linguistic markers of the interpersonal mode include the first and second person pronouns *I* and *you*; terms of address; deictic elements; and speech acts such as questions and directives" (Wales 2011, 233).

Some of the linguistic markers mentioned in Wales, are exemplified below. In example 1, the pronoun *I* refers to the writer, and in example 2, the pronoun *you* refers to the reader. Both examples make the writer/reader visible, and they also attribute different roles to the writer/reader; in example 1, the pronoun *I* allows the writer to participate actively in the text, functioning as a text organizer or reader guide (more on this in chapter 2), while in example 2, the writer gives the reader the role of the visible addressee, which is not as active as the text organizer, but still more active than the invisible reader. While these are quite obvious writer/reader references, the third example is rather a reference to the text itself (a deictic element), which also makes the writer and the reader visible, but in a more subtle way. Finally, in example 4, the reader becomes visible through the question form of the sentence which implies a reader-address. The writer also becomes visible to some extent, as he is obviously the person who poses the question.

1. “Firstly, **I** will look at the development and distribution of TV, radio, music and Internet (...)” (NICLE)
2. “If **you** ask yourself what the most important thing in life is, **you** will probably say: family.” (NICLE)
3. “**In this essay** I will give my own view of this statement.” (NICLE)
4. “Are coloured people afraid of speaking to the press?” (NICLE)

The four examples above are all taken from texts written by Norwegian learners of English in the Norwegian component of the International Corpus of Learner English (ICLE/NICLE)¹. Many studies of learner writing in ICLE have concluded that the level of writer/reader visibility in this corpus is quite high when compared to control corpora of native speaker writing. Some of these studies have considered writer/reader visibility in general, looking at many different features, while other studies have focused on specific features, sometimes in relation to other topics than writer/reader visibility, but still contributing to our knowledge of learners’ use of these features. Table 1 presents a list of features based on previous studies, along with examples of such features, and references to the relevant studies. These studies will be discussed in chapter 2, and used as sources of comparison in the present study. The list of features will be referred to again in chapter 3, which discusses the material and methods in the present study, and all the features listed here will be investigated in chapter 4, in terms of their frequency in corpora of Norwegian learner writing and corpora of native speaker writing.

¹ More information about ICLE and NICLE in chapter 3.

Table 1: Overview of writer/reader visibility features

Writer/reader features	Examples	Studies
1 st person singular pronouns ²	I Me My Myself	Ädel (2008), Petch-Tyson (1998), Hasselgård (2009), Hyland (2002a), Coffin and Mayor (2004), Harwood (2005), Fløttum (2006), Paquot, Hasselgård and Ebeling (forthcoming)
Patterns of subjective stance	I believe I think	Hasselgård (2009), Paquot (2010)
1 st person plural pronouns	We Us Our Ourselves	Petch-Tyson (1998), Herriman (2009), Coffin and Mayor (2004), Kuo (1999), Hyland (2005), Kitagawa and Lehrer (1990), Paquot, Hasselgård and Ebeling (forthcoming)
2 nd person pronouns	You Your Yourself	Petch-Tyson (1998), Coffin and Mayor (2004), Smith (1983), Paquot, Hasselgård and Ebeling (forthcoming)
Patterns introducing other voices	Many think Let's say that	Hasselgård (2009)
Disjuncts	Of course Maybe	Ädel (2008), Paquot (2010), Paquot, Hasselgård and Ebeling (forthcoming)
Emphatic particles	Just Really	Petch-Tyson (1998)
Reference to situation of writing / reading	Here Now	Petch-Tyson (1998)
Sentence types (questions and exclamations)	Does this all sound utopian to you? Terrible, I know!	Virtanen (1998), Ädel (2008), Paquot, Hasselgård and Ebeling (forthcoming)

² Pronoun groups are defined differently in different studies. Personal pronouns like *I* and *me* are always included. In addition, reflexive pronouns and/or possessive pronouns and determiners may be included.

1.2. Aims and scope

What the present study aims to contribute, regarding the list of features in table 1, is an expansion of our knowledge of EFL learners' usage of these features. The studies which have already dealt with this topic have considered learners with different first languages. The present study will use this list of features and find out how Norwegian learners use them, compared to native speakers of English and compared to other learner groups. Only the features and specific items which have been investigated in previous studies will be included in this study. Thus, Norwegian learners may be compared to the learners in previous studies.

The main research question in the present study is as follows: **How and to what extent do Norwegian advanced learners of English use features of writer/reader visibility in their writing?** Based on previous research, the hypothesis is that a study of writer/reader visibility features in Norwegian learner writing will show tendencies of overuse of some features, compared to native speakers. However, it remains to be seen to what extent the Norwegian learners overuse these features in the corpora that are investigated in the present study. It also remains to be seen how the use of writer/reader visibility features in Norwegian learner writing relates to that of other learners. These questions will be explored by studying corpora of Norwegian learner writing at an advanced level, in comparison with other corpora.

Advanced learner writing, in the present study, refers to writing at university level. Two Norwegian learner corpora have been used: NICLE, which is a corpus of argumentative learner writing, and N-VESPA, which is a corpus of academic learner writing³. Argumentative writing, in the present study, refers to general argumentative writing which discusses more or less controversial topics that are not related to any particular academic discipline. Academic writing, on the other hand, which may also be argumentative in form, refers to discipline-specific writing related to university-level education. Since the N-VESPA corpus only consists of academic writing in the linguistics discipline, the present study does not claim to present findings about academic writing in general, but only writing in the linguistics discipline.

The present study performs numerous corpus comparisons and often focuses on frequencies and frequency differences. When doing so, it is important not to get lost in the

³ More about the corpora in chapter 3.

vast amounts of available data. Meyer's advice is "to begin the process with a very clear research question in mind, so the analysis involves more than simply 'counting' linguistic features" (Meyer 2002: 136). In the present study, the main research question has been investigated through the use of several sub-questions. These questions, and the methods applied to explore them are presented in the following section.

1.3. Research questions

In order to be able to answer my main research question, I have formulated the following sub-questions:

1. To what extent do Norwegian advanced learners of English overuse features of writer/reader visibility in their writing?

- a) To what extent is the general level of writer/reader visibility higher in Norwegian learner writing than in native speaker student writing?
- b) Does the level of writer/reader visibility seem to be higher in Norwegian learner writing than in the writing of other learner groups?
- b) Which features are significantly overused?

These questions are dealt with in chapter 4, which is referred to as the preliminary study. Here, all the types of features which were listed in table 1 (section 1.1.) are investigated, in terms of their frequency in the Norwegian learner corpora NICLE and N-VESPA, compared to their frequency in the comparable native speaker student corpora LOCNESS and BAWEling. In this part of the study, comparisons with other learner groups are made by drawing on previous research. One of the aims of this preliminary study is to be able to describe the general level of writer/reader visibility in Norwegian learner writing. The hypothesis is that this level will be higher in Norwegian learner writing than in native speaker writing. Based on previous studies, I also hypothesize that the level of writer/reader visibility may be higher in the Norwegian learner group than in other learner groups. This part of the study also aims to identify the most significantly, and most frequently, overused features of writer/reader

visibility. These findings were among the factors which contributed to my choice of features to study more in depth, namely 1st and 2nd person pronouns.

2. What may be the causes of the Norwegian learners' overuse in argumentative and academic writing?

- a) To what extent may the overuse have been caused by transfer from the learners' first language (L1)?
- b) To what extent may the differences between the overuse in argumentative and academic learner writing be due to differences in academic experience?

Previous studies have suggested many different reasons for learners' overuse of writer/reader visibility features. The present study investigates the possibility that L1 transfer partly explains the overuse. The main hypothesis is that overuse of writer/reader visibility may be a general language learner problem, related to the challenges of writing in a foreign language, which is intensified for some learners, due to transfer from the L1. This is investigated through an in-depth study of pronoun frequencies in the EFL writing of learners with different L1s, as well as a comparison of pronoun frequencies in L1 Norwegian and L1 English. Furthermore, since the level of writer/reader visibility was found to be higher in the argumentative genre than in the academic genre, I try to find out to what extent this difference may be due to differences in academic experience.

3. How do learners and native speakers use first and second person pronouns to express different functions of the writer and the reader?

In order to better understand Norwegian learners' use of first and second person pronouns, a qualitative investigation was performed, aiming at identifying different functions of these pronouns. In this investigation, samples from the main corpora were studied manually and the frequencies of different functions were measured. This investigation is presented in chapter 6, and is referred to as the qualitative in-depth study.

1.4. Outline of the thesis

Some of the theoretical background for the present study was given in section 1.1., in order to provide a basis for the description of the aims and scope. Chapter 2 gives more information about theoretical background and previous studies on writer/reader visibility in learner writing and native speaker writing, and also about functions of the writer and the reader. In chapter 3, the corpus comparisons of the present study will be discussed, in terms of aims and research methods. In addition, the main corpora will be described, and the advantages and disadvantages of each corpus will be discussed. Chapter 4-6 present and analyze the results from the corpus comparisons. Chapter 4 considers the preliminary study of writer/reader visibility, and chapter 5 presents the quantitative in-depth study of pronouns, which aims at identifying reasons for the high level of writer/reader visibility in Norwegian learner writing. Chapter 6 considers the qualitative in-depth study of pronouns, which aims at identifying the functions of the writer and the reader in learner and native speaker writing. In chapter 7, all the findings will be summed up and briefly discussed in terms of pedagogical implications, and future research topics will be suggested.

2. Theoretical background and previous studies

In this chapter, previous studies about writer/reader visibility and writer/reader functions are presented, along with studies which do not deal with these topics per se, but which still contribute to our knowledge about the writer and the reader, by for example providing findings about the use of a certain word belonging to a writer/reader visibility category. I have selected the studies based on the topics of the present study, so that the present study may build on the findings from the previous studies mentioned here.

Since the present study aims to describe the general level of W/R (writer/reader) visibility in argumentative and academic learner writing, I will present some studies which cover similar topics. Section 2.1. is an introduction to W/R visibility in the argumentative genre, while section 2.2. presents the actual findings from studies of argumentative learner writing. Similarly, section 2.3. introduces the topic of W/R visibility in the academic genre, and section 2.4. presents relevant findings from studies of academic learner writing. Most of the studies find that overuse of W/R visibility features is very common in learner writing. The suggested reasons for this overuse are presented in section 2.5.

Since the present study also has an in-depth part which focuses on first and second person pronouns, which are some of the most obvious signs of W/R visibility, I have also included some studies which consider pronouns more in detail (in section 2.6.). Some of these studies claim to be studies of pronouns, while other studies focus on the terms “the writer” and “the reader” – but in reality they tend to study the same features, namely 1st and 2nd person pronouns. The purpose of section 2.6. is to give insight into previous research on the pronouns *I*, *we* and *you* in argumentative and academic writing, so that the present study, which studies these three pronouns in depth, can build on the knowledge from previous research. The studies are put together with the intention to create an overall impression of how the different pronouns are used in the two genres, and how the writer and the reader may function in the two genres. Some studies focus on learner writing, while other studies focus on professional writing; in my opinion, both types of studies are relevant with respect to the present study.

2.1. Expected level of writer/reader visibility in argumentative writing

In argumentative writing, the writer often needs to have a voice. Sometimes, it may be sufficient to argue for and against a statement, but most of the time the writer is expected to have an opinion. Tasks given in educational settings usually encourage the writer to take a stance towards the topic of discussion. This applies to the tasks in the sources of argumentative learner writing in the present study, and to ICLE tasks in general. Recski (2004) comments the irony that lies in accusing learners of being too personally involved in texts that are written on the basis of tasks such as those given below. “In such cases, personal references and subjective attitudes are certainly hard to avoid,” Recski argues, and there seems to be no doubt about it – personal opinions are directly requested.

- Some people say that in our modern world, dominated by science, technology and industrialism, there is no longer place for dreaming and imagination. What is YOUR opinion?
- In the 19th century, Victor Hugo said: “How sad it is to think that nature is calling out but humanity refuses to pay heed.” Do YOU think it is still true nowadays?

Since this is the nature of the NICLE tasks, we can only expect the texts to be quite involved, which means that the writer will be visible to some extent. Therefore, we cannot criticize learners for being involved in their argumentative writing, but we can point to any tendency of learners being *more* or *less* involved than native speakers when given similar tasks.

Unfortunately, the LOCNESS information site⁴ (Université catholique de Louvain, Centre for English Corpus Linguistics 2010) does not provide the full task descriptions for this corpus. However, we can see from the discussion topics, and from reading the native speaker texts, that the LOCNESS texts are also involved to some extent. Examples of such topics, and of native speaker students referring to their own opinions, are given below.

Examples of LOCNESS topics:

- Euthanasia
- Capital punishment
- Nuclear power
- Pride or segregation

⁴ See <http://www.uclouvain.be/en-cecl-locness.html>

Examples of writer visibility in LOCNESS:

- Personally, I think there will be some loss of sovereignty for all member states. (LOCNESS)
- I agree with Hirschberg's remarks. (LOCNESS)
- All things considered, I believe that genetic manipulation will be a benefit to modern society and to future societies to come. (LOCNESS)

Based on this, we might expect argumentative writing to have a visible writer, to some extent, and perhaps also a visible reader, as it is often the writer's intention to convince the reader about something. The following section will deal with the extent to which the writer and the reader are visible in learners' and native speakers' argumentative writing.

2.2. Previous research on writer/reader visibility in argumentative learner writing

As the present study uses Norwegian learner writing from ICLE and native speaker student writing from LOCNESS when studying W/R visibility in argumentative writing, the previous studies to be presented here mostly use the same corpora. Thus, when the results are presented in chapter 4, the potential overuse of the different W/R visibility features in Norwegian learner writing can be compared to the overuse among learners with other L1s. ICLE consists of writing by learners with various L1 backgrounds, and the previous studies, which will be discussed in the following sections, deal with several of these learner groups. However, most of the features have not been investigated in *Norwegian* learner writing, so the comparisons in the present study will contribute new information.

The features of W/R visibility presented in table 1 (section 1.1.) have all been considered in previous studies based on learner writing in ICLE. However, the studies do not cover the same learner groups (with regard to the learners' L1), so, even though the studies all show overuse of W/R visibility features among learners of English, the knowledge about learner overuse is fragmented; we know that the overuse varies from learner group to learner group, and also from feature to feature, but we have no complete overview of the actual overuse in each learner group. Nevertheless, the knowledge that we do have from previous

research is indeed valuable and will be considered as thoroughly as possible in the following sections. These sections will deal with one feature of W/R visibility at a time, referring to the relevant studies. Thus, the more wide-ranging studies will be referred to several times, and the more specific studies of a certain feature will be taken into account when that feature is being discussed.

Petch-Tyson (1998) is one of the more wide-ranging studies of W/R visibility features, which has been an inspirational source to many later studies of W/R visibility in learner writing. The study is called “Writer/reader visibility in EFL written discourse” and it appears in *Learner English on Computer*, edited by Granger – a book which, as the author of the preface puts it, “(...) is the first book devoted to the idea of collecting a corpus, or computer textual database, of the language produced by foreign language learners: a collection known as a learner corpus” (Leech 1998, xiv). The pioneering feel of the book makes it a natural place to begin when looking into previous research into learners’ overuse of writer/reader references. Overuse is one of the major focus points in the book, and overuse of writer/reader references or involvement, as some choose to call it, is the topic of several articles in it. Of these articles, Petch-Tyson’s covers the widest range of features, including different types of pronouns, fuzziness words⁵, emphatic particles and reference words. However, although the article is wide-ranging, Petch-Tyson calls it a “preliminary study” and emphasizes the need of further research (Petch-Tyson 1998, 117). Because of this, many other studies along with Petch-Tyson’s study form the basis for my own research, and several of these will be referred to in the following sections.

2.2.1. First person singular pronouns

The investigation of 1st person sg (singular) pronouns in Petch-Tyson (1998) include the items *I*, *me*, *my* and *mine*. For some reason Petch-Tyson does not consider the reflexive pronoun *myself*, which has been included in other studies (see Paquot, Hasselgård and Ebeling forthcoming, Coffin and Mayor 2004). Ädel (2008) has chosen to study the same items as Petch-Tyson in the category “first person singular pronouns,” probably because her study is based on, and partly criticizes, Petch-Tyson’s study.

⁵ Fuzziness words have been left out of the present study, since they have not been considered typical features of writer/reader visibility.

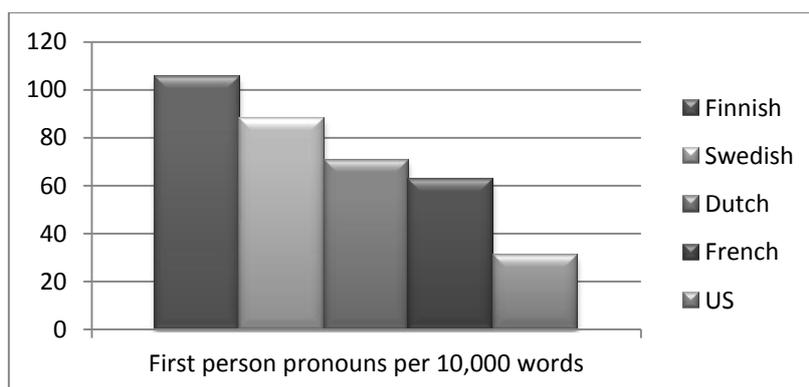
Petch-Tyson finds that different learner groups overuse 1st person pronouns to varying extents. The learners in her study come from four different L1 backgrounds – Dutch, Finnish, French and Swedish. Although all these learner groups overuse 1st person sg pronouns compared to the native speaker control group, the overuse among Finnish learners is higher than the overuse among French learners, for example. Table 2 shows the raw frequencies of 1st person sg pronouns in the different sub-corpora of ICLE, taken from the original table 8.1 (Petch-Tyson 1998, 112).

Table 2: 1st person sg pronouns, Petch-Tyson (1998)

	Dutch 55,314 words	Finnish 56,910 words	French 58,068 words	Swedish 50,872 words	US 53,990 words
First person singular pronouns	391	599	364	448	167

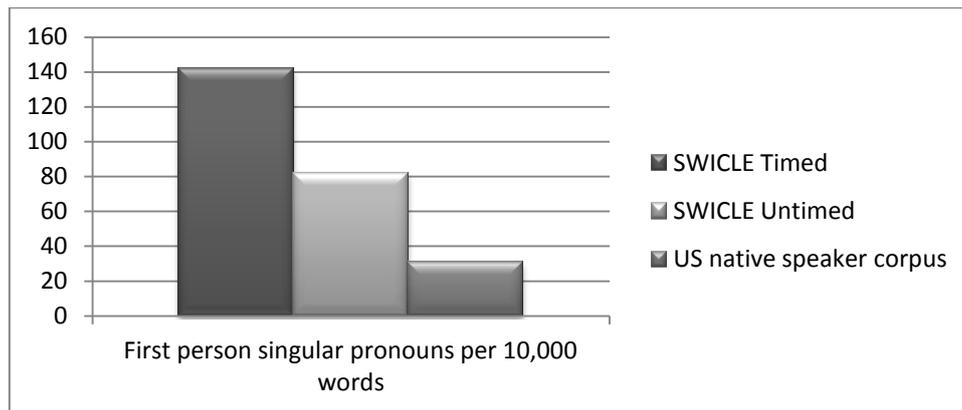
Since these figures are raw frequencies, and since the different corpora do not have the same total amount of words, I have taken the liberty of calculating the frequencies of 1st person sg pronouns per 10,000 words when creating the diagram below. This also makes it easier to relate Petch-Tyson’s findings to other findings which will be present later. As the diagram shows, the Finnish learners have the highest overuse in this study, but all the learner groups have higher frequencies of 1st person sg pronouns than the US control group.

Figure 1: 1st person sg pronouns per 10,000 words, based on table 8.1 in Petch-Tyson (1998)



Ädel's (2008) study "Involvement features in writing" does not aim towards detecting overuse of involvement features, but rather to find out which factors influence learners' register awareness. Such factors are timing – whether or not the students were timed when writing their contributions to the corpus – and interaction during the writing session – meaning access to secondary sources. By comparing texts from three different Swedish corpora of learner English, which are different with respect to these two factors, Ädel finds that learner texts are generally less involved when they are untimed and when the learners have had access to secondary sources. However, the different involvement features are not affected in the same way by these two factors; Ädel sums it up like this: "(i) untimed essay tasks reduce the number of first person pronouns, and (ii) tasks that include input from secondary sources reduce the number of disjuncts, questions and exclamations" (Ädel 2008, 46). So, having access to secondary sources does not affect the frequency of pronouns significantly. The impact that timing has on the use of 1st person sg pronouns is illustrated in the diagram below, showing frequencies per 10,000 words. The SWICLE frequencies are taken from Ädel's table 7 (2008, 45), and the US corpus frequencies are from Petch-Tyson's findings (1998, 112) – added for comparison.

Figure 2: 1st person sg pronouns in SWICLE, based on table 7 in Ädel (2008)



2.2.2. Patterns of subjective stance

Patterns of subjective stance are W/R visibility markers in that they all consist of word patterns including the 1st person sg pronoun *I*. Additionally, they say more about the function of *I* in the text, as they all refer to the writer's personal opinions. Patterns of subjective stance are considered by Hasselgård (2009b) in the article "Thematic choice and expressions of stance in English argumentative texts by Norwegian learners," which is a more recent publication than Petch-Tyson (1998) and may, therefore, be expected to be more up-to-date. The patterns of subjective stance included in Hasselgård (2009b) are *I believe*, *I think*, *I don't think*, *I guess*, *I suppose*, *I would say*, *I would like to say*, *I'm sorry to say*, *I would argue*, *I must emphasize*, *I know*, *I remember* and *I'm not saying*.

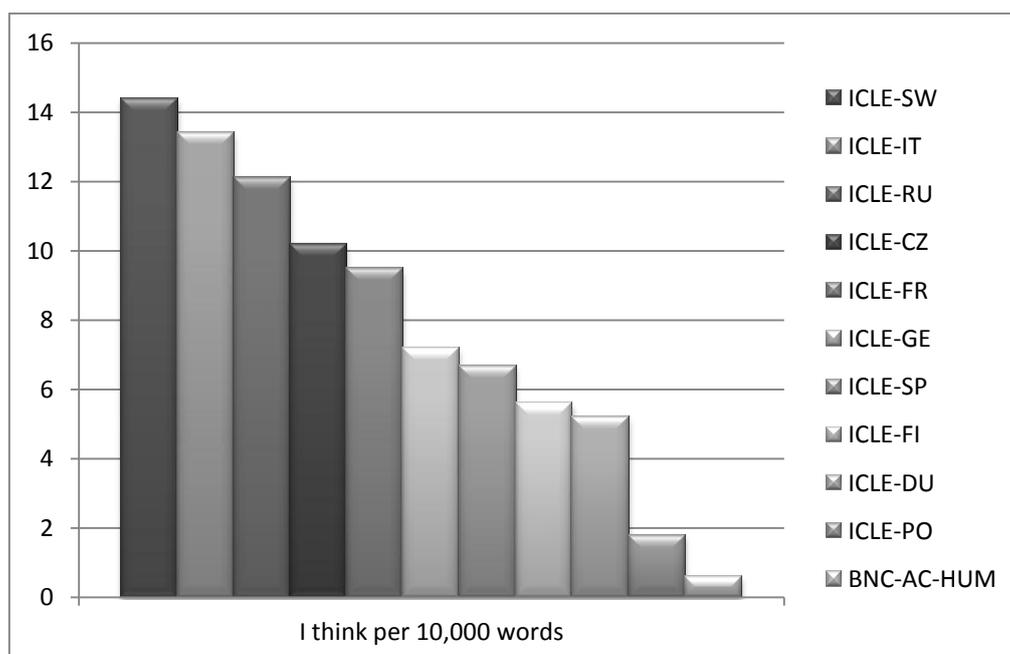
Hasselgård's study is not mainly a study of W/R visibility, but rather a study of thematic choice and expressions of stance. Still, the findings from this study complement Petch-Tyson's study of W/R visibility by including more detailed information about some usage areas of the pronoun *I* and by studying a "new" learner group – Norwegian learners. To complement Hasselgård's study as well, I have included Paquot's findings (2010) about learner overuse of the word pattern *I think*, which includes learners with different L1s, so that we can see tendencies across learner groups.

Starting with Hasselgård (2009b), this is a particularly interesting study because it considers the Norwegian learner group (NICLE), as will also the present study. Hasselgård finds that *I* often co-occurs with certain types of verbs, more specifically "(...) mental and verbal processes (typically 'think' and 'say', respectively)" (2009b, 132). These co-

occurrences, or word patterns, are what she refers to as patterns of subjective stance. The most frequent of these patterns in NICLE are those expressing belief or probability, namely *I believe* and *I think*. Quantitative investigations of these two expressions along with *I would say* and *I don't think* show that “(...) the frequency of subjective stance expressions in NICLE is far higher than in the academic (written) genres of ICE-GB, and approaching the frequency for conversation” (Hasselgård 2009b, 133).

Word patterns similar to those in Hasselgård (2009b) are studied in Paquot (2010) – a study of academic vocabulary in learner writing from ICLE. Paquot finds that learners typically overuse patterns which express personal opinions, like *I think*, *to my mind*, *from my point of view* and *it seems to me* (2010, 151). She studies learners with ten different L1s, and finds overuse among all these learner groups. However, there is a great extent of variability between the groups – the phrase *I think*, for example, was used 14.4 times per 10,000 words by Swedish learners and only 1.8 times by Polish learners. Figure 3, which is based on table 5.17 (Paquot 2010, 154), shows the frequencies per 10,000 words in the ten learner corpora and the control corpus. According to Paquot, all these learner frequencies are significantly higher than the frequencies in the BNC-AC-HUM, which is the humanities and arts section in the academic sub-corpus of the British National Corpus.

Figure 3: *I think* in learner writing, based on table 5.17 in Paquot (2010)

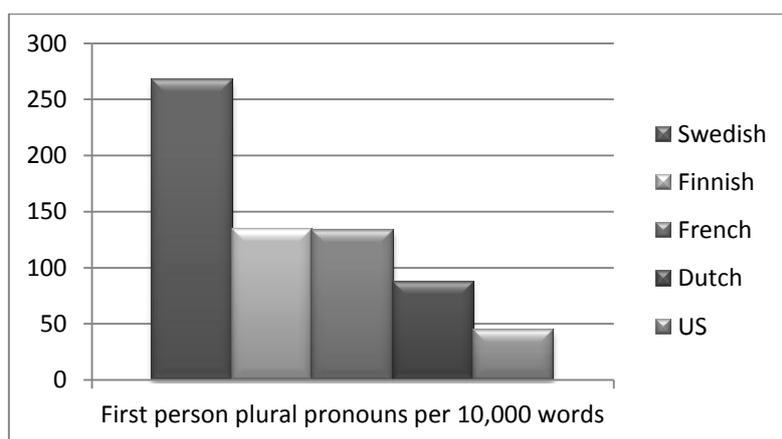


The remarkably high frequency in the Swedish sub-corpus is particularly fascinating when we take into account the findings in Hasselgård (2009b). Her investigation of subjective stance markers, including *I think*, shows that the NICLE frequency is “(...) slightly higher than the corresponding figures from SWICLE” (Hasselgård 2009b, 133). Knowing this, we can assume that if the Norwegian sub-corpus of ICLE was included in Paquot’s study, it would have been ranged somewhere at the top of the list, alongside the Swedish sub-corpus.

2.2.3. First person plural pronouns

1st person pl (plural) pronouns are investigated by Petch-Tyson (1998) in the same way as 1st person sg pronouns, by counting the frequencies of the items *we*, *us*, *our* and *ours* in texts by Dutch, Finnish, French and Swedish learners and US native speakers. Petch-Tyson’s study shows that the overuse of 1st person *plural* pronouns is as variable as the overuse of 1st person *singular* pronouns. However, the distribution of these pronouns is different from the former; with regard to 1st person pl pronouns, the Swedish learner group has the highest frequencies. The Swedish learners use twice as many 1st person pl pronouns as the Finnish and French learners, almost three times as many as the Dutch learners, and six times more than the US students. This is shown in figure 4, which is based on Petch-Tyson’s table 8.1 (1998, 112).

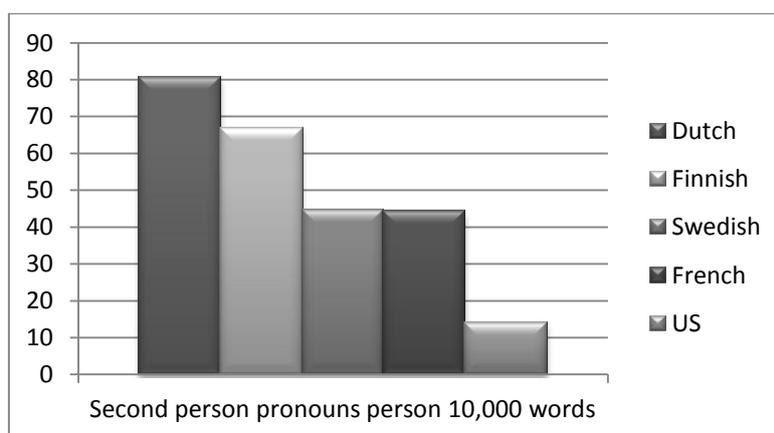
Figure 4: 1st person pl pronouns, based on table 8.1 in Petch-Tyson (1998)



2.2.4. Second person pronouns

“Second person pronouns” in Petch-Tyson (1998) refers to the items *you*, *your* and *yours*. When it comes to these items, the Dutch learners have the highest overuse, compared to French, Finnish and Swedish learners. This is quite the opposite of what was the case for 1st person pronouns, where the Dutch learners had quite low frequencies compared to the other learner groups.

Figure 5: 2nd person pronouns, based on table 8.1 in Petch-Tyson (1998)



2.2.5. Patterns introducing other voices

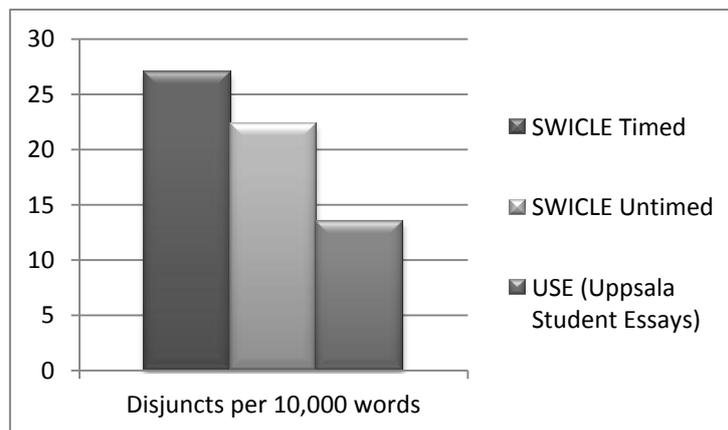
Closely related to the usage of 2nd person pronouns is the usage of “patterns introducing other voices,” referred to in Hasselgård (2009b). Hasselgård’s investigation of some Norwegian learner texts resulted in the identification of these patterns: *You could say that*, *you may say that*, *some may say that*, *some might say that*, *one can say that*, *many think*, *let’s say that*, *you think*, *you could have guessed*, *you may ask* and *I hear you saying*. As we can see from these patterns, the “other voice” is often *you*, but also sometimes *some*, *many*, *one* or *us*. Hasselgård distinguishes between two main functions of “other voices”; they can be used to 1) “introduce a counterargument” and 2) “to represent a pseudo-interaction with the reader” (2009b, 136). Hasselgård’s article does not refer to any quantitative studies of these patterns, but the examples she provides from the NICLE excerpt that she studied manually, make it possible for others to investigate the frequencies of these examples in larger corpora.

2.2.6. Disjuncts

Unlike Petch-Tyson, Ädel has chosen to include disjuncts on her list of involvement features. “The list of disjuncts includes hedges (*perhaps, maybe*), emphatic markers (*obviously, naturally, of course*) and opinion markers (*frankly, unfortunately*), displaying the speaker/writer’s degree of certainty, and her attitude and stance taken towards what is said,” Ädel comments (2008, 37). Since Ädel’s study aims to find out what affects learners’ use and overuse of different involvement features, the article does not compare learner writing to native speaker writing. Therefore, I have included an additional study of the disjunct *maybe* in Paquot (2010), which compares learner writing to native speaker student writing, as well as expert academic writing and native speaker speech, just to have a source which shows potential for overuse of disjuncts among learner writers.

Instead of comparing learners to native speakers, Ädel compares different groups of Swedish learners, to look for differences which might be related to setting variables. As mentioned earlier (in section 2.2.1), some involvement features are affected by timing and other features are affected by available resources. It is mainly the access to secondary sources that helps reduce the frequency of disjuncts in Swedish learner texts. This is illustrated in figure 6, which is based on Ädel’s table 7 (2008, 45).

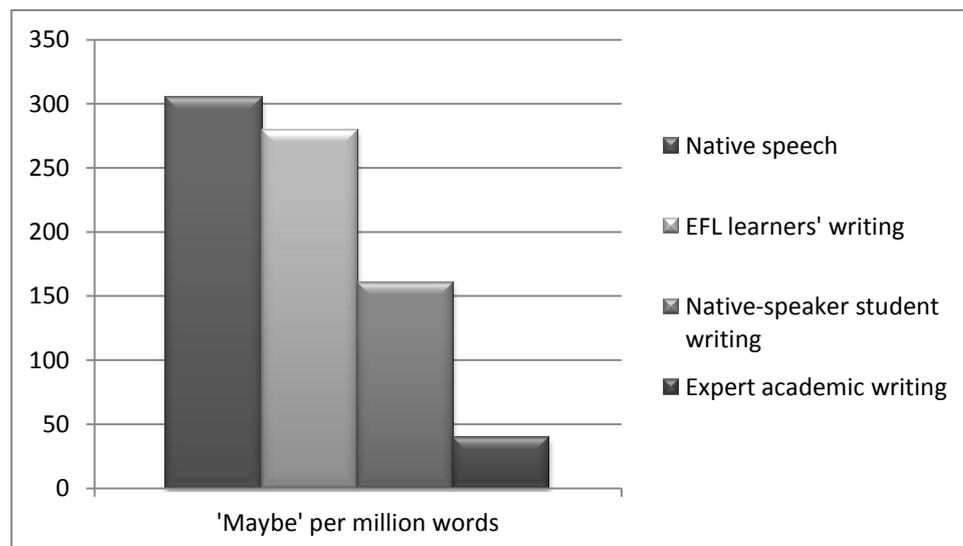
Figure 6: Disjuncts in Swedish learner writing, based on table 7 in Ädel (2008)



Ädel comments that the difference between the two SWICLE corpora is not statistically significant, whereas the difference between the SWICLE corpora and the USE corpus *is* significant. Since USE only differs from the two other corpora in terms of access to secondary sources, it appears as if this access is the relevant impact factor when it comes to learners' (over)use of disjuncts.

Paquot's study (2010) of academic vocabulary includes a study of *maybe* which might be of interest here, as *maybe* is also one of the disjuncts in Ädel's study. I have reconstructed a figure from Paquot's study (figure 7 below), showing the frequencies of *maybe* per million words – a case where learner writing clearly is closer to native speaker speech than to expert academic writing (Paquot 2010, 196). The figure also shows that learners overuse *maybe*, compared to native speaker students. (Note that the frequencies here are normalized per million words, like Paquot chose to portray them, and not per 10,000 words as in the figures above.)

Figure 7: *Maybe* in learner writing, based on figure 5.7 in Paquot (2010)



2.2.7. Emphatic particles

Petch-Tyson (1998) also includes a study of what she terms *emphatic particles*. In her study she only considers two items belonging to this group, namely *just* and *really*, and she does not explain the reasons for her choice of examples, besides noting that “the list is by no means

exhaustive and is merely intended to pave the way for a more comprehensive analysis” (Petch-Tyson 1998, 110). Because the computer-based corpus method is still a relatively new addition to the field of linguistics and second language acquisition research, this is, unfortunately, the case in most corpus studies; that is, most studies only explore a field; they sometimes study what seems to be random examples; they often find out that their corpora are inadequate; and they almost always refer to the intention of wanting to encourage more extensive studies. This is, of course, natural in the onset of a new methodology, but it means that, when referring to previous research, there are not many firm conclusions to refer to. Rather, we can refer to suggestions and non-exhaustive findings – like the results from Petch-Tyson’s study of *just* and *really* presented in figure 8 and 9, which are based on the original table 8.1 (Petch-Tyson 1998, 112).

Figure 8: *Just* in different learner groups, based on table 8.1 in Petch-Tyson (1998)

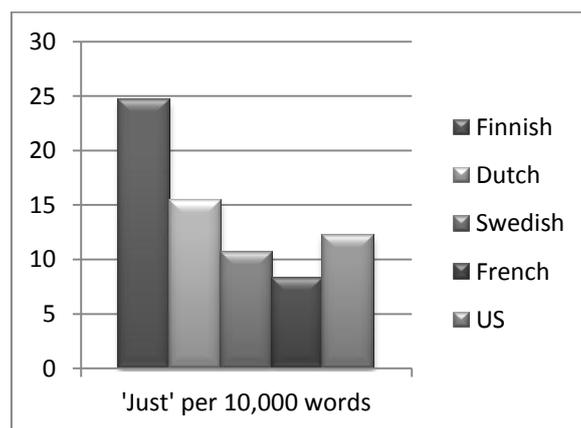
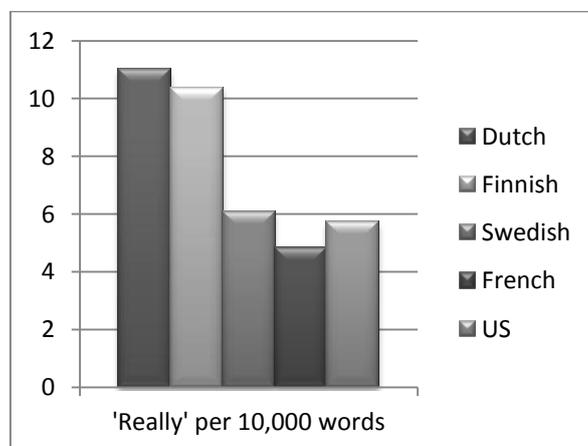


Figure 9: *Really* in different learner groups, based on table 8.1 in Petch-Tyson (1998)



These figures may not tell us anything certain about learner's use of emphatic particles in general, since they only cover the use of two examples from what we can imagine is quite a long list of words. However, if studies of other emphatic particles would show approximately the same results as these, there might be a tendency among Finnish and Dutch learners in particular to overuse emphatic particles. As the figures show, Finnish learners clearly overuse both *just* and *really*, and Dutch learners clearly overuse *really*, while the Swedish and French learners do not seem to overuse any of these words, compared to the US control group.

2.2.8. Reference to situation of writing/reading

The items explored in Petch-Tyson's study of words which refer to the situation of writing/reading are *here*, *now* and *this essay*. Because of the low raw frequencies of *this essay*, I have only included *here* and *now* in figures 10 and 11, which are based on table 8.1 (Petch-Tyson 1998, 112).

Figure 10: *Here* in different learner groups, based on table 8.1 in Petch-Tyson (1998)

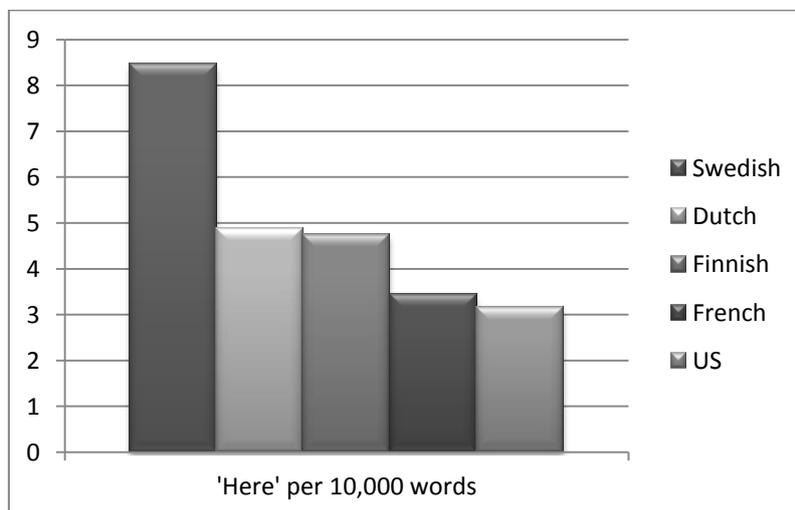
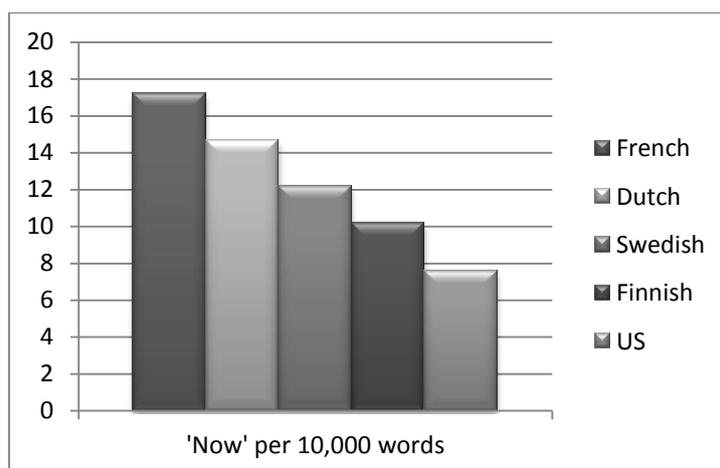


Figure 11: *Now* in different learner groups, based on table 8.1 in Petch-Tyson (1998)



Once again, the frequencies are variable across learner groups. What stands out the most in these figures, is the overuse of *here* among Swedish learners, which is significantly higher than in any other group in this study. This high degree of overuse might be caused by transfer from the learners' L1, as we shall see in section 2.5. As for the investigation of *now*, figure 11 shows quite varying results in the different learner groups; the French learners have the highest overuse, while the Swedish learners have the lowest overuse, but all learners have higher frequencies of *now* than the US control group.

2.2.9. Sentence types

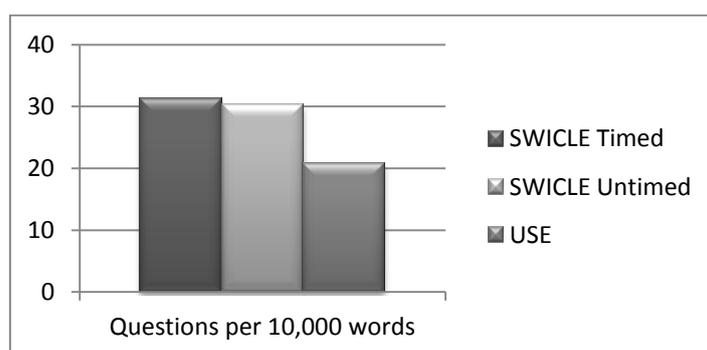
Ädel (2008) considers two types of sentences which have a high degree of W/R visibility. First of all, there are questions, which are included because they "(...) tend to be explicitly audience-oriented and generally contribute to making a discourse more interactive" (Ädel 2008, 36). Then, there are exclamations, which are either used as "addressee-oriented directives" (ex. 1) or used to express "the extent to which she [the writer] is impressed by something" (ex. 2) (Ädel 2008, 36-37). Both usages make the writer and/or the reader visible to some extent.

1. Terrible, I know! (SWICLE)
2. But let's hope for peace in Europe! (SWICLE)

(Ädel 2008, 36)

By comparing different groups of Swedish learners, Ädel finds that having access to secondary sources when writing in English helps Swedish learners reduce the number of questions in their texts. Figure 12 shows how the USE learners, who had access to secondary sources, use a significantly smaller amount of questions than the SWICLE learners, who did not have access to secondary sources (Ädel 2008, 45). The timing factor, on the other hand, does not seem to have any effect on the use of questions, as the two SWICLE corpora have similar frequencies of questions. The same tendencies were found when looking at exclamations – the only difference being that the USE writers did not use any exclamations at all, which is why I did not illustrate this finding with a figure.

Figure 12: Questions in Swedish learner writing, based on table 7 in Ädel (2008)



Since Ädel does not compare learners to native speakers, I have found an additional source to make sure that there are tendencies of overuse of certain sentence types. I could not find a study which specialized in learners' overuse of exclamations, but I found a study which considers questions in learner writing. Virtanen (1998) compares the use of questions in different sub-corpora of the ICLE and in LOCNESS, and finds that, as a whole, the ICLE learners overuse questions when compared to the native speaker students in LOCNESS (see figure 13). However, one learner group – the L1 Spanish group – uses fewer questions than the native speakers. Moreover, the rest of the learner groups also show variable frequencies. The differences between the learner groups are shown in figure 14. Both figures below are based on Virtanen's table 7.2 (1998, 97).

Figure 13: Questions in learner writing, based on table 7.2 in Virtanen (1998)

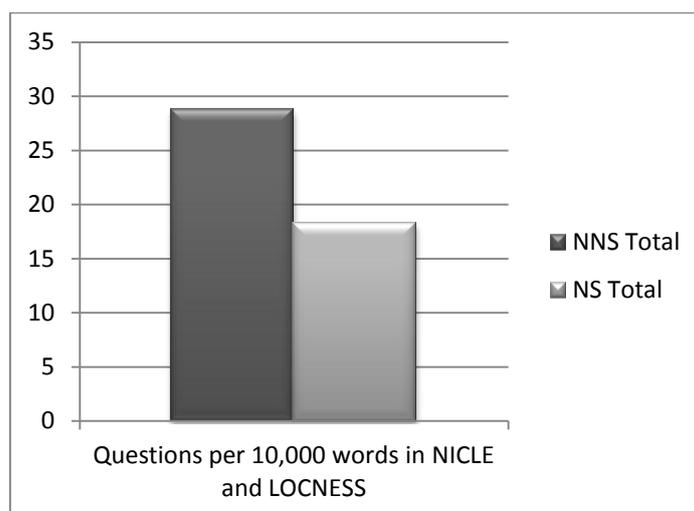
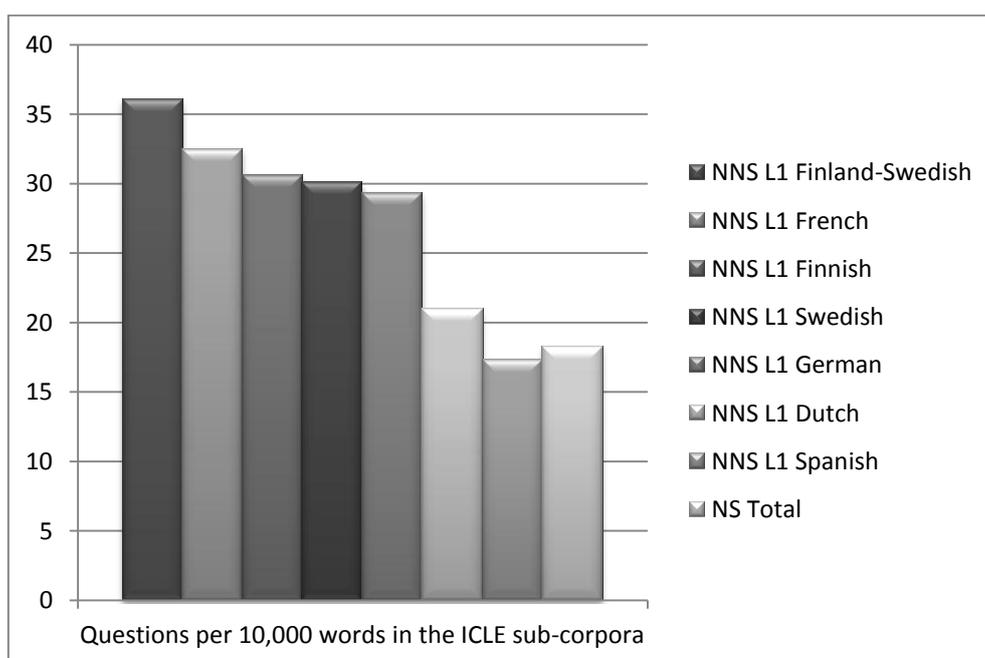


Figure 14: Questions in different learner groups, based on table 7.2 in Virtanen (1998)



Commenting on the high frequency of questions in the Finland-Swedish essays, Virtanen is concerned that “[a]n overuse of questions can reduce their argumentative value and increase the often more informal style of their writing” (Virtanen 1998, 105). It is good to know, then, that there is hope for the learners if they only have access to the right resources, as shown in Ädel’s study.

2.2.10. Summing up writer/reader visibility in argumentative learner writing

What the previous studies can tell us about learners' overuse of W/R visibility features, in the argumentative genre, is that there seems to be a general tendency towards overuse. However, this overuse varies from feature to feature, and not all learner groups overuse all the features.

Starting with the most direct W/R references, 1st and 2nd person pronouns were found to be overused by all learner groups in Petch-Tyson (1998). However, the degree of overuse varied both according to learner group and type of pronoun. Patterns of subjective stance were found to be highly overused in Norwegian learner writing (Hasselgård 2009b), and the findings in Paquot (2010) suggest that many other learner groups also tend to overuse such patterns, although there is great variation between the learner groups. With regard to patterns introducing other voices, which were identified in Hasselgård (2009b), there is no quantitative information about these patterns in any of the studies mentioned.

Continuing with the, arguably, more subtle W/R references, disjuncts may be overused by learners. At least an investigation of *maybe* in Paquot (2010) showed overuse in learner writing compared to native speaker writing. Moreover, the emphatic particles *just* and *really* were found to be overused in Dutch and Finnish learner writing, but not in Swedish and French learner writing (Petch-Tyson 1998). The items *here* and *now*, which are used refer to the situation of writing/reading, were shown to be overused to varying extents by learners with different L1s (Petch-Tyson 1998). A study by Virtanen (1998) showed that questions are also overused by many learners, and that the learner overuse varies according to the learners L1 in this case as well.

Finally, Ädel adds to our knowledge about learner overuse of W/R visibility features by concluding that learner overuse of pronouns can be reduced if the learners are not timed, and that their overuse of disjuncts, questions and exclamations can be reduced if they are given access to secondary sources.

2.3. Expected level of writer/reader visibility in academic writing

The role of the writer and the reader in academic writing is not the same as in argumentative writing, at least in the way that these genres are represented in the present study. In the academic world, several perceptions exist of how the writer should behave, and also whether or not he/she should refer to himself/herself, as well as to the reader. The traditional view is

formulated in the citation below, which is taken from a textbook about academic writing and reading.

In general, academic writing aims at being ‘objective’ in its expression of ideas, and thus tries to avoid specific reference to personal opinions. Your academic writing should imitate this style by eliminating first person pronouns ... as far as possible.

(Arnaudet and Barrett 1984, 73, cited in Hyland 2002b, 351)

There are at least two problems with the advice given in this book; the first problem is pointed out by Hyland (2002b) who argues that this view is oversimplifying, since it tries to consider academic writing as one type of writing. He emphasizes the variety between academic disciplines when it comes to the role of the writer. For example, a study of the use of personal pronouns in academic texts across disciplines showed that “[b]roadly, writers in the hard sciences and engineering prefer to downplay their personal role to highlight the issue under study, while a stronger identity is claimed in the humanities and social sciences papers” (Hyland 2002b, 352). Adding to this knowledge, Fløttum (2006) found that, in the linguistics discipline, which is of particular interest for the present study, the writer is more visible than in the medical discipline, where writers tend to use passive constructions and less personal phrases instead of direct writer references like personal pronouns. By comparing scientific articles from different disciplines and in different languages, the study showed that writer identity is more closely related to discipline than to language. It was found that Norwegian and French medical articles were more similar in form than Norwegian medical articles and Norwegian linguistics articles.

Another problem with the above statement is that it might be outdated. Although it might have been considered wise, in the past, not to refer to personal opinions in academic texts and to avoid certain pronouns, this may not be the academic norm today. McCrostie points to the recent trend in academic writing “(...) towards an increase in the use of personal pronouns, self-mention, and other informal features in nearly every academic discipline, a fact reflected in many style manuals published after the 1980’s” (McCrostie 2008, 107). So what we can expect, with regard to the general level of W/R visibility in academic writing by students, is perhaps that they behave differently depending on the instructions they have been given in school and at university. Those who have been taught the rules of such textbooks as

the one cited above might have quite low degrees of W/R visibility in their academic writing, because they have aimed at objectivity. In contrast, those who have been taught the newer trend, which implies more self-reference and a less formal style, might have quite high degrees of W/R visibility in their academic writing, because they aim at a more personal style. Differences in instruction might be one of the explanatory factors with regard to the differences in the academic writing by students in the studies which are presented in the following section.

2.4. Previous research on writer/reader visibility in academic learner writing

While the ICLE project has resulted in numerous studies on argumentative learner writing, which complement each other in that they are based on the same corpus and often focused on similar matters (but with different perspectives), there does not seem to be any similar research project on the field of academic learner writing. The Varieties of English for Specific Purposes dAtabase (VESPA) project is perhaps an attempt to fill this gap. At least the initiators of the VESPA project aim to collect academic texts from learners with different L1 backgrounds, from different disciplines and with different levels of writer expertise, which sounds similar to the ICLE project, only even more wide-ranging (Paquot and de Cock 2011). Since the VESPA corpus is still a work in progress, not much research has been published about it, or based on it. However, two studies on the current VESPA material (focusing on French and Norwegian learner writing) will be referred to in the following chapters, and, to add to this, two studies based on other sources of learner writing (by Chinese, Greek and Hong Kong learners) will be presented.

2.4.1. Underuse of academic vocabulary in learner writing

Granger and Paquot (2009) is not a study of W/R visibility, but can still contribute to our understanding of W/R visibility in learner writing. Since we know from previous studies that learners tend to *overuse* informal features, such as W/R references, it might be interesting to know if they *underuse* more formal features, such as academic vocabulary. By comparing the most frequent words in professional scientific articles and academic writing by French

learners of English in VESPA, Granger and Paquot are able to find out which words the learners underuse. They find that more than half of the verbs which are underused by learners belong to a list of key academic vocabulary (Granger and Paquot 2009, 6). Examples of such words are listed below:

- Achieve
- Contribute
- Demonstrate
- Establish
- Examine
- Identify
- Provide
- Report
- Suggest
- Support

(Granger and Paquot 2009, 6)

We can relate the underuse of academic words to the overuse of features of W/R visibility, because both these phenomena contribute to making the learners' texts more informal and interpersonal, and both can be related to learners' lack of register awareness. However, no specific conclusions about W/R visibility in academic learner writing can be drawn on the basis of this study.

2.4.2. Overuse of writer/reader visibility features in academic (and argumentative) writing by Norwegian and French learners

Paquot, Hasselgård and Ebeling (forthcoming) investigate a list of linguistic features, covering some of the same features as the ICLE studies in the former chapter. The features are presented in table 3, along with the frequencies per 100,000 words in each of the corpora. The three corpora to the left consist of argumentative writing from respectively the French part of ICLE, the Norwegian part of ICLE and the native speaker corpus LOCNESS. The three corpora to the right consist of academic writing from the French part of VESPA, the

Norwegian part of VESPA and the native speaker corpus BAWE (British Academic Written English)⁶.

Table 3: W/R visibility in French and Norwegian learner writing, Paquot, Hasselgård and Ebeling (forthcoming)

	ICLE-FR	ICLE-NO	LOCNESS	VESPA-FR	VESPA-NO	BAWE
1 st person pronouns ⁷	1898	2550	791	1178	1366	575
2 nd person pronouns ⁸	344	797	227	41	112	29
<i>Let's</i> imperatives	72	15	5	9	8	1
Epistemic modal adverbs						
<i>Certainly</i>	52	24	9	10	4	6
<i>Maybe</i>	20	93	16	14	19	2
<i>Of course</i>	73	66	13	14	18	5
<i>Probably</i>	32	75	19	22	51	9
<i>Perhaps</i>	41	34	25	9	39	24
Questions	364	387	184	51	55	33

Some of the conclusions which were reached after comparing frequencies across learner groups and genres are that there are “generally fewer W/R visibility markers in academic (discipline-specific) student texts (VESPA+BAWE) than in argumentative essays (ICLE+LOCNESS),” although learners tend to overuse W/R visibility markers in both genres (Paquot, Hasselgård and Ebeling forthcoming). As we can see from the table, there are also some differences between the French and the Norwegian learners; the Norwegian learners, for example, overuse 1st and 2nd person pronouns to a greater extent than the French learners. The French learners, on the other hand, overuse *let's*-imperatives to a greater extent. The overuse of questions seems to be of similar degrees in the two learner groups, whereas the overuse of adverbs differs from one adverb to the other.

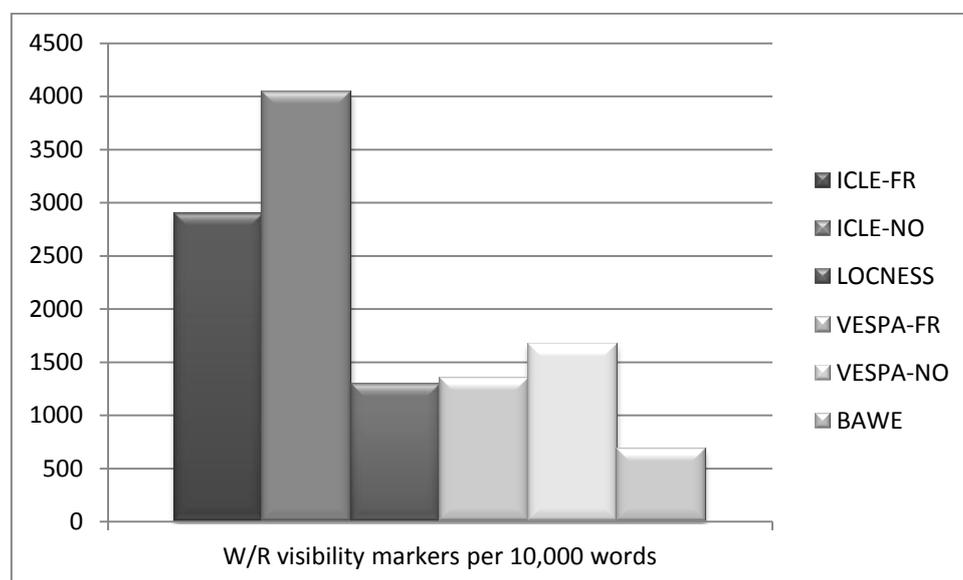
⁶ For a description of these corpora, see chapter 3.

⁷ Including the personal pronouns *I, me, we* and *us*, and the reflexive pronouns *myself* and *ourselves*.

⁸ Including the personal pronoun *you* and the reflexive pronouns *yourself* and *yourselves*.

If we were to sum up all these frequencies, we would see that Norwegian learners, in general, use W/R visibility features to a greater extent than French learners in both genres. Figure 15 shows the total frequencies of such features in all the corpora. Note that the list of features is not exhaustive, so the total frequencies do not account for all possible features – only the ones in Paquot, Hasselgård and Ebeling’s study. The idea to sum up the frequencies was my own, so any potential problematic issues concerning this are all on me. What this figure illustrates is the great difference between the argumentative corpora and the academic corpora, and the tendency of Norwegian learners’ to overuse W/R visibility features to a greater extent than French learners.

Figure 15: W/R visibility features in total per 100,000 words, based on Paquot, Hasselgård and Ebeling (forthcoming)



2.4.3. Underuse of personal reference in academic writing by Hong Kong learners

Hyland (2002a) investigates the use of personal reference in learner writing by Hong Kong students. Although this study is not defined as a study of W/R visibility, it contributes to our knowledge about this topic, especially with respect to writer visibility. Hyland focuses on the use of 1st person pronouns and determiners⁹ to express authorial identity in academic writing

⁹ Hyland includes the singular and plural forms *I, me, we* and *us*, as well as the possessive pronouns/determiners *my* and *our*.

by novice and professional writers¹⁰. The corpora of novice (student) writing and professional writing cover several academic disciplines. The novice corpus consists of 64 project reports written by Hong Kong learners of English, whereas the professional corpus consists of 240 published research articles. By comparing the two corpora, Hyland finds that professional writers use 1st person pronouns four times more often than Hong Kong novice writers. The frequencies are shown in table 4.

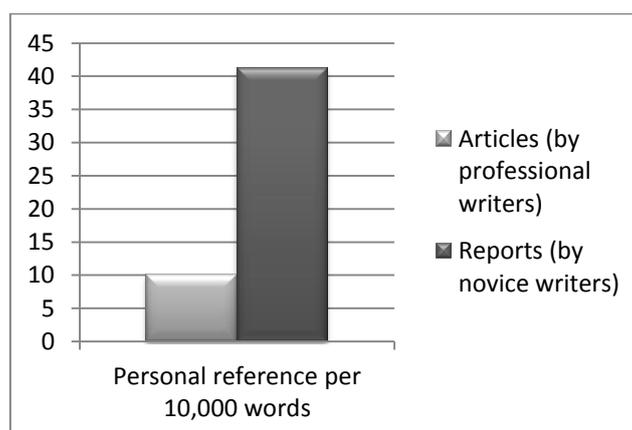
Table 4: Personal reference¹¹ per 10,000 words in novice and professional writing, Hyland (2002a)

Field	Totals		Singular references		Plural references	
	Articles	Reports	Articles	Reports	Articles	Reports
Science and Engineering	32.7	9.4	0.1	4.9	30.6	4.5
Business and Professional	46.9	10.5	22.2	6.7	24.7	3.8
Overall	41.2	10.1	14.4	6.1	26.8	4.1

(Hyland 2002a, 1099)

As table 4 shows, students underuse 1st person pronouns and determiners in most cases. There is one exception, however; in the science and engineering field students overuse singular references. Still, we can see from the frequencies in the left column that the total underuse is significant in both fields. The total frequencies of 1st person pronouns in novice and professional writing are illustrated in figure 16, which is directly based on the findings we saw in table 4 (in the present study).

Figure 16: Personal reference per 10,000 words in novice and professional writing, based on table 3 in Hyland (2002a)

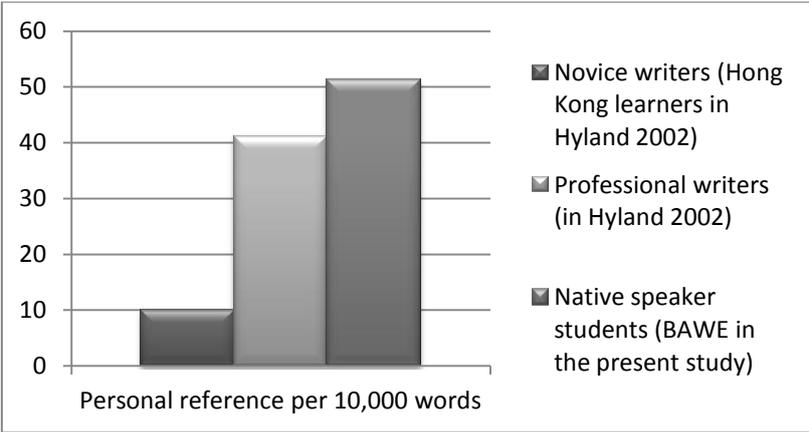


¹⁰ It is unclear whether the professional writers are native speakers of English or not.

¹¹ First person pronouns and possessive pronouns/determiners.

The problem about Hyland’s study, with regard to the present study, is that it is mainly a comparison of novice and professional writing, and not a comparison of learner and native speaker writing, even though the novice writers are learners who write in their L2. Therefore, to find out whether Hyland’s findings are relevant to the present study, I have chosen to compare Hyland’s findings to some of my own findings – that is, the frequency of self-reference (which is defined in the same way here as in Hyland) in the BAWE corpus. I have used the whole BAWE corpus (L1 English) in order to include a variety of disciplines, since Hyland’s sources also represent many disciplines. The writers in BAWE are all native speaker students, and should constitute a valid control group as they are all university students, and the Hong Kong learners are defined as undergraduate students.

Figure 17: Personal reference per 10,000 words in learner (novice) writing, professional writing and native speaker student (novice) writing, partly based on table 3 in Hyland (2002a)



As we can see from figure 17, which compares the Hong Kong learners to the native speaker students in BAWE (and also to professional writers), the tendency towards underuse of self-reference in Hong Kong learner writing seems to be present in this comparison as well. This is interesting to the present study, as all other studies mentioned here show the opposite tendencies in their learner groups.

2.4.4. Overuse of personal reference in academic writing by Chinese and Greek learners

Coffin and Mayor's study (2004) of the writer and the reader in academic learner writing is in some ways similar to Hyland's study, but the findings are the opposite – they detect overuse of personal reference, instead of underuse. In Coffin and Mayor, learners with Chinese and Greek L1 backgrounds, who are considered novice writers, are compared to professional academics, as was also the case in Hyland (2002a). The features which are studied are also similar to the ones in Hyland's study, as they focus on personal reference. However, in this study, references to the writer *and* the reader are included. When comparing frequencies of first and second person references¹² in corpora of novice L2 writing and professional academic writing, Coffin and Mayor find overuse among novice writers. The frequencies per 10,000 words in each corpus are shown in table 5, which is a modified version of the original table (Coffin and Mayor 2004, 249).

Table 5: 1st and 2nd person references per 10,000 words in learner writing and academic writing, based on table 3 in Coffin and Mayor (2004)

	1 st person singular	2 nd person	1 st person plural
Biber <i>et al</i> academic prose corpus: average across disciplines	25	10	40
Chinese L1 scripts	94	62	237
Greek L1 scripts	58	27	234

Coffin and Mayor's study is slightly irrelevant in itself, with regard to the present study, since it does not specify that the learners are compared to native speakers, but rather focuses on the comparison of novice writers to professional writers, as was also the case in Hyland (2002a). However, the findings from the learner texts may still be interesting to the present study. Therefore, I would like to compare the Chinese and Greek learners to native speaker students, like I did with the Hong Kong learners in Hyland.

¹² Here, first and second person references are defined as personal pronouns, possessive pronouns/determiners and reflexive pronouns, which includes the items *I, me, my, mine, myself* for 1st person sg reference, for example.

However, there is an additional problem with Coffin and Mayor's study, related to the definition of genre, which should be sorted out before choosing an appropriate control group. Coffin and Mayor claim to have studied academic writing, but the material they have used are actually argumentative essays, which the students were only given 40 minutes to complete. Even though these essays were part of a so-called academic writing task, there are no discipline-specific elements to the tasks, and the genre seems to be more closely related to the argumentative genre as it is defined in the present study, and as it is also found in ICLE and LOCNESS. Therefore, I chose to compare the results from Coffin and Mayor's study to the LOCNESS frequencies in the present study, this time including the frequencies of reflexive pronouns to make the findings more comparable.

Table 6: 1st and 2nd person references per 10,000 words in learner writing (in Coffin and Mayor 2004) and student writing in LOCNESS

	1st person singular	2nd person	1st person plural
Native speaker student writing in LOCNESS	42	21	55
Chinese L1 scripts	94	62	237
Greek L1 scripts	58	27	234

As shown in table 6, there are still clear indications of overuse in the learner corpora, particularly in Chinese learner writing, and in the 1st person pl category for both learner groups. However, the degree of overuse is lower when the learners are compared to native speaker students, instead of professionals. The remaining overuse may partly be caused by the fact that the learners were only given a limited amount of time to write their essays; as shown by Ädel (2008) learner overuse of pronouns is reduced when the learners are not timed. Noting that the comparison of timed learners to native speaker students, who may or may not have been timed, is slightly unfair, it seems quite strict of Coffin and Mayor to compare a corpus of timed learners' writing in the argumentative genre to a corpus of academic prose at

a professional level. Still, their findings are useful, since we again find tendencies of overuse of W/R visibility features in learner writing.

2.4.5. Summing up writer/reader visibility in academic learner writing

The previous studies of learner writing in the academic genre, have shown that although many learners seem to have a high level of W/R visibility, compared to native speakers, some learners may also have lower levels of W/R visibility than native speakers. The learners who were found to underuse W/R visibility features – at least personal pronouns and possessive pronouns/determiners – were the Hong Kong learners in Hyland's study. The Norwegian and French learners in Paquot, Hasselgård and Ebeling (forthcoming) showed the opposite tendency, namely overuse compared to native speakers, both with respect to personal pronouns and other features of W/R visibility. Granger and Paquot (2009) showed that in addition to overusing features of W/R visibility, French learners tend to underuse academic vocabulary, which means that their texts are less academic, and more informal, in two different ways. The final previous study to be mentioned here was Coffin and Mayor's study of academic learner writing. However, I found the genre in the learner writing to be more similar to the argumentative genre, according to the definitions used in the present study. In any case, the learners in Coffin and Mayor were found to overuse personal reference, which correlates well with all the other previous studies, except Hyland's study, which stands alone as an investigation of underuse of personal reference in learner writing.

2.5. Suggested reasons for overuse and underuse of writer/reader visibility features

The previous studies suggest some possible reasons for the overuse of W/R visibility features in learner writing, such as general language learning issues, transfer from the learners' L1, lack of register awareness, issues related to the tasks and the setting in which the texts were written, and differences in instruction. The following sections will present the arguments for each of these possible explanations of learner overuse. The underuse found in Hong Kong learner writing, might be related to some of the same issues, particularly differences in instruction, if the case is that the EFL/ESL (English as a foreign/second language) instruction

in Hong Kong focuses more on objectivity than the instruction in the other countries which are represented in this chapter.

2.5.1. General language learning issues

Learners seem to experience difficulties in adapting to the detached style of formal writing, and, consequently, use features of W/R visibility more often than what is expected in the argumentative and academic genres at university level. According to Ädel (2008), this might be related to the very nature of communicative language learning: “Features of involvement are learnt first, while features of detachment are adopted at a later stage” (2008, 37).

Therefore, it is only natural that advanced learners, who are proficient users of English in informal settings, have problems when applying their language skills in more formal settings, particularly with respect to personal involvement or W/R visibility.

Another argument which speaks for the theory that general language learning issues might have caused some of the overuse in learner writing is found in Virtanen (1998). When discussing learners’ overuse of questions, Virtanen suggests that, in addition to task-related issues, “(...) the fact that they are writing in a foreign language, [is] bound to have an effect upon the confidence with which they address and perform the writing task and it is this lack of confidence and reluctance to assert themselves which may well result in a tendency to ask questions rather than make statements” (1998, 94).

Although Virtanen’s comment only concerns the learners’ frequent use of questions, we might imagine that the same lack of confidence could be the cause of some of the uses of other W/R visibility features as well. For example, the most highly overused patterns of subjective stance in Hasselgård (2009b) are *I think* and *I believe*, which may be signs of lack of confidence rather than assertiveness. Some of the adverbs in Paquot, Hasselgård and Ebeling (forthcoming) also express lack of confidence, such as *maybe*, *probably* and *perhaps*, which are overused by Norwegian learners in both argumentative and academic writing, and by French learners in some cases. If this apparent lack of confidence is partially caused by the fact that the learners are writing in a foreign language, which is what Virtanen suggests, then the high level of W/R visibility in learner writing might indeed be partially related to general language learning issues.

2.5.2. L1 transfer

In addition to general language learning issues, the learners' high level of W/R visibility, compared to native speakers, may partially have been caused by transfer from the learners' L1. Whether due to L1 transfer or other reasons which may be specific to one or several learner groups, the difference between the learner groups implies that there are other explanatory reasons for the overuse of W/R visibility features, in addition to general language learning issues. None of the studies referred to in this chapter explore this possibility further, however.

2.5.3. Lack of register awareness

Hasselgård finds that the frequency of subjective stance markers in Norwegian learner writing, in addition to being far higher than the frequency in native speaker writing by students and academics, approaches the frequency in English native speaker conversation (2009b, 133). Hasselgård does not make any firm conclusions based on these findings, but she tentatively suggests that the overuse of subjective stance markers may have been caused by influence from native speaker conversation. This could imply that learners are not aware of the register differences between the language of conversation and the language of argumentative/academic writing.

Paquot also finds that learners often overuse informal words and word patterns to the extent that their frequencies are closer to native speaker speech than to native speaker academic prose (2010, 193). Paquot refers to many studies which have commented on learners' lack of register awareness (Granger and Rayson 1998, Lorenz 1999b, Altenberg and Tapper 1998, Meunier 2000 and Ädel 2006 in Paquot 2010, 150). However, what Paquot's study showed, was that native speaker students also overuse many informal speech-like features in their writing, although not to the same extent as EFL learners. Thus, she concludes that "(...) as a general rule, the findings suggest that the main feature shared by native and non-native novice writers is a lack of register awareness" (Paquot 2010, 195). It seems, then, that the lack of register awareness in learner writing may be caused by the fact the learners are novice writers, combined with the fact that they are writing in a foreign language.

2.5.4. Tasks and setting

As already mentioned, setting issues like timing and access to secondary sources have been pointed out as important sources of impact on the learners' level of W/R visibility (Ädel 2008). An additional setting issue, which is particularly relevant with regard to learner writing in ICLE, is pointed out by Virtanen: "The writing process is also affected by the unnatural setting, the ill-defined or fictional audience and the teaching / testing purpose of the task" (1998, 94). Virtanen also points out that if the topic of the task is not interesting to the students, they may not know what to write about. In many ICLE texts, this seems to be the case; not only are the texts often very short, but the writers sometimes specifically refer to the fact that they do not know what to write about. To illustrate this problem, I have selected two examples from NICLE, in which the writers clearly do not know what to write about – the second writer expressing this even more clearly than the first writer. Perhaps the learners' lack of interest in the topics, combined with the unnatural setting, has partially caused the high level of W/R visibility in the learner writing in ICLE.

1. The only college education I've ever done is the one I'm doing here at the teacher training college at the moment, so I have nothing to compare this statement with. (NICLE)
2. I have been sitting by the computer for more than one and a half hour now, trying to write an essay about how science technology and industrialisation have influenced our lives, whether there is still a place for dreaming and imagination. I am beginning to get pretty desperate, as the day will soon be over and I ought to be in bed. But I have so many things I should have done, and now I have been sitting here for more than one and a half hour and come up with nothing! (NICLE)

2.5.5. Differences in instruction

Differences in the writing instruction at schools and universities might have caused the differences between the learners and the native speakers of English, as well as the differences between learner groups. For example, this factor might explain why the Hong Kong learners in Hyland (2002a) differ from the French and Norwegian learners in the VESPA study (Paquot, Hasselgård and Ebeling forthcoming). According to one of the university supervisors in Hyland's study, the invisibility of the author in the student's writing was not caused by

recommendations by university supervisors, but might have been part of the students' previous instructions at school:

Yes, I like to see students use the first person. Their own interpretations are important but often it is difficult to see what is theirs and what is lifted from sources. Maybe this is something to do with how they are taught to write essays at school. They hide themselves. (TESL supervisor)

(Hyland 2002a, 1105)

2.6. Functions of the writer and the reader

According to Hyland, “[w]hile frequency of occurrence is important in determining the scale of underuse, we can learn a lot more about authorial identity by exploring the rhetorical functions the first person is used to perform” (2002a, 1098). This probably applies to the functions of 1st person pronouns as well as 2nd person pronouns, and I expect that we can learn a lot about learners' identity both in the cases where personal references are underused, and where they are overused.

Since the present study deals with general argumentative writing as well as discipline-specific academic writing, it is important to define the roles of the writer and the reader in the two genres. The following sections will deal with findings about the functions of the pronouns *I*, *we* and *you* in the argumentative and the academic genre.

2.6.1. *I* in argumentative writing

Petch-Tyson (1998) finds that the functions of pronouns are not equally distributed across the learner groups. In the control corpus of native speaker (US) writing “(...) almost half of all occurrences of *I* were with verbs in the past tense, and many of these represented chains which together recounted personal experiences” (1998, 111). Such functions of *I* also occurred quite frequently in the Finnish corpus, and sometimes in the Swedish corpus. In the Dutch and French corpora, on the other hand, there are almost no such instances of *I*. Instead, “(...) a main function of *I* seems to be to talk about the writer within the context of the piece of discourse, either saying something about the writer functioning within the text or what the

writer thinks” (Petch-Tyson 1998, 111-114). The examples below show typical uses of *I* in Dutch, French and Swedish learner texts. According to Petch-Tyson, “[t]his type of use is almost completely absent both from the native American texts and from the Finnish texts” (Petch-Tyson 1998, 114).

- ... I am not of the same opinion... (from the Dutch corpus)
- Through this paper I want to explain that... (from the French corpus)
- Before I will give my arguments for... (from the Swedish corpus)

(Petch-Tyson 1998, 114)

If we try to define specific functions of *I* in argumentative writing, based on Petch-Tyson’s descriptions, we might refer to 1) *I* as narrator (the typical native-like use), 2) *I* as text organizer (referred to as “the writer functioning within the text”), and 3) *I* as arguer (“what the writer thinks”). Hasselgård’s description of self-reference in NICLE also correlate with these definitions; she points out the frequent use of reference to the writer as “participant in a story”, which correlates with the narrating function of *I* (1), and reference to “the writer as writer”, which correlates with the organizing function of *I* (2) (2009b, 131). In addition, she comments on the overuse of patterns of subjective stance, which probably belong to category 3) *I* as arguer.

2.6.2. *I* in academic writing

According to Harwood (2005), the main functions of the academic writer are as follows:

- Organizing the text
- Stating personal opinions and claims
- Recounting experimental procedure and methodology
- Acknowledging funding bodies, institutions, and individuals

(Harwood 2005, 1210)

Harwood's functions apply to all academic fields; however, writer identity is shown to vary according to discipline. Fløttum finds differences between writers in the disciplines linguistics, economics and medicine, when it comes to their functions in the texts. She considers three main functions – 1) the writer as *arguer*, 2) the writer as *reader guide* and 3) the writer as *researcher*. Except for the missing *acknowledging* writer function, these functions are similar to Harwood's functions above. The findings show that, while writers of medical articles usually stay in the researcher role, writers of economic articles act as researchers as well as reader guides, and linguists make use of all three writer roles. Fløttum suggests some reasons for these differences; firstly, medical articles do not call for guiding readers through the text because they usually follow the same strict rules for article structuring. Secondly, linguists may need to discuss and argue more often because linguists have a smaller common knowledge base, and because there is less general agreement as to how certain terms are understood (Fløttum 2006, 67). The findings about writing in the linguistics discipline is particularly important in the present study, as the corpus of academic learner writing (N-VESPA) only consists of linguistics texts.

While Harwood and Fløttum's studies deal with academic writing by professional writers only, Hyland (2002a) considers functions of the writer in novice writing, as contrasted by professional writers. Hyland sorts the rhetorical functions into five groups and finds that novice writers most commonly use 1st person pronouns to state a goal or purpose (36 % of the cases) and to explain a procedure (31 % of the cases). Professional writers also use 1st person pronouns to explain a procedure (38 %), but in addition they like to use them to state results or claims (26 %) and to elaborate an argument (21 %). The ranging of the five rhetorical functions in each group is given in table 7 below, which is based on the original tables 4 and 5 (Hyland 2002a, 1099-1100).

Table 7: Functions of 1st person pronouns in novice and professional writing, based on tables 4 and 5 in Hyland (2002a)

Novice writers	Professional writers
Stating a goal/purpose (36 %)	Explaining a procedure (38 %)
Explaining a procedure (31 %)	Stating results/claims (26 %)
Stating results/claims (16 %)	Elaborating an argument (21 %)
Expressing self-benefits (9 %)	Stating a goal/purpose (15 %)
Elaborating an argument (8 %)	Expressing self-benefits (0 %)

Hyland explains the novices' reluctance to use 1st person pronouns to elaborate an argument with the high risk of this rhetorical function: "Most students sought to disguise their responsibility when elaborating arguments and giving opinions" (Hyland 2002a, 1103). He also finds that that the students in his study rarely use 1st person pronouns with verbs like *think*, *believe* and *assume* (Hyland 2002a, 1103), a notion which clearly separates Hyland's Hong Kong learners from learners in ICLE, as we have just seen. What is more, Hong Kong learners also underuse 1st person pronouns when they state results and claims. According to Hyland, they apply strategies where the author is invisible when they interpret results (Hyland 2002a, 1105).

2.6.3. *We* in argumentative writing

The pronoun *we* is ambiguous, since it may or may not include the reader. The *Longman Student Grammar of Spoken and Written English* explains how *we* can refer to 1) the writer(s), 2) the writer and the reader, and 3) to people in general (Biber, Conrad and Leech 2002, 95). All these different uses of *we* are found in the corpora used in this study. The examples below are taken from the native speaker corpus BAWE and the learner corpus N-VESPA.

1. ***We* referring to the writer:** "Finally **we** discuss the generalisability of the findings." (BAWE)
2. ***We* referring to the writer and the reader:** "As **we** have seen, the word 'blank' in English do correspond to the word 'blank' in Norwegian (...)" (N-VESPA)
3. ***We* referring to people in general:** "One of the most typical cohesive devices **we** use in English is word repetition." (N-VESPA)

The present study will refer to these three uses as 1) exclusive *we*, 2) inclusive *we*, and 3) generic *we*. Herriman (2009), who studies argumentative writing by Swedish learners of English, in addition distinguishes between inclusive specific and inclusive authorial *we*, and exclusive specific and exclusive authorial *we*. In her study of these different usage of *we*, she compares Swedish learners in SWICLE to native speaker students in LOCNESS, and also to professional writers in the British newspaper corpus COMMENT. As shown in table 8, almost all uses of *we* in both SWICLE and LOCNESS are generic, while in COMMENT quite

a large proportion of the uses are exclusive specific, which implies that “(...) the writer represents a group or organisation which exists in the external world outside the discourse” (Herriman 2009, 112). This means that students differ from professional newspaper writers when it comes to the functions of the pronoun *we*, regardless of their L1. Herriman attributes the difference between student writing and newspaper writing to the different social roles, prestige and power of the writers, and to the difference in genre (2009, 113).

Table 8: Different uses of *we*, Herriman (2009)

<i>We</i>	SWICLE	LOCNESS	COMMENT
Inclusive specific	-	4 (1%)	-
Inclusive authorial	14 (1%)	5 (2%)	1 (>1%)
Generic	1195 (95%)	258 (91%)	232 (60%)
Exclusive specific	29 (2%)	7 (3%)	150 (39%)
Exclusive authorial	17 (1%)	8 (3%)	4 (1%)
Total	1255	282	387

(Herriman 2009, 111)

2.6.4. *We* in academic writing

The distinction between exclusive, inclusive and generic *we* also applies to the academic genre. As we saw in the previous section, what Herriman termed the exclusive authorial function of *we* was not used very frequently in the argumentative genre, be it student writing or professional writing. In academic writing, on the other hand, *we* is often used in this way – to refer to the writer, that is. Previous studies on academic writing have shown that exclusive *we* referring to the writer(s) is not only used in articles with more than one author, but also appears in single-authored articles (Kuo 1999). According to Kuo, “the fact that the writer of a single-authored article uses *we*, instead of *I*, as he/she is referring to himself/herself, may suggest an intention to reduce personal attributions” (Kuo 1999, 125). So, although exclusive *we* is one of the most obvious ways to refer to the writer, the writer may still try to hide behind a *we*, which appears less personal than an *I*.

Inclusive *we*, which is also less common in argumentative writing, has been shown to be a popular ingredient in academic writing. The inclusive *we*, which refers to the reader as

well as the writer, can be used by writers “to invite readers into their arguments and presuppose readers’ knowledge” (Kuo 1999, 126). Hyland finds that, in research articles from different disciplines “(...) there is enormous emphasis on binding the writer and reader together through inclusive *we*, which is the most frequent engagement device in academic writing” (Hyland 2005, 182). This conclusion was reached through quantitative corpus studies, and was confirmed through interviews with scientists from different disciplines. One of the scientists that Hyland interviewed for this study, comments on his own usage of inclusive *we*:

I often use ‘we’ to include readers. I suppose it brings out something of the collective endeavour, what we all know and want to accomplish. I’ve never thought of it as a strategy, but I suppose I am trying to lead readers along with me.

(From an interview with a mechanical engineer in Hyland 2005, 183)

The generic *we*, which refers to people in general, can be related to what Kitagawa and Lehrer refer to as “vague *we*” or “impersonal *we*” (1990, 745). The difference between the vague and the impersonal *we* is illustrated in the examples below, where example 1 is an example of vague *we* (or *us*), and example 2 is an example of impersonal *we*, which according to Kitagawa and Lehrer could be substituted by *one*. Both examples refer to people in general, but in the first example this is restricted to Americans, while in the second example there are no restrictions.

- 1) Nationwide only 7.8% of *us* are without a telephone at home.
- 2) *We* are obliged to make the world a better place to live.

Both examples are taken from Kitagawa and Lehrer (1990, 745)

2.6.5. You in both genres

We have already dealt with one of the ways in which the reader becomes visible in academic texts, namely through the inclusive *we*. However, there are other, more explicit ways of addressing the reader, which applies to all genres. For example, the writer can address the reader by posing questions or by using the 2nd person pronoun *you*. The pronoun *you* may

have different functions in a text; it may be used to 1) address the reader(s) in both singular and plural form or 2) to refer to people in general (generic *you*), which is similar to the generic function of *we* above (Biber, Conrad and Leech 2002, 95). Since it appears as if few corpus-based studies deal specifically with the different functions of *you*, I have chosen to illustrate the difference between the two main functions, by providing examples from the native speaker corpus LOCNESS:

- 1) **Reader-addressing you:** “This is all very interesting **you** may say (...)” (LOCNESS)
- 2) **Generic you:** “I’ve heard that if **you** ask a child who they admire, many answer with the name of a teacher.” (LOCNESS)

In previous research, it appears as if the role of the reader in argumentative and academic writing has not been dealt with to a similar extent as the role of the writer. This is probably because the writer usually is present to a greater extent than the reader, in these genres. Still, the role of the reader is not less important; in fact, the degree of reader visibility in a text, can tell us more about the total level of involvement than the degree of writer visibility can. According to Smith, we can be confident that the writer is always more visible than the reader in a text because, as he puts it, “[r]eferences to the reader presuppose a writer making those references, though self-referential statements by the writer do not necessarily entail an explicitly acknowledged reader” (Smith 1983, 6). Smith supposes that “the greater the degree of interpersonality in a text (...), the more the presence of the participants in the discourse will be explicitly acknowledged” (Smith 1983, 6). Thus, when the reader is explicitly acknowledged, in addition to the writer, the degree of interpersonality increases.

2.7. What the present study aims to contribute

The previous studies which have been referred to in this chapter all contribute to our knowledge of the roles of the writer and the reader in argumentative and academic writing. Some of the studies focus on learner writing, while other studies focus on professional writing. Furthermore, some of the studies focus on W/R visibility, while other studies focus on the functions of the writer and/or the reader. The present study aims to contribute to our

knowledge about the writer and the reader in argumentative and academic advanced learner writing, by building on the knowledge from these previous studies.

The present study will consider features of W/R visibility which have already been investigated in previous studies (of other learner groups). In this way the findings from the present study, which only considers Norwegian learners, may be related to the findings in other studies. This makes it possible to determine whether the potential overuse of W/R visibility features in Norwegian learner writing is more or less significant than the overuse which has been found in the writing of learners with other L1 backgrounds.

It has already been found, by Paquot, Hasselgård and Ebeling (forthcoming), that Norwegian learners seem to overuse W/R visibility features to a greater extent than French learners. This made it possible to begin the present study with the hypothesis that Norwegian learners probably have a high level of W/R visibility, and that it might be interesting to compare Norwegian learners with other learner groups. What distinguishes the present study from that of Paquot, Hasselgård and Ebeling is that several additional features of W/R visibility are investigated, and that additional learner groups are considered. Also, the present study includes a study of Norwegian L1 sources. The combination of comparing different learner groups, and studying the learners' L1, may help identify to what extent L1 transfer might have influenced learners' use of W/R visibility features. Other possible reasons for the overuse, suggested in previous studies (in section 2.5), are interesting, but will not be dealt with in detail in the present study. However, these issues will be considered briefly in chapter 7.

When it comes to writer/reader roles, or the functions of 1st and 2nd person pronouns, the present study will build on the previous studies mentioned in this chapter. Many of the functions found in the qualitative investigations of the present study, are similar to the functions referred to in section 2.6. The inclusive, exclusive and generic functions of *we* were found, both in argumentative and academic writing, as were the reader-addressing and generic functions of *you*. The pronoun *I* was used as arguer, organizer and narrator in both genres, as anticipated after considering Petch-Tyson's (1998) and Hasselgård's (2009b) findings. In addition, *I* as researcher was found in the academic genre, in accordance with the findings in Fløttum (2006) and Harwood (2005). Although several studies have mentioned some functions of pronouns, or functions of the writer and the reader, the present study contributes

new information by studying the use, and potential overuse, of these functions in Norwegian learner writing. This information may be useful to EFL teaching in Norway, since it reveals some of the learners' skills and some of their problem areas in a more precise way than the quantitative information from the frequency comparisons.

3. Method and material

In this chapter, my methodological choices, regarding features, corpora and comparisons, are explained and discussed. First, I will consider the preliminary study, which is based on the list of features presented in table 1 (section 1.1.). Then, the in-depth study of personal pronouns will be presented, explaining and discussing the procedures for both the quantitative and the qualitative parts of the study. Finally, the corpora are presented in more detail, and the pros and cons of each corpus are discussed. A schematic comparison of the different corpora is given in the last section (3.5) of this chapter.

All the corpus comparisons in the present study focus on frequency differences between corpora, usually learner and native speaker corpora. In order to make sure that the differences are statistically significant, I have used an online log-likelihood calculator which is directly aimed at frequency-based corpus studies¹³. I have based the log-likelihood calculation on a 5 % level ($p < 0.05$), which means that the estimated significance is at least 95 % certain. In our case, this means that log-likelihood values of 3.84, or more, indicate significant differences between two corpora. Significant differences may again be described as significant overuse or underuse when the corpus comparisons consider learners versus native speakers. Thus, the statistical significance test, provided by the log-likelihood calculator, is an invaluable tool in the present study, making it possible to be more confident about any potential overuse.

3.1. The preliminary study of writer/reader visibility

In this preliminary study, the level of W/R visibility in Norwegian learner writing is described by contrasting the use of W/R visibility features in Norwegian learner writing and native speaker student writing, and also by drawing on the findings from previous studies, which consider other learner groups. In addition to adding to our knowledge about the general level of W/R visibility in Norwegian learner writing, the detailed information about the frequencies, and potential learner overuse, of each feature, form the basis for the choice of features to study further in the in-depth study.

¹³ Available at <http://ucrel.lancs.ac.uk/llwizard.html>. More information about these calculations can be found in the article “Comparing Corpora using Frequency Profiling” (Rayson og Garside 2000) at http://www.comp.lancs.ac.uk/~paul/publications/rg_acl2000.pdf.

3.1.1. Features studied

When investigating the overall frequency of features of W/R visibility in learner corpora and native speaker corpora, I used a list of features found in a collection of previous studies. The features were sorted into groups, some of which were considered closed and some of which were considered open. The group which I called “first person singular pronouns” is an example of a closed group. For all groups of pronouns I decided to include possessive pronouns/determiners as well as personal pronouns; thus the group called “first person singular pronouns” consists of the items *I, me, my* and *mine*¹⁴. This is considered a closed group because the definition of the group does not open for any additional items. Because it is a closed group, I can treat the items together, and choose to refer to the total frequency of 1st person sg pronouns, instead of the frequency of each individual item.

The group called “reference to situation of writing/reading” is an example of an open group. In the present study it consists of the items *here, now* and *this essay*, because these are the items used in previous studies. However, we can imagine several other ways in which to refer to the situation writing and/or reading, such as *this paper* or *the present study*, which means that the list of items is not exhaustive. Items in such groups are therefore treated only as examples of the types of features that they represent. Consequently, I do not consider the total frequency of items belonging to open groups, but only the frequency of each item.

The groups of features are listed below, along with all the items which I have explored in the different corpora. I have also marked all the groups as either closed or open.

¹⁴ Possessive determiners were included in the pronoun groups for simplicity’s sake, as they often have meanings similar to those of possessive pronouns and personal pronouns. Reflexive pronouns were excluded from the group, in accordance with previous studies such as Petch-Tyson (1998). In this way comparisons can easily be made across studies.

- **First person singular pronouns** (closed group): *I, me, my, mine*.
- **Patterns of subjective stance** (open group): *I believe, I think, I don't (do not) think, I guess, I suppose, I would say, I would like to say, I'm (I am) sorry to say, I would argue, I must emphasize, I know, I remember, I'm (I am) not saying*.
- **First person plural pronouns** (closed group): *We, us, our, ours*.
- **Second person pronouns** (closed group): *You, your, yours*.
- **Patterns introducing other voices** (open group): *You could say that, you may say that, some may say that, some might say that, one can say that, many think, let's (let us) say that, you think, you could have guessed, you may ask, I hear you saying*.
- **Disjuncts** (open group): *of course, naturally, perhaps, maybe, unfortunately, obviously, frankly*.
- **Emphatic particles** (open group): *just, really*.
- **Reference to situation of writing/reading** (open group): *here, now, this essay*.
- **Sentence types** (open group): questions (question marks), exclamations (exclamation marks)

Notes on the choice of features

There are some limitations to the list of features in the present study. First of all, the list is not exhaustive, so it is not possible to say anything certain about the overall W/R visibility in the corpora. However, I do not believe that it would have been possible to include all conceivable ways of referring to the writer and the reader, because our imagination lets us find all sorts of creative ways to express ourselves in our writing, which makes systematizing very difficult.

Secondly, the groups from Hasselgård's study (2009b) – patterns of subjective stance and patterns introducing other voices – are slightly problematic to use in the present study. This is because the patterns were originally found through a close reading of Norwegian learners' argumentative texts, which may influence the results from the comparisons with other corpora. It may be that native speakers express subjective stance through similar, but still not exactly the same, patterns, and this is not accounted for in the quantitative corpus investigations of the present study. It should also be mentioned that most of the patterns have very low frequencies in all the corpora, including the corpus from which they were taken. These patterns may be exclusive to one particular learner, and not representative of the

Norwegian learner group. Nevertheless, I found it interesting and important to study all these patterns, because the results from the different corpora showed that some of the patterns were indeed very frequent in both learner and native speaker writing. These patterns were also, in some cases, shown to be overused by the learners.

Finally, a note about the open groups; it would probably have improved the total value of this study to have included even more items in the corpus investigations, in order to be able to close some of the open groups. When investigating the frequency of different disjuncts in learner and native speaker corpora, for example, we cannot rule out the possibility that the results may have been different if more disjuncts were studied. In the present study, some disjuncts were underused and others overused by learners, and we cannot really say anything certain about the overall use of disjuncts, since many are left out. However, as already mentioned, it seems an impossible task to include all sorts of features and to be able to say something general about W/R visibility. Also, I wanted to base the choice of items in my own study on those in previous studies, both to be sure that my investigations were valid, and to be able to compare my own corpus results to the results from other studies.

Notes on ambiguous items

When counting the frequencies of the many items belonging to the different categories of writer/reader visibility, the frequencies will not always tell us what we want to know. In the present study, the only items we want to count are those which refer to the writer, the reader or the situation of writing/reading. Therefore, it is a problem that computer-based investigations do not automatically distinguish between different uses of a word. The result of automatic investigations is that not all the items which have been counted will be relevant. In some cases, it was considered important to check for relevance, since the amount of irrelevant hits might constitute a major part of the automatically calculated frequency.

Here and *now* are examples of words with different functions; when they refer to the text or something that is happening in the text they have a writer/reader involving function, but when *here* refers to a country and *now* refers to these days, for example, their function is not involving in that sense. Examples 1 and 2 show the relevant uses of *here* and *now*.

1. All of the examples I mentioned **here** are, in my opinion, positive changes. (NICLE)¹⁵
2. I have **now** looked upon advantages and disadvantages of marriage compared to (...) (NICLE)

When investigating the frequencies of *here* and *now* in the four corpora, only the relevant cases were to be counted. Therefore, the automatic frequency calculations that the concordancers offer could not be applied. A manual sorting of relevant and irrelevant cases led to the frequencies which are presented in chapter 4. Uses of *here* which were considered irrelevant for the present study are those which refer to places in the world (ex. 3). Irrelevant uses of *now* include references to the present time (ex. 4) and also *now* as a discourse marker (ex. 5).

3. **Here** in Norway most people eat bread with cheese and drink milk or orange juice in the morning. (NICLE)
4. They often worked much more than we do **now** here in social democratic Norway. (NICLE)
5. **Now**, what the hell is the problem? (NICLE)

The phrase *this essay* is often used to refer to the situation of writing/reading by way of informing the reader what the essay is going to deal with or how the essay is structured. In NICLE – the corpus of Norwegian learners’ argumentative essays – all the uses of *this essay* has this function. This is also the case in LOCNESS – the comparable native speaker corpus. So, when investigating the use of *this essay* in argumentative writing, no extra measures were needed to establish the correct frequencies. The example below is a typical use of *this essay* in argumentative student essays.

6. In **this essay** I will deal with the period starting from about 1960 lasting till today. (NICLE)

¹⁵ Emphasis is added in most of the examples in the present study. Bold font is always an indicator of added emphasis.

In the academic corpora, however, there are two main uses of *this essay* – one which is similar to that of the above example, and one which is typical to the linguistics discipline. The linguistics usage of phrases like *this essay* is related to the fact that linguists often present studies of texts. Therefore, we find cases in N-VESPA and BAWE-ling – the academic learner and native speaker corpora in this study – where *this essay* refers to an essay which is being analyzed by the writer. In these cases, *this essay* cannot be said to make the writer and/or the reader visible. Therefore, manual sorting is needed in these two corpora. The example below exemplifies the cases which were not included in the final raw frequencies for *this essay*.

7. The high number of misspelled words in **this essay** seems to be due to slips rather than actual errors. (N-VESPA)

A final remark on ambiguous items regards words with more than one meaning. In the present study, such words were identified by studying the first concordance lines in each corpus search. *Mine* and *us* were identified as such words, as *mine* was found to be used as a pronoun and a noun, and *us* can refer to the pronoun *us* as well as the *US* (United States). When investigating these words, the whole list of concordance lines was studied manually, in order to identify the relevant hits. In some cases, where the list of concordance lines was too long, samples of 100 concordance lines were studied instead, and an estimated amount of relevant items was calculated.

Certainly, other words have irrelevant hits as well. The qualitative study of pronouns, which considers the functions of *I*, *we* and *you*, revealed quite a few irrelevant hits. For example, the item *I* may have been used in the abbreviation *i.e.* However, it would have been too time-consuming to study the concordance lines thoroughly in all the corpus investigations in the present study. Therefore, we must take into account that a certain amount of irrelevant hits have been counted. However, this is the case in both the learner and the native speaker corpora, so it may not have affected the investigation to a great extent.

3.1.2. The corpus investigations

In order to explore the use of features of W/R visibility in Norwegian learner writing, and to detect potential overuse of these features, I have compared learner writing in NICLE and N-VESPA to native speaker student writing in LOCNESS and BAWE-ling. Due to comparability issues in terms of study level and genre, NICLE has only been compared to LOCNESS, and N-VESPA only to BAWE-ling. Explanatory information about these choices is given in chapter 3.3., which discusses the four main corpora more in detail. In addition to comparing Norwegian learner writing and native speaker writing, I have compared the findings in the present study to relevant findings from previous studies. By doing so, I hope to add to our understanding of Norwegian learner writing, as compared to the writing of learners with other L1s.

3.2. The in-depth study of pronouns

The in-depth study of pronouns is a mixture of quantitative and qualitative investigations which are intended to give further insight into Norwegian learners' use of W/R visibility features, exemplified by 1st and 2nd person pronouns. In the quantitative part, the two main objectives are identifying any signs of L1 transfer related to the overuse of pronouns, and investigating the effect of academic experience, so as to explain the differences between the overuse in NICLE and N-VESPA. This will be done by performing several corpus investigations. In addition to these quantitative investigations, a qualitative study considers the functions of 1st and 2nd person pronouns in Norwegian learner writing and native speaker student writing. The study of the functions of 1st and 2nd person pronouns will also focus on differences between learners and native speakers, and the aim is to identify overuse and/or underuse of the different functions in learner writing.

3.2.1. Choosing personal pronouns

The results from the preliminary corpus investigations identified pronouns as significantly overused features of W/R visibility. More precisely, this refers to 1st and 2nd person pronouns, both singular and plural, and both personal pronouns and possessive pronouns/determiners. Other features also presented themselves as possible objects for an in-depth study, but

because of the high raw frequencies of pronouns, these were considered the most reliable features to investigate further, since the results from such an investigation might lead to more confident conclusions. It is also an advantage for later investigations that pronouns are relatively similar in Norwegian and English, particularly when it comes to investigating the possibility that transfer from the learners' L1 might have caused the learners to overuse features of W/R visibility.

3.2.2. Studying pronouns quantitatively

The main aim of the quantitative in-depth study of pronouns was to explore the reasons for the overuse which was detected in the preliminary study. In order to investigate the potential transfer effect from the learners' L1, sources of Norwegian L1 writing were used to search for possible differences between the English language and the Norwegian language. Also, to investigate the probability of L1 transfer, corpora of learner writing by learners with other L1s than Norwegian were considered. Next, to investigate the effect that academic experience may have on learner overuse of pronouns, the N-VESPA learners were divided into bachelor students and master students, and the differences between the two groups were studied. This was done in BAWE-ling, as well, to see if the developmental tendencies were the same for the L1 students as for the L2 students. The following sections deal with the specific procedures related to each corpus investigation.

The effect of L1 transfer

In an effort to add to our knowledge about Norwegian learners' overuse of pronouns, I investigated the potential effect of L1 transfer related to this. Two corpora of Norwegian L1 writing were studied, namely NOESS (NORwegian ESSays) and NOBA (NORwegian Bachelor's Assignments¹⁶). The frequency of 1st and 2nd person pronouns in these corpora were compared to the frequencies in the corpora of L1 English and the corpora of L2 (learner) English. If 1st and 2nd person pronouns were found to be more frequent in Norwegian L1 writing than in English L1 writing, there would be reason to believe that transfer from the L1 could partly explain Norwegian learners' overuse of these pronouns.

¹⁶ More about these corpora in section 3.4.

Comparing Norwegian learners to other learners of English would also help in identifying the role of transfer. If Norwegian learners overuse pronouns to a greater extent than learners with other L1s, the probability of L1 transfer is higher. Here, Norwegian learners in NICLE are compared to ICLE learners in general, as well as to some specific learner groups in ICLE. N-VESPA learners, on the other hand, are not compared to other learner groups through any corpus comparisons, because, to my knowledge, no comparable corpora of academic learner writing are publicly available at this point. However, Norwegian and French learners in VESPA have already been compared in Paquot, Hasselgård and Ebeling (forthcoming), and this will be referred to where relevant.

The effect of academic experience

The effect that academic experience may have on learners' overuse of pronouns was investigated by comparing students at different university levels in two corpora. Since NICLE and LOCNESS students do not represent the whole range of university levels, these corpora were excluded from this investigation. In N-VESPA, the Norwegian learners were divided into bachelor students and master students, and in BAWE, the native speaker students were divided into two groups based on the levels defined in the BAWE corpus – levels 1-2 and levels 3-4. The hypothesis is that the more experienced students may have a lower degree of W/R visibility than the less experienced students. If this is the case, this might partly explain the differences between NICLE and N-VESPA.

3.2.3. Studying pronouns qualitatively

For the qualitative study, the pronouns *I*, *we* and *you* have been selected to represent 1st person sg, 1st person pl and 2nd person pronouns, as they are the most frequently occurring forms in each group. Samples ranging from 100-200 concordance lines from each corpus were studied in order to find the functions of these pronouns in learner and native speaker writing. Adding to the quantitative studies which showed that learners overuse all these pronouns, a qualitative study can identify which functions of the pronouns are overused. Since different functions of a pronoun may imply different degrees of W/R visibility, identifying the

overused functions may tell us more about the total degree of W/R visibility in learner writing when compared to native speaker writing.

In the present study, the functions of *I*, *we* and *you* were identified by studying samples of each pronoun, looking at the contexts in which they appear. The studies referred to in section 2.6, which identified different functions of pronouns were used to label some of the functions, such as exclusive and inclusive *we* (Herriman 2009), and *I* as organizer, researcher and arguer (Fløttum 2006, Harwood 2005). After identifying functions and calculating the estimated frequency of each function of the three pronouns per 10,000 words, the results from the learners were compared with the results from the native speakers. Since the functions of the writer and the reader differ according to genre, argumentative texts and academic texts were studied separately, so NICLE was compared to LOCNESS, and N-VESPA to BAWEling. Note that, in this part of the study, any identified overuse or underuse of functions in learner writing will be based on estimated frequencies, which means that the results are not as reliable as when using actual raw frequencies. The statistical significance of the differences between the corpora, regarding the different functions, is also based on the estimated frequencies, which means that the findings are not necessarily as certain as the statistical significance value signals.

3.3. The main corpora

3.3.1. NICLE

Contents

NICLE is the Norwegian part of the International Corpus of Learner English¹⁷ (ICLE), and it is our source of argumentative essays by Norwegian learner writers. The essays are not discipline-specific; they rather discuss general topics which are more or less controversial, such as the ones in the examples below.

¹⁷ For more information about ICLE, visit UCL's website at: <http://www.uclouvain.be/en-cecl-icle.html>

1. Some people say that in our modern world, dominated by science and technology and industrialisation, there is no longer a place for dreaming and imagination. What is your opinion?
2. Feminists have done more harm to the cause of women than good.
3. In his novel "Animal Farm" George Orwell wrote "All men are equal but some are more equal than others". How true is this today?

The learners who have contributed to the ICLE corpus share some features, such as age, learning context, level, medium, genre and technicality, and vary when it comes to other features, such as sex, mother tongue, region, other known foreign languages, topic and task setting (Granger 1998, 9). Sometimes it is useful to sort the corpus according to the different variables, instead of studying the corpus as a whole. A recent ICLE version (ICLE 2 on CD-ROM) allows users to make their own corpus selection based on numerous variables, enabling researchers to study linguistic differences between male and female writing or between learners with different L1s, for example.

In the present study, only two criteria were set to define the Norwegian learner group – the home country (Norway) and the L1 (Norwegian). The selection of essays which conform to these two criteria is what the present study refers to as NICLE. Consisting of 316 essays, totaling 210,367 words, NICLE is a relatively small corpus, but this is often the case for L1-specific learner corpora. In comparison, the Dutch, Finnish, French and Swedish sub-corpora of ICLE used in Petch-Tyson (1998), only consist of about 50,000 words each.

When it comes to setting issues, it has been pointed out by Ädel (2008) that timing and access to secondary sources may impact the use of involvement (W/R visibility) features. In NICLE, only 2 % of the essays were timed, so timing is not a major issue. The issue of access to secondary sources, however, could be a problem because 37 % of the essays were written without access to secondary sources. According to Ädel this means that the learners' use of disjuncts, questions and exclamations can be expected to be higher than what would have been the case if the learners were able to use such sources (Ädel 2008, 46). Fortunately, when it comes to pronouns, which is the main focus of the present study, only the timing factor is an issue, so no significant influence from setting variables will be expected in these cases (Ädel 2008, 46).

Advantages

One of the advantages with the NICLE corpus is that the learner group is quite homogenous in many ways. They all attend university, and most of them (93 %) have studied English at university level for less than a year, which means that we might expect similar levels of English proficiency. The essays are also quite similar; almost all of them (98 %) are argumentative¹⁸, and all the texts are relatively short, ranging from 200 to 1500 words, most of them (86%) belonging to the range of 500-1000 words. All these similarities make the NICLE corpus representative for a certain learner group and allow us to get a better understanding of this particular learner group without being distracted by numerous unshared features.

Another advantage with the NICLE corpus is related to the importance of adding to an existing field of research. Any study which is based on ICLE or a sub-corpus of ICLE will add to the total amount of knowledge shared by the research community. Not only is this important in order to learn more about learners of English in general; it is also helpful for those who investigate one particular learner group to be able to compare results with similar studies of other learner groups. This cannot be done with the same accuracy if the studies of other learner groups are not based on similar corpora. Quoting Meyer, who also values the use of already existing corpora where it is possible, “[i]t is therefore most desirable to work with a corpus already available not just to decrease the work time but to add to the growing body of work that has been based on a given corpus” (Meyer 2002, 103).

Disadvantages

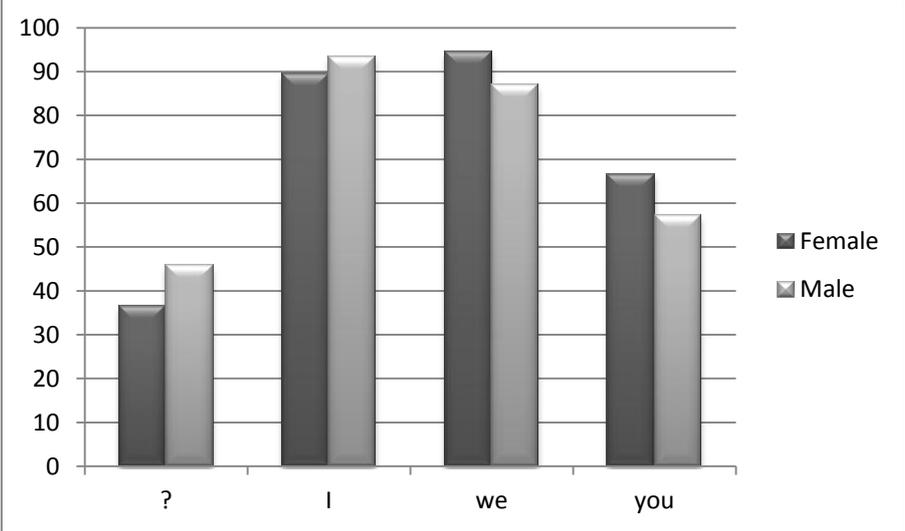
What appears to be a disadvantage with NICLE is the unbalanced gender distribution. Almost three quarters of the writers are female, and the potential problem, relevant to this study, is that there might be differences in the ways and degrees in which male and female writers are involved in their writing. Some sociolinguistic studies show differences between men and women’s language, some of which can be related to the degree of W/R visibility. For example, women are found to use tag questions (e.g. – *isn't it?*) more frequently than men because they “(...) put more emphasis than men on the polite or affective function of tags” (Holmes 1999, 320). Since questions are typical markers of W/R visibility, such notions about

¹⁸ The rest are essays relating to English literature.

the female use of questions are relevant. However, these findings do not necessarily apply to the argumentative genre, and, therefore, an investigation of the male and female contributions to NICLE was required.

Since ICLE 2 allows the user to distinguish between male and female writers, it was an easy task to find out about the distribution of certain items in male and female writing. The items that I chose to investigate in this small corpus comparison are questions, since they were suggested as a problem, along with some of the most frequently occurring features of W/R visibility, namely the pronouns *I*, *we* and *you*. Figure 18 shows the frequencies of these features per 10,000 words in male and female NICLE writing.

Figure 18: W/R visibility in texts by male and female NICLE writers



As shown in figure 18, there are some minor differences between male and female writers, but these differences do not point in the direction of a generally higher level of W/R visibility in female writing. Statistical tests showed that the differences in the use of the pronouns *I* and *we* were not significant, whereas the differences in the use of questions and the pronoun *you* were significant. Thus, it seems like questions are actually overused by male writers, instead of female writers, and, on the whole, it does not seem like female writers use W/R visibility features more often than male writers. Judging from these findings, it does not seem likely

that the overrepresentation of female writers in NICLE has had any significant effect on the present study.

3.3.2. LOCNESS

Contents

The Louvain Corpus of Native English Essays¹⁹ (LOCNESS) is a corpus of student writing by American and British students, consisting of 290 essays and totaling 324,400 words. The corpus is said to be directly comparable to ICLE (Granger 1998, 13), and is often used as a control corpus when studying learner language, as we have seen in chapter 2. In the present study, an online version of LOCNESS has been used, made available for concordance searches by the Oslo Interactive English (OIE) interface. This is a simple search interface, which provides frequency counts and concordance lists for any item, but no extra information about the writer or the task. It does, however, give the user access to the source text so that he/she can study the context of the item in question. The immediate context will also appear in the concordance list, as exemplified below. This is similar to the way that concordances are presented in ICLE 2.

Figure 19: Concordance list for *me* in LOCNESS

o use another argument used. How it can be sport is beyond me .

...it is not as open as it was in the sixties. It is hard for me , a white middle class male, to say what prejudice is like ...

... his citizens (and expressly to the audience). It seems to me a sort of perverse cruel-to-be-kind type of rule.

... an old car. I kept remembering what my father always told me about "crime not paying" and I knew for sure that they wou...

... my shoulder wondering if this was ever going to happen to me again or to someone else. My talk with them helped. They d...

Another case and point that would make me agree with an abortion is if the mother was unable to carr...

Advantages

Some of the reasons why LOCNESS is considered a valid control corpus for ICLE are that the writers of the two corpora are of the approximate same age and at similar study levels, and that most of the essays are of the same type and length. Many of the tasks are of the same

¹⁹ For more information about LOCNESS, visit UCL's website at: <http://www.uclouvain.be/en-cecl-locness.html>

nature as those in NICLE, and, although they are not exactly the same, they are at least mostly argumentative, and they discuss more or less controversial topics.

Disadvantages

As we saw from the examples in section 2.1., the LOCNESS writers have been asked to write about topics such as euthanasia and capital punishment, which might lead to argumentative essays similar to those in NICLE. However, it might have affected the NICLE writers that they were given questions, not merely topics, in which they were asked specifically to offer their opinions. Whether or not this was the case for the LOCNESS writers, I have not been able to find out. Therefore, there is some uncertainty related to the possibility that the learners might have been more encouraged to refer to themselves than the native speakers, which might have affected this investigation directly. Uncertainty is a major problem when studying LOCNESS, since too little information about the students has been made available for users of the corpus.

Another potential problem is that LOCNESS includes some literary essays, in addition to the argumentative essays. It might have been preferable to exclude these essays from the investigation, since this would have made NICLE and LOCNESS more suitable for comparison, but the online interface used to measure frequencies in the LOCNESS does not allow users to make selections of texts. This means that I had to use the whole original LOCNESS corpus, which in addition to literary essays also includes some A-level essays from the British students. A-level students are less comparable to the NICLE students because they are at a lower educational level. The advantage of including both literary and A-level essays in the study, however, is that it makes the corpus larger. LOCNESS is already quite a small control corpus, so it is preferable not to reduce its size.

3.3.3. N-VESPA

Contents

N-VESPA is a corpus of Norwegian learners' discipline-specific academic writing. The corpus is the Norwegian part of the Varieties of English for Specific Purposes dAtabase²⁰ (VESPA). VESPA is a corpus in progress, and I have gratefully been permitted to use the current contributions from the University of Oslo. From these texts I have selected those which were written by Norwegian students who have Norwegian as their L1. This is what I call N-VESPA, which is not an official corpus, but a selection specific to the present study. This selection consists of 194 essays totaling 231,675 words, and the essays all belong to the academic genre and the linguistics discipline.

The texts in N-VESPA were studied through the corpus analysis program Wordsmith Tools. Similar to all the other concordancers used in this study, Wordsmith Tools allows for concordance searches, automatically calculates the raw frequencies of any item, and presents a list of hits with some immediate context. It is also possible to enter the texts in which the items appear, and to access different sorts of statistical information about the item and the texts. An optional function in Wordsmith Tools, which was applied in the N-VESPA investigation, is to exclude parts of the texts from the automatic analysis. Such exclusion requires pre-tagging of these elements; in the N-VESPA documents several tags have been inserted, making it possible to leave out elements which are not representative of the learners' own language, such as quotes and linguistic examples, as well as parts of the texts which are not regular, grammatical sentences, such as lists.

Leaving out tagged elements has proved to be particularly important with regard to texts within the linguistics discipline, because these texts tend to have many linguistic examples which can completely change the frequencies of certain words. To illustrate this point, a search for *me* in N-VESPA, before filtering out quotes and linguistic examples, resulted in 148 hits, while the same search after filters resulted in 71 hits, because in most of the cases *me* was part of a text sample which was the object of analysis in the N-VESPA text. This means that in order to identify the actual level of W/R visibility in the N-VESPA texts, it is crucial that we use the tag filters. N-VESPA and NICLE are different in this respect,

²⁰For more information about VESPA, visit UCL's website at: <http://www.uclouvain.be/en-cecl-vespa.html>

because argumentative texts like those in NICLE are not analyzing in nature, so they usually do not have as many quotes and examples.

Advantages

Since the present study focuses on advanced learner writing by Norwegian students, it was most useful to have two independent sources of Norwegian learner writing at this level. For a long time, NICLE has been the only available source of Norwegian learner English, so I feel very fortunate to be able to include N-VESPA as an additional source. Studying two similar corpora makes it possible to see whether a high degree of W/R visibility is typical for just one corpus, or whether it may be typical for advanced Norwegian learner writing in general. When making general conclusions, only tendencies can be estimated on the basis of two corpora, but such conclusions are nevertheless more confident than conclusions based on only one corpus.

In addition to just being another corpus of advanced Norwegian learner writing, N-VESPA offers possibilities for investigating more detailed issues about W/R visibility in learner writing. This is because the N-VESPA contributors represent all university levels from first-years to master students. By studying differences in the writing of bachelor students and master students in N-VESPA, for example, we gain insight into the development of the use of W/R references at different stages.

N-VESPA also has the advantage of complying with the requirements in Ädel (2008) about time and setting, which means that none of the N-VESPA texts are timed, and that all the writers have been free to use secondary sources when writing their texts. In addition to this, it is important to note that the N-VESPA texts were not originally written with the purpose of contributing to a corpus. Instead, they were submitted as obligatory assignments for different English courses at the University of Oslo, and later added to the corpus after the students had given their consent to this. This has two important implications; firstly, the students probably put a lot of work into the texts since their grades partly depended on the results, which means that the texts show what the learners are able to do when they put some effort into the task. Secondly, texts are more authentic when they are not written solely for the purpose of being research material, and corpus linguistics is all about studying authentic or

“real” language: “In short, linguists of various persuasions use corpora in their research, and are united in their belief that one’s linguistic analysis will benefit from the analysis of ‘real’ language” (Meyer 2002, xiv).

Disadvantages

Gender distribution was at first considered a potential problem in this corpus as well, since there is a significantly higher number of female than male contributors in N-VESPA. However, the uneven gender balance probably reflects the actual gender distribution in the linguistics discipline, so it may not have been more appropriate to have a fifty-fifty gender distribution in the corpus. Also, as we have already seen in the NICLE investigation of male and female writing (in section 3.3.1), female writers did not seem to have a higher level of W/R visibility than male writers. For these reasons, the gender distribution in N-VESPA has not been considered an actual disadvantage.

The subjects of the essays, however, constitute a more significant disadvantage. To this date, the N-VESPA corpus only covers linguistics subjects. To compensate for this, only the linguistics part of the native speaker student corpus was included in the corpus comparison. However, this means that the control corpus is significantly smaller than what it could have been. It also means that we cannot make any general conclusions about academic learner language based on this study.

3.3.4. BAWE(-ling)

Contents

The British Academic Written English²¹ (BAWE) corpus has been chosen as the control corpus for N-VESPA. Since the N-VESPA writers are more advanced than the NICLE writers, it was important to find a control corpus for N-VESPA which was more advanced than LOCNESS. The BAWE contributors range from first year students to master students and are thus similar to the N-VESPA contributors.

²¹ For more information about BAWE, visit Coventry University’s website at: <http://wwwm.coventry.ac.uk/researchnet/BAWE/Pages/BAWE.aspx>

The BAWE corpus is fairly large, consisting of 1953 essays and totaling 4,534,873 words²². In comparison, the linguistics section of BAWE, which was used for comparison to N-VESPA, is quite small, consisting of only 76 essays and 174,840 words. Nevertheless, since studies have shown that there are significant differences in the way that W/R references are used in different disciplines (Hyland 2002b, Fløttum 2006), it was considered important to include only linguistics texts in the main corpus comparisons with N-VESPA. The online concordancer Sketch Engine, which allows users to define their own corpus selection, was used to make a sub-corpus consisting of linguistics texts written by students with English as their L1. This corpus is referred to as BAWE-ling. BAWE and BAWE-ling are used in different corpus comparisons, in order to ensure the best possible comparability level. However, BAWE-ling is definitely the main corpus of academic native speaker writing in the present study.

Advantages

When using the whole BAWE corpus, the main advantage is the size of the corpus, and the wide range of disciplines. Using the smaller sub-corpus BAWE-ling, however, is also advantageous in many ways. BAWE-ling is highly comparable to N-VESPA when it comes to genre and discipline, which are two important variables. Furthermore, BAWE-ling contributors, like BAWE contributors in general, represent four different levels of university experience, and they have written academic texts which are described as “proficient” and “good-standard” (Coventry University 2011). It benefits the present study to use good-standard essays when we are comparing learners of a language to native speaker students, because the native speaker source should offer a standard that the learners would want to strive towards.

Disadvantages

A potential problem when comparing the tagged version of N-VESPA to the online version of BAWE and BAWE-ling, is that the frequencies might not be fully comparable. While all quotes are removed from N-VESPA, this is not the case in BAWE where block quotes are

²² When only L1 English writers are included.

removed from the texts, but not quotes which appear inside paragraphs. This means that when the concordancer automatically calculates the frequency of an item, the items which are parts of a quote will also be counted. This may pose a problem to the investigation if some items appear frequently inside quotes. However, it would be too time-consuming in this context to read through all the hits in the BAWE corpus, because of the high amount of items that are investigated and the high frequency of these items in the corpus.

3.4. The sources of Norwegian L1 writing

3.4.1. NOESS

Contents

NOESS, which is short for NORwegian ESSays, is a small corpus of argumentative essays in Norwegian. It was collected from the website skoleforum.no by Hasselgård to be used as a source of Norwegian L1 writing in a study of Norwegian learner English (Hasselgård 2009a). The present study performs similar comparisons as those in Hasselgård – the use of certain features are studied in NOESS, LOCNESS and NICLE to check for transfer from the L1 in NICLE.

NOESS was investigated by using the search function in Microsoft Word, which was unproblematic because of the small size of the corpus – NOESS only consists of 13 essays which total 14,761 words. The essays are written by students in upper secondary school, and deal with topics like the media and the environment. Hasselgård comments that although the texts are all argumentative, they “(...) often contain elements of exposition” (Hasselgård 2009a, 95).

Advantages

Clearly, NOESS is not a perfect corpus, but it has the advantage of being authentic. The essays in NOESS were written for educational purposes and published online by the writers themselves to inspire other students. In this way, NOESS is similar to N-VESPA – it is a corpus through which we can study real language.

Disadvantages

There are some disadvantages related to using NOESS as a corpus in a comparison with other larger corpora, such as NICLE and LOCNESS. Firstly, the small size of the NOESS corpus is not ideal; it makes our conclusions less confident. Secondly, the NOESS writers are younger than the NICLE writers and most of the LOCNESS writers to whom they are compared. Thirdly, as was also the case with the LOCNESS writers, we have no extra information about the NOESS writers, apart from their study level which is upper secondary school. Still, I found it useful to compare NOESS to LOCNESS in a study of W/R visibility across languages, in order to see if there were any tendencies towards a higher degree of W/R visibility in Norwegian L1 essays than in English L1 essays.

3.4.2. NOBA

Contents

The Norwegian Bachelor Assignments (NOBA) collection is compiled for the present study to be a comparable corpus to N-VESPA and BAWE, in the same way that NOESS is comparable to NICLE and LOCNESS. It consists of 15 texts written in Norwegian by Norwegian bachelor students. These texts were found by using the BIBSYS search interface, which is used at the University of Oslo and accessed online. The total size of the corpus is 168,850 words, which means that the texts average about 11,000 words. The texts are quite lengthy for a corpus, as the standard length of a text sample is often 2,000 words (Meyer 2002, 33). However, since it is possible that pronouns may be unevenly distributed throughout the texts, it is important to study as set of complete texts (Meyer 2002, 30).

Since studies have suggested that the use of W/R visibility features in academic writing has changed over the years (McCrostie 2008), I chose to include quite recent publications only – from 2003 to 2010 – in order to ensure the best possible representation of academic language as it is used today. The texts cover different disciplines, so as to be comparable to the BAWE corpus. Furthermore, the writers of the texts are both male and female; the distribution of male and female writers, as well as topics, text length and publication year is shown in table 9.

Table 9: Information about the NOBA texts

	Field	Year	Male/Female	Words
Text 1	Health Care	2010	Unknown	10,271
Text 2	Health Care	2010	Unknown	9,668
Text 3	Economics	2010	F + M	15,741
Text 4	Health Care	2010	F	10,626
Text 5	Computer Science	2003	M	10,262
Text 6	Sports Management	2009	M	8,582
Text 7	Sports Management	2010	F	5,993
Text 8	Economics	2010	M	14,975
Text 9	Economics	2010	F	12,458
Text 10	Project Management	2010	F	12,163
Text 11	Marketing	2010	M + F	14,151
Text 12	Economics	2010	F	22,848
Text 13	Technology	2010	M	7,803
Text 14	Pedagogy	2009	F	8,166
Text 15	History	2010	M	5,143

As an alternative to tagging the text parts which should be left out of the analysis, a simpler method was applied – deletion of quotes, diagrams, tables, lists, prefaces, abstracts and information about the writers. In this way, when the frequency of an item was counted in MS Word, only the items which were parts of the writers’ own language were included.

Advantages

Like the NOESS essays, the NOBA assignments are authentic – they are the results of bachelor students’ hard work and part of their education. Having two sources of Norwegian L1 writing is also an advantage, because if the results are similar in the two corpora, this will increase our confidence about the foundation for, and plausibility of, transfer from the L1, with respect to the use of writer/reader references.

Disadvantages

Unfortunately, NOBA is not a discipline-specific linguistics corpus. This is because it was difficult to find linguistics bachelor's assignments online. The disciplines with the highest number of bachelor's assignments published online were economics and health care, and I also managed to include some other disciplines, such as history and pedagogy, but I found no linguistics assignments in Norwegian. Due to the more general academic nature of NOBA, the comparisons with BAWE will include both the linguistics sub-corpus BAWE-ling and the whole BAWE corpus.

Another disadvantage with NOBA is that, although the total number of words is relatively high, it only consists of 15 texts, which means that there will be some uncertainty regarding the representativeness of the corpus. However, NOBA is by no means meant to be reliable for extensive research on student academic writing. The corpus can still be useful in a small research project such as this, especially since we study personal pronouns, which are relatively frequent elements in the Norwegian language.

3.5. Comparable information about the corpora and collections

Table 10 present the four main corpora and the two sources of Norwegian L1 writing. Most of the information in this table has already been discussed in the previous sections.

Table 10: Comparable information about NICLE, LOCNESS, NOESS, N-VESPA, BAWE and NOBA

	NICLE	LOCNESS	NOESS	N-VESPA	BAWE / BAWE-ling	NOBA
Language in the texts	L2 English	L1 English	L1 Norwegian	L2 English	L1 English	L1 Norwegian
Number of essays	316	290	13	194	<u>BAWE:</u> 1,953 <u>BAWE-ling:</u> 76	15
Number of words	210,367	324,400	14,761	231,675	<u>BAWE:</u> 4,534,973 <u>BAWE-ling:</u> 174,840	168,850
Genre	Argumentative	Argumentative	Argumentative	Academic	Academic	Academic
Topic / discipline	General argumentative topics	General argumentative topics (some literary topics)	General argumentative topics	Linguistics	<u>BAWE:</u> 30 disciplines <u>BAWE-ling:</u> Linguistics	9 disciplines
Home country	Norway	The UK and the US	Assumed to be Norway	Norway	Assumed to be the UK for most of the contributors	Assumed to be Norway
L1	Norwegian	English	Assumed to be Norwegian (L1 proficiency)	Norwegian	English	Assumed to be Norwegian (L1 proficiency)
Gender	Female: 74 % Male: 26 %	Unknown	Unknown	Female: 77 % Male: 23 %	<u>BAWE:</u> Female: 60 % Male: 40 % <u>BAWE-ling:</u> Female: 86 % Male: 14 %	Male: 33 % Female: 40 % Both (group work): 13 % Unknown: 13 %
Years of English at University	Less than 1 year: 93 %			More than 1 year: 91 %		

4. Results and analysis part 1: The preliminary study

In this preliminary study, I have compared the frequency of W/R visibility features in Norwegian learner writing and native speaker student writing, in order to get an impression of the general level of W/R visibility in Norwegian learner writing, as well as to identify the overused features. The differences between the learners and the native speakers have been tested for statistical significance, and this is referred to where relevant. Both general argumentative writing and discipline-specific academic writing has been considered, and these two genres have been studied separately. NICLE and LOCNESS represent respectively Norwegian learners' and native speaker students' argumentative writing, while N-VESPA and BAWE-ling represent Norwegian learners' and native speaker students' academic writing in the linguistics discipline. The list of items to be investigated in these four corpora, which was also presented in chapter 3, is repeated below.

- **First person singular pronouns** (closed group): *I, me, my, mine.*
- **Patterns of subjective stance** (open group): *I believe, I think, I don't (do not) think, I guess, I suppose, I would say, I would like to say, I'm (I am) sorry to say, I would argue, I must emphasize, I know, I remember, I'm (I am) not saying.*
- **First person plural pronouns** (closed group): *We, us, our, ours.*
- **Second person pronouns** (closed group): *You, your, yours.*
- **Patterns introducing other voices** (open group): *You could say that, you may say that, some may say that, some might say that, one can say that, many think, let's (let us) say that, you think, you could have guessed, you may ask, I hear you saying.*
- **Disjuncts** (open group): *of course, naturally, perhaps, maybe, unfortunately, obviously, frankly.*
- **Emphatic particles** (open group): *just, really.*
- **Reference to situation of writing/reading** (open group): *here, now, this essay.*
- **Sentence types** (open group): questions (question marks) and exclamations (exclamation marks).

Starting with the big picture, we will first have a brief look at the overall W/R visibility in learner writing, compared to native speaker writing. Moving on, a more detailed overview of the different features of W/R visibility in learner and native speaker writing will be presented, starting with the argumentative genre and continuing with the academic genre. Finally, the significant cases of overuse in both genres will be identified and discussed.

4.1. Overall writer/reader visibility in learner writing

When we look at the overall level of W/R visibility, it is important to keep in mind that all possible features have not been included in the study, so we cannot really claim that this is the actual overall level of W/R visibility, but this is as close as we get to the overall level in the present study. Figures 20 and 21 show the frequency of features of W/R visibility per 10,000 words in learner writing compared to native speaker writing. When adding up the frequencies of the different features, overlapping was prevented by excluding the frequency of the patterns which have 1st or 2nd person pronouns in them, so that these patterns were not counted twice. What is more, questions and exclamations are not included in the comparisons, as these features should not normally be considered in terms of frequency per 10,000 words, but rather in terms of frequency per a certain number of sentences or s-units.

The figures show the amount of W/R visibility features per 10,000 words, but since the list of features is not exhaustive, the most important factor to notice is not the actual frequencies, but the relationship between learner writing and native speaker writing. The overall level of W/R visibility in argumentative writing, which is illustrated in figure 20, is 2.8 times higher in NICLE than in LOCNESS²³, which indicates overuse of many features of W/R visibility in Norwegian learner writing in this genre. In the discipline-specific academic genre, the overuse is more moderate – the overall level of W/R visibility here is 1.7 times higher in N-VESPA than in BAWE-ling²⁴. Note also that the level of W/R visibility is higher in the argumentative genre than in the academic genre, both for learners and native speakers²⁵.

²³ The difference between NICLE and LOCNESS is statistically significant with 99.99 % certainty.

²⁴ The difference between N-VESPA and BAWE-ling is statistically significant with 95 % certainty.

²⁵ The difference between NICLE and N-VESPA as well as between LOCNESS and BAWE-ling were found to be statistically significant with at least 95 % certainty.

Figure 20: Overall level of W/R visibility in argumentative writing

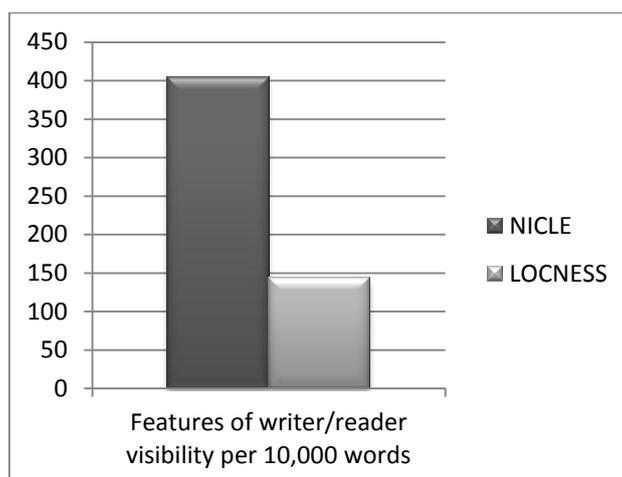
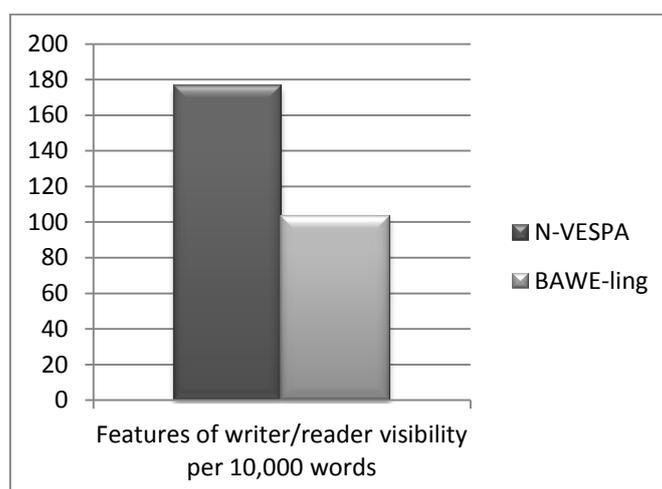


Figure 21: Overall level of W/R visibility in academic writing



4.2. Detailed overview of writer/reader visibility in argumentative writing

The detailed findings from the argumentative texts are presented in the following sections. Here, raw and normalized frequencies from NICLE and LOCNESS are shown in similarly structured tables, along with the degree of overuse in NICLE. NICLE overuse is identified by dividing the normalized frequency of an item in NICLE by the normalized frequency of that item in LOCNESS. Thus, values above 1 indicate overuse. For example, an overuse value of 2 for a certain item means that NICLE writers have used that item twice as frequently as

LOCNESS writers. Overuse values have been calculated for each individual item, but also for some groups – the closed groups, that is. (See section 3.1.1. for definitions of closed and open groups.) The differences between the corpora are checked for statistical significance. In some cases, the overuse value may be quite high and, at the same time, not statistically significant due to low raw frequencies which make the overuse value less reliable.

4.2.1. First person singular pronouns

As shown in table 11, NICLE writers overuse all the items in the “first person singular pronouns” group; consequently, the whole pronoun group is also overused. The total overuse value of 1st person sg pronouns in NICLE is relatively high, at 2.93. According to the log-likelihood calculator, the overuse is statistically significant²⁶. Furthermore, with a normalized frequency of 120.84, the Norwegian learners’ frequency is higher than the frequencies in Petch-Tyson’s (1998) learner groups, which ranged from 62.69 to 105.25 1st person sg pronouns per 10,000 words. Compared to Ädel’s study of timed and untimed Swedish learners, where the untimed learners had the lowest frequencies of pronouns, the Norwegian learners in the present study have frequencies approaching the timed Swedish learners, even though the Norwegian learners, on the whole, were not timed. This says even more about the Norwegian learners’ high level of W/R visibility.

Table 11: 1st person sg pronouns in NICLE and LOCNESS

Type of feature	Search words	NICLE		LOCNESS		NICLE (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
First person singular pronouns	I	1902	90.41	978	30.15	3.00
	Me	158	7.51	96	2.96	2.54
	My	474	22.53	260	8.01	2.81
	Mine	8	0.38	6	0.18	2.06
	Total	2542	120.84	1340	41.31	2.93

²⁶ 99.99 % certain.

4.2.2. Patterns of subjective stance

The patterns of subjective stance were found to have highly varying frequencies in argumentative learner writing. As also pointed out in Hasselgård (2009b), the most frequent patterns are *I believe* and *I think* with respectively 104 and 246 hits in NICLE. In addition to being frequent, these patterns are also found to be overused by as much as 2.97 and 4.36, and the overuse was found to be statistically significant²⁷. The less frequent patterns *I don't think*, *I guess* and *I would say* were also found to be significantly overused²⁸, as were *I know* and *I remember* – but with a lower degree of certainty²⁹. The rest of the patterns were very infrequent in both corpora, and not significantly overused in NICLE.

Table 12: Patterns of subjective stance in NICLE and LOCNESS

Type of feature	Search words	NICLE		LOCNESS		NICLE (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
Patterns of subjective stance	I believe	104	4.94	54	1.66	2.97
	I think	246	11.69	87	2.68	4.36
	I don't think	36	1.71	8	0,25	6.94
	I guess	21	1.00	6	0.18	5.40
	I suppose	4	0.19	3	0.09	2.06
	I would say	27	1.28	5	0.15	8.33
	I would like to say	2	0.10	1	0.03	3.08
	I'm sorry to say	1	0.05	0	0.00	# ³⁰
	I would argue	4	0.19	1	0.03	6.17
	I must emphasize	0	0.00	0	0.00	0
	I know	34	1.62	27	0.83	1.94
	I remember	14	0.67	5	0.15	4.32
	I'm not saying	4	0.19	0	0.00	#

²⁷ 99.99 % certain.

²⁸ 99.99 % certain.

²⁹ 99 % certain.

³⁰ The # sign is used to signalize that the overuse value of an item cannot be calculated because no uses of that item were found in the native speaker corpus.

4.2.3. First person plural pronouns

The investigation of 1st person pl pronouns shows that learners overuse all the items belonging to this group³¹, as was also the case for 1st person singular pronouns. However, since this is a closed group, the total frequency and the total overuse are what I have chosen to focus on. As shown in table 13, the overuse is quite high (2.84)³² and the raw frequency for this pronoun group is the highest raw frequency in this study (3258), which ensures the statistical significance of the overuse. Compared to the frequencies in Petch-Tyson (1998), the frequencies are relatively high, but not as high as the Swedish learners' frequencies. While Norwegian learners use 154.87 1st person pl pronouns per 10,000 words, Swedish learners use as many as 266.94. Still, Norwegian learners' level of overuse is higher than the overuse in Finnish, French and Dutch learner writing.

Table 13: 1st person pl pronouns in NICLE and LOCNESS

Type of feature	Search words	NICLE		LOCNESS		NICLE (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
First person plural pronouns	We	1944	92.41	922	28.42	3.25
	Us	401	19.06	258	7.95	2.40
	Our	905	43.02	584	18.00	2.39
	Ours	8	0.38	2	0.06	6.17
	Total	3258	154.87	1766	54.44	2.84

³¹ Statistically significant.

³² 99.99 % certain.

4.2.4. Second person pronouns

2nd person pronouns are less frequent than 1st person pronouns, but the raw frequency is relatively high (1712), compared to the frequency of the other items in this study. The overuse value is 3.97³³, which is higher than the overuse values for 1st person singular and plural pronouns. The item *yours* does not appear to have been overused, but this has no effect on the present study, since this is considered a closed group, which means that we focus on the total frequencies of 2nd person pronouns, and the total overuse. Compared to the learner groups in Petch-Tyson (1998), the total frequency of 2nd person pronouns is quite high in Norwegian learner writing. Whereas Petch-Tyson's learner groups had normalized frequencies between 44.26 and 80.81 per 10,000 words, the Norwegian learners in NICLE have a normalized frequency of 81.38 – in this case quite close to the learner group with the highest overuse in Petch-Tyson (1998), which was the Dutch learner group.

Table 14: 2nd person pronouns in NICLE and LOCNESS

Type of feature	Search words	NICLE		LOCNESS		NICLE (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
Second person pronouns	You	1346	63.98	540	16.65	3.84
	Your	365	17.35	122	3.76	4.61
	Yours	1	0.05	3	0.09	0.51
	Total	1712	81.38	665	20.50	3.97

³³ 99.99 % certain.

4.2.5. Patterns introducing other voices

The word patterns which were studied in the group called “patterns introducing other voices” were found to have very low raw frequencies in NICLE – from 0 to 8 hits per pattern, and even lower frequencies in LOCNESS. The differences which were found, were not statistically significant; therefore these patterns are not suitable for quantitative analysis. It might be better to analyze such patterns more extensively, including all possible patterns which introduce other voices, by manually reading through a sample of texts, which was how Hasselgård (2009b) identified these patterns in the first place.

Table 15: Patterns introducing other voices in NICLE and LOCNESS

Type of feature	Search words	NICLE		LOCNESS		NICLE (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
Patterns introducing other voices	You could say that	2	0.10	0	0.00	#
	You may say that	0	0.00	0	0.00	0
	Some may say that	1	0.05	0	0.00	#
	Some might say that	1	0.05	0	0.00	#
	One can say that	8	0.38	0	0.00	#
	Many think	1	0.05	0	0.00	#
	Let’s say that	0	0.00	0	0.00	0
	You think	7	0.33	6	0.18	1.80
	You could have guessed	1	0.05	0	0.00	#
	You may ask	2	0.10	0	0.00	#
	I hear you saying	1	0.05	0	0.00	#

4.2.6. Disjuncts

Disjuncts were found to vary in frequency in both corpora; the disjuncts *of course*, *naturally*, *unfortunately* and *obviously* were not overused at all, while *perhaps* appeared to be slightly overused, but this overuse was not statistically significant. The overuse value of *frankly* could not be calculated since the raw frequency in LOCNESS was 0, but, since *frankly* was not frequent in NICLE either, the potential overuse was not considered important. The only disjunct which was actually overused with statistical significance³⁴ was *maybe*, which had a very high overuse value (5.68).

Compared to the Swedish learners in Ädel's (2008) untimed SWICLE corpus, Norwegian learners generally seem to use disjuncts less frequently. However, *maybe* is used more often in Norwegian learner writing, than in all the corpora of Swedish learner writing. The level of overuse of *maybe* in Norwegian learner writing is also high compared to the general learner frequency in Paquot (2010), which was 2.80 per 10,000 words, compared to 8.94 in this investigation of NICLE.

Table 16: Disjuncts in NICLE and LOCNESS

Type of feature	Search words	NICLE		LOCNESS		NICLE (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
Disjuncts	Of course	1	0.05	75	2.31	0.02
	Naturally	8	0.38	19	0.59	0.65
	Perhaps	83	3.95	109	3.36	1.17
	Maybe	188	8.94	51	1.57	5.68
	Unfortunately	23	1.09	43	1.33	0.82
	Obviously	26	1.24	42	1.29	0.95
	Frankly	3	0.14	0	0.00	#

³⁴ 99.99 % certain.

4.2.7. Emphatic particles

The emphatic particles *just* and *really* are quite frequent in both corpora. Both items are overused in NICLE, and the overuse was found to be statistically significant with a high degree of certainty³⁵. Compared to the Finnish, Dutch, Swedish and French learners in Petch-Tyson (1998), the Norwegian learners do not have exceptionally high frequencies of emphatic particles. The learner frequencies of *just* in Petch-Tyson ranged from 8.27 to 24.60 per 10,000 words, and the Norwegian learner frequency was found to be 16.97. Petch-Tyson's learner frequencies for *really* ranged from 4.82 to 11.03, and the Norwegian learner frequency was found to be 8.89. Thus, the Norwegian learners' overuse seems to be quite average, or perhaps slightly above average, compared to the other learners' overuse, in this case.

Table 17: Emphatic particles in NICLE and LOCNESS

Type of feature	Search words	NICLE		LOCNESS		NICLE (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
Emphatic particles	Just	357	16.97	358	11.04	1.54
	Really	187	8.89	148	4.56	1.95

³⁵ 99.99 % certain.

4.2.8. Reference to situation of writing/reading

The items which refer to the situation of writing/reading are all overused in NICLE. Although these items have relatively low raw frequencies compared to the other items in this NICLE investigation, the overuse was found to be statistically significant³⁶. Compared to Petch-Tyson's (1998) findings, the frequencies in the present study seem very low, but this might be due to methodological choices. It appears as if Petch-Tyson may have counted all the occurrences of *here*, *now* and *this essay*, while I have only counted the occurrences which refer directly to the situation of writing/reading. This is also apparent when comparing the results from Petch-Tyson's control group and the control group used in the present study, which differ quite substantially. Accordingly, a comparison of the results in the present study and Petch-Tyson's findings, would not be reasonable, with respect to these particular items.

Table 18: Reference to situation of writing/reading in NICLE and LOCNESS

Type of feature	Search words	NICLE		LOCNESS		NICLE (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
Reference to situation of writing / reading	Here ³⁷	26	1.24	17	0.52	2.36
	Now ³⁸	16	0.76	7	0.22	3.52
	This essay ³⁹	57	2.71	11	0.34	7.99

³⁶ The statistical significance is 99 % certain for the overuse of *here* and *now*, and 99.99 % certain for the overuse of *this essay*.

³⁷ Only the uses of *here* which actually refer directly to the text have been counted.

³⁸ Only the uses of *now* which actually refer directly to the text have been counted.

³⁹ All the uses of *this essay* in NICLE and LOCNESS referred directly to the text.

4.2.9. Sentence types

Questions and exclamations, which, according to Ädel (2008), are sentence types with a high level of W/R visibility, are found to have high raw frequencies in NICLE, as well as quite high overuse values⁴⁰. The question form was found to be the most frequent of these sentence types (819 hits), while the exclamation form was the most overused sentence type in NICLE (overuse value: 5.03)⁴¹. Compared to the seven different learner groups in Virtanen (1998), the Norwegian learners were found to have the highest normalized frequency of questions, with 38.9 questions per 10,000 words, while Virtanen's learners had frequencies ranging from 17.3 to 36.1. The Norwegian learners in the present study were also found to have higher frequencies of both questions and exclamations than the Swedish learners in Ädel (2008).

Table 19: Sentence types (questions and exclamations) in NICLE and LOCNESS

Type of feature	Search items	NICLE		LOCNESS		NICLE (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
Questions	?	819	38.93	634	19.54	1.99
Exclamations	!	212	10.08	65	2.00	5.03

⁴⁰ Both sentence types were significantly overused with 99.99 % certainty.

⁴¹ Note that the sentence types are actually normalized per 10,000 words like all the other writer/reader visibility markers, and not per 1,000 sentences or s-units as would have been preferable. This is because I could not find information about the total number of s-units in NICLE and LOCNESS. However, Virtanen also uses this scale, so the advantage is that our respective findings are comparable.

4.3. Detailed overview of writer/reader visibility in academic writing

This section presents the findings from the academic writing in N-VESPA and BAWE-ling in the same way and order as the findings from NICLE and LOCNESS were presented in the former section.

4.3.1. First person singular pronouns

The general tendency in N-VESPA is that the overuse is more moderate than in NICLE, when comparing the Norwegian learners to their native speaker peers in BAWE-ling. As shown in table 20, 1st person sg pronouns are overused in N-VESPA, with the overuse value 1.87. In comparison, the same pronoun group was overused by 2.93 in NICLE. However, within N-VESPA this overuse value is relatively high, and, in addition, the overuse was found to be statistically significant with 99.99 % certainty.

With respect to previous research on academic learner writing, there are no directly comparable findings to refer to, but it was shown in Paquot, Hasselgård and Ebeling (forthcoming), that Norwegian learners in VESPA have a slightly higher frequency of 1st person pronouns⁴² than the French learners in VESPA. However, both Norwegian and French learners were found to overuse 1st person pronouns compared to native speaker students in the academic genre. The Hong Kong learners in Hyland's study (2002a), on the other hand, were found to underuse personal pronouns. Thus, the use of pronouns seems to be a general learner problem which may result in either underuse or overuse – possibly depending on EFL instruction and L1 norms.

Table 20: 1st person sg pronouns in N-VESPA and BAWE-ling

Type of feature	Search words	N-VESPA		BAWE-ling		N-VESPA (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
First person singular pronouns	I	1668	72.00	640	36.60	1.97
	Me	71	3.06	63	3.60	0.85
	My	338	14.59	133	7.61	1.92
	Mine	0	0.00	1	0.06	0
	Total	2077	89.65	837	47.87	1.87

⁴² This comparison includes reflexive pronouns, and both the singular and the plural forms of first person pronouns.

4.3.2. Patterns of subjective stance

In the academic corpora, most of the patterns of subjective stance have quite low frequencies. Some of the patterns were not found in any of the corpora, whereas others were only found in N-VESPA, which makes it impossible to calculate an overuse value and also to check whether the difference between the two corpora is statistically significant. However, it seems likely that the use of *I would say* in N-VESPA is a potential case of overuse, with the raw frequency 24 in N-VESPA versus 0 in BAWE-ling. Where it was possible to calculate overuse values, the overuse was only statistically significant for one pattern – *I would argue*⁴³ – which, in comparison, had the raw frequency 8 in N-VESPA and 1 in BAWE-ling.

Table 21: Patterns of subjective stance in N-VESPA and BAWE-ling

Type of feature	Search words	N-VESPA		BAWE-ling		N-VESPA (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
Patterns of subjective stance	I believe	38	1.64	19	1.09	1.51
	I think	39	1.68	24	1.37	1.23
	I don't think	3	0.13	0	0.00	#
	I guess	4	0.17	0	0.00	#
	I suppose	3	0.13	1	0.06	2.26
	I would say	24	1.04	0	0.00	#
	I would like to say	0	0.00	0	0.00	0
	I'm sorry to say	0	0.00	0	0.00	0
	I would argue	8	0.35	1	0.06	6.04
	I must emphasize	0	0.00	0	0.00	0
	I know	2	0.09	2	0.11	0.75
	I remember	0	0.00	0	0.00	0
	I'm not saying	0	0.00	0	0.00	0

⁴³ 95 % certain.

4.3.3. First person plural pronouns

1st person pl pronouns are relatively frequent, although only half as frequent as 1st person sg pronouns. The overuse value (1.58) is slightly lower than the overuse of 1st person sg pronouns in N-VESPA (which is 1.87), and much lower than the overuse of 1st person pl pronouns in NICLE (which is 2.84). Nevertheless, the overuse was found to be statistically significant with 99.99 % certainty.

Table 22: 1st person pl pronouns in N-VESPA and BAWE-ling

Type of feature	Search words	N-VESPA		BAWE-ling		N-VESPA (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
First person plural pronouns	We	831	35.87	334	19.10	1.88
	Us	149	6.43	74	4.23	1.52
	Our	54	2.33	85	4.86	0.48
	Ours	0	0.00	1	0.06	0
	Total	1034	44.63	494	28.25	1.58

4.3.4. Second person pronouns

2nd person pronouns have lower frequencies than 1st person pronouns, but are still quite frequent compared to the other markers of W/R visibility in this investigation of N-VESPA, when we look at the total frequency, that is. The overuse value is quite low, at 1.30, but so are most of the overuse values in the N-VESPA investigation. Furthermore, even though the overuse value is only 1.30, the log-likelihood test showed that the difference between the two corpora was statistically significant⁴⁴. Although the results from the present study may not be completely comparable to the findings in Paquot, Hasselgård and Ebeling (forthcoming), it may be interesting to note that their findings showed higher frequencies of 2nd person pronouns in the Norwegian part of VESPA than in the French part of VESPA.

Table 23: 2nd person pronouns in N-VESPA and BAWE-ling

Type of feature	Search words	N-VESPA		BAWE-ling		N-VESPA (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
Second person pronouns	You	256	11.05	131	7.49	1.47
	Your	22	0.95	30	1.72	0.55
	Yours	0	0.00	0	0.00	0
	Total	278	12.00	161	9.21	1.30

⁴⁴ 99 % certain.

4.3.5. Patterns introducing other voices

The patterns introducing other voices are the least frequent W/R visibility markers in this investigation. In N-VESPA none of the patterns appear more than 3 times, which means that there is no reason for studying them further in the present study.

Table 24: Patterns introducing other voices in N-VESPA and BAWE-ling

Type of feature	Search words	N-VESPA		BAWE-ling		N-VESPA (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
Patterns introducing other voices	You could say that	0	0.00	0	0.00	0
	You may say that	0	0.00	0	0.00	0
	Some may say that	0	0.00	0	0.00	0
	Some might say that	0	0.00	0	0.00	0
	One can say that	3	0.13	0	0.00	#
	Many think	0	0.00	0	0.00	0
	Let's say that	1	0.04	0	0.00	#
	You think	2	0.09	4	0.23	0.38
	You could have guessed	0	0.00	0	0.00	0
	You may ask	0	0.00	0	0.00	0
	I hear you saying	0	0.00	0	0.00	0

4.3.6. Disjuncts

When it comes to disjuncts, the overuse in N-VESPA is actually more significant than the overuse in NICLE. While, in NICLE only *maybe* was found to be significantly overused, in N-VESPA *maybe*, *of course* and *perhaps* were significantly overused⁴⁵. These three items were also studied in Paquot, Hasselgård and Ebeling (forthcoming)⁴⁶. Their study also found overuse of *maybe*, *of course* and *perhaps* in Norwegian academic learner writing, and, in addition, they found that the frequencies of these items in the Norwegian part of VESPA were higher than the frequencies in the French part of VESPA. Finally, returning to the findings in the present study, the other disjuncts were either not overused at all, or not significantly overused.

Table 25: Disjuncts in N-VESPA and BAWE-ling

Type of feature	Search words	N-VESPA		BAWE-ling		N-VESPA (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
Disjuncts	Of course	41	1.77	10	0.57	3.09
	Naturally	15	0.65	20	1.14	0.57
	Perhaps	88	3.80	39	2.23	1.70
	Maybe	41	1.77	6	0.34	5.16
	Unfortunately	6	0.26	1	0.06	4.53
	Obviously	32	1.38	15	0.86	1.61
	Frankly	0	0.00	0	0.00	0

⁴⁵ With respectively 99.99 %, 99.9 % and 99 % certainty.

⁴⁶ Here defined as epistemic modal adverbs.

4.3.7. Emphatic particles

While emphatic particles were significantly overused in NICLE, this is not the case in N-VESPA. *Just* is not overused at all, and the overuse of *really* is not statistically significant.

Table 26: Emphatic particles in N-VESPA and BAWE-ling

Type of feature	Search words	N-VESPA		BAWE-ling		N-VESPA (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
Emphatic particles	Just	95	4.10	83	4.75	0.86
	Really	45	1.94	23	1.32	1.48

4.3.8. Reference to situation of writing/reading

The reference words *here*, *now* and *this essay* were used more frequently about the situation of writing/reading in N-VESPA than in BAWE-ling. However, the differences between the corpora were only found to be statistically significant for the item *here*⁴⁷, which also has a relatively high overuse value (3.10).

Table 27: Reference to situation of writing/reading in N-VESPA and BAWE-ling

Type of feature	Search words	N-VESPA		BAWE-ling		N-VESPA (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
Reference to situation of writing / reading	Here	230	9.93	56	3.20	3.10
	Now	31	1.34	13	0.74	1.80
	This essay	67	2.89	42	2.40	1.20

⁴⁷ 99.99 % certain.

4.3.9. Sentence types

Whereas, in NICLE, both questions and exclamations were found to be overused, in N-VESPA neither of these sentence types are used more frequently than in the native speaker corpus BAWE-ling. These findings are slightly surprising, since Paquot, Hasselgård and Ebeling's study (forthcoming) of VESPA and BAWE showed overuse of questions in the Norwegian part of VESPA (section 2.4.2.). The differences between this study and the present study may be related to differences in corpus selections, or due to methods. For example, in the present study, the raw frequencies are based on the automatic calculations by the concordancers, and not studied in detail, manually.

Type of feature	Search words	N-VESPA		BAWE-ling		N-VESPA (over)use
		Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
Questions	?	144	6.22	167	9.55	0.65
Exclamations	!	4	0.17	17	0.97	0.18

4.4. Summing up the preliminary study

The corpus comparisons in this part of the study have shown that Norwegian learners tend to overuse features of W/R visibility compared to native speakers. Comparisons of the results from the present study and the findings from previous studies show that Norwegian learners, in many cases, have a higher degree of overuse than learners with other L1s. However, the Norwegian learners do not overuse all the features in the study. In the tables below, all the features of W/R visibility which have been shown to be significantly overused in learner writing are listed, ranged according to their overuse value. The overused features in NICLE are shown in table 28, and the overused features in N-VESPA are shown in table 29. The mere size of the tables indicates the difference between the two learner corpora – a larger amount of features are significantly overused in NICLE than in N-VESPA. However, the feature groups with the highest raw frequencies, namely the different pronoun groups, are

overused in both NICLE and N-VESPA. Only two additional features are overused in both corpora of learner writing, namely *maybe* and *here*.

In addition to being significantly overused in both argumentative and academic learner writing, 1st and 2nd person pronouns are also very frequent – more frequent than any of the other features. Furthermore, 1st and 2nd person pronouns are very obvious W/R visibility features. These reasons, combined, explain why 1st and 2nd person pronouns have been chosen for the in-depth study. Another advantage of studying pronouns more in depth, is that pronouns are quite easily compared across languages, at least in Norwegian and English, because most of the pronouns have close equivalents in the other language, such as *jeg/I* and *meg/me*. This is not to say that the pronouns are necessarily used in the same way in Norwegian and English, but their apparent similarity makes them easier to compare than for example *maybe*, which might be translated into both *kanskje* and *muligens*, which might again be translated into *maybe*, *perhaps* and *possibly*.

Table 28: The significantly overused features in NICLE

Features	NICLE		LOCNESS		NICLE (over)use
	Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
I would say	27	1.28	5	0.15	8.33
This essay	57	2.71	11	0.34	7.97
I don't think	36	1.71	8	0.25	6.94
Maybe	188	8.94	51	1.57	5.68
I guess	21	1.00	6	0.18	5.40
Exclamations	212	10.08	65	2.00	5.03
I think	246	11.69	87	2.68	4.36
I remember	14	0.67	5	0.15	4.32
2nd person pronouns	1712	81.38	665	20.50	3.97
Now	16	0.76	7	0.22	3.45
I believe	104	4.94	54	1.66	2.97
1st person sg pronouns	2542	120.84	1340	41.31	2.93
1st person pl pronouns	3258	154.87	1766	54.44	2.84
Here	26	1.24	17	0.52	2.38
Questions	819	38.93	634	19.54	1.99
Really	187	8.89	148	4.56	1.95
I know	34	1.62	27	0.83	1.94
Just	357	16.97	358	11.04	1.54

Table 29: The significantly overused features in N-VESPA

Features	N-VESPA		BAWE-ling		N-VESPA (over)use
	Raw frequencies	Per 10.000 words	Raw frequencies	Per 10.000 words	
I would argue	8	0.35	1	0.06	6.04
Maybe	41	1.77	6	0.34	5.16
Here	230	9.93	56	3.20	3.10
Of course	41	1.77	10	0.57	3.09
1st person sg pronouns	2077	89.65	837	47.87	1.87
Perhaps	88	3.80	39	2.23	1.70
1st person pl pronouns	1034	44.63	494	28.25	1.58
2nd person pronouns	278	12.00	161	9.21	1.30

5. Results and analysis part 2: The quantitative in-depth study

In this part of the study, 1st and 2nd person pronouns are investigated more thoroughly with the purpose of explaining the high level of W/R visibility in learner writing. As we have seen from the results in the preliminary study, Norwegian learners overuse 1st and 2nd person pronouns significantly, both in the argumentative and in the academic genre. What I have tried to determine in the second part of the study is to what extent this overuse may have been caused by L1 transfer. I also found it interesting to explore the possibility that increasing levels of academic experience might lead the students towards a more native-like level of W/R visibility, and that this might be the reason why the NICLE students have a higher overuse of W/R visibility features than the N-VESPA students.

5.1. The effect of L1 transfer

I have two approaches to the investigation of the L1 transfer effect on L2 writing, which are inspired by Gilquin's (2008) article about the DEE transfer model (Detection – Explanation – Evaluation). First, L1 English writing is compared to L1 Norwegian writing to see whether there is any reason why L1 transfer from the Norwegian language would lead to a higher level of W/R visibility. Since the present study is mainly a study of student writing, I chose to use a Norwegian source of student writing when comparing Norwegian and English. Consequently, the English student writing in LOCNESS was compared to NOESS, and the academic student writing in BAWE and BAWE-ling was compared to my own collection of Norwegian bachelor assignments called NOBA. In theory, if the level of W/R visibility is higher in Norwegian L1 writing than in English L1 writing, there is reason to believe that L1 transfer could be a reason for the overuse in English L2 writing by Norwegian learners.

The second approach to the investigation of L1 transfer in Norwegian learner writing is to compare Norwegian learners to learners with other L1 backgrounds. In the present study, the NICLE corpus of Norwegian learner English is compared to the ICLE corpus, which comprises writing by learners with numerous L1 backgrounds. NICLE writers are also compared to specific L1 groups from Petch-Tyson's study (1998), which was presented in chapter 2. N-VESPA writers will not be compared to any other learner groups in the present study, due to lack of comparable learner corpora. However, the comparison of Norwegian and French learners in VESPA by Paquot, Hasselgård and Ebeling (forthcoming) has already

showed that the Norwegian learners' overuse of 1st and 2nd person pronouns is higher than the French learners' overuse (see section 2.4.2.). Paquot, Hasselgård and Ebeling also consider the Norwegian and French parts of ICLE, and find similar tendencies there. If the present study finds that the overuse of pronouns in Norwegian learner writing is more significant than the overuse in other learner groups, in addition to the French learner group, this increases the credibility of the hypothesis that Norwegian learners transfer some of their L1 norms onto their L2 writing.

5.1.1. Argumentative writing in Norwegian and English

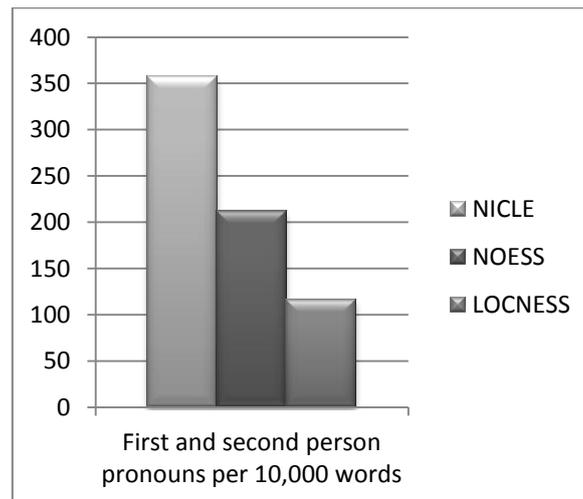
In Gilquin's (2008) DEE transfer model, comparing L1 writing in two languages is just one of six steps which help detect, explain and evaluate the pedagogical implications of L1 transfer. The comparison of L1 writing in two languages belongs to the explanatory part of Gilquin's transfer model, and may be classified as a regular contrastive analysis. If I were to follow Gilquin's recommendations, I would also have performed a contrastive analysis with a parallel corpus. However, I would then need to have access to a corpus of argumentative writing (and academic writing in the next part) with translations from Norwegian into English and vice versa, and, to my knowledge, no such corpus exists. Consequently, only L1 original texts were compared.

L2 English by Norwegian learners was also included in the comparison. Thus, another step from Gilquin's DEE transfer model was used – the comparison of native English versus learner English to examine overuse in learner writing, which belongs to the part about evaluation of the pedagogical implications of L1 transfer in Gilquin. In the present study, this particular comparison has already been made, and the overuse has already been examined. Still, including learner English in the present comparison was considered important in order to be able to evaluate the plausibility of L1 transfer as a reason for the Norwegian learners' overuse of pronouns.

Figure 22 shows the findings from the comparison of L2 English by Norwegian learners (NICLE), L1 Norwegian (NOESS) and L1 English (LOCNESS). The frequency of 1st and 2nd person pronouns per 10,000 words is shown to be 1.7 times higher in L2 English by Norwegian learners than in Norwegian L1 writing, which means that L1 transfer is probably not the only reason for the learners' overuse of pronouns. However, the differences between

L1 Norwegian and L1 English are statistically significant⁴⁸, so there is reason to believe that there is more room for using 1st and 2nd person pronouns in Norwegian argumentative writing than in English argumentative writing, which might have affected the Norwegian learners' use of these pronouns in their L2 writing.

Figure 22: 1st and 2nd person pronouns per 10,000 words in L2 English, L1 Norwegian and L1 English (argumentative writing)



When looking at the different pronoun groups individually, however, the L1 transfer investigation becomes more complicated. With respect to 1st person sg pronouns (figure 23), the high frequency in English L2 writing does not appear to have been caused by L1 transfer, as the frequency in L1 Norwegian writing is not significantly higher than the frequency in L1 English writing. The same seems to be the case with respect to 2nd person pronouns (figure 25), which are, in fact, less frequent in Norwegian L1 than in English L1⁴⁹. When it comes to 1st person pl pronouns (figure 24), on the other hand, transfer-induced overuse seems quite plausible, since there is a significant difference between the frequencies in Norwegian L1 writing and English L1 writing⁵⁰, whereas the frequencies in Norwegian L1 writing and English L2 writing by Norwegian learners are similar.

⁴⁸ 99.99 % certain.

⁴⁹ 99 % certain.

⁵⁰ 99.99 % certain.

Figure 23: 1st person sg pronouns per 10,000 words in L2 English, L1 Norwegian and L1 English (argumentative writing)

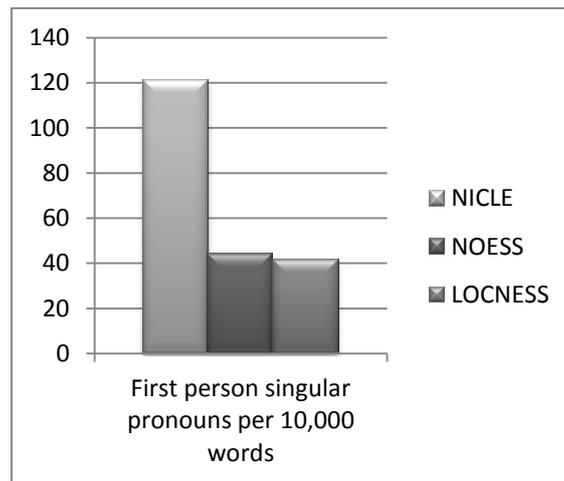


Figure 24: 1st person pl pronouns per 10,000 words in L2 English, L1 Norwegian and L1 English (argumentative writing)

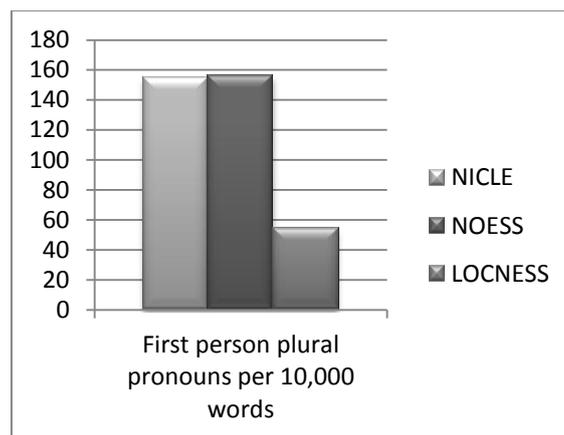
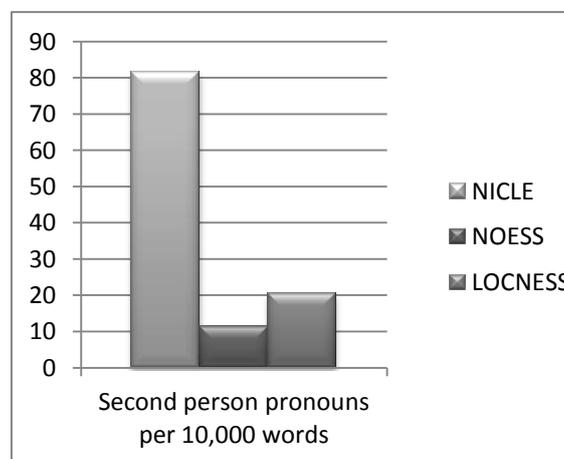


Figure 25: 2nd person pronouns per 10,000 words in L2 English, L1 Norwegian and L1 English (argumentative writing)



In total, the findings from these comparisons are more confusing than they are explanatory of the L1 transfer effect. However, the overuse of 1st person plural pronouns seems to have been caused by L1 transfer. However, the NOESS corpus is quite small, so the reliability of any findings is questionable. Perhaps a study of a larger Norwegian L1 corpus would solve the mystery, or a similar corpus investigation of academic student writing, which will be presented in the following section.

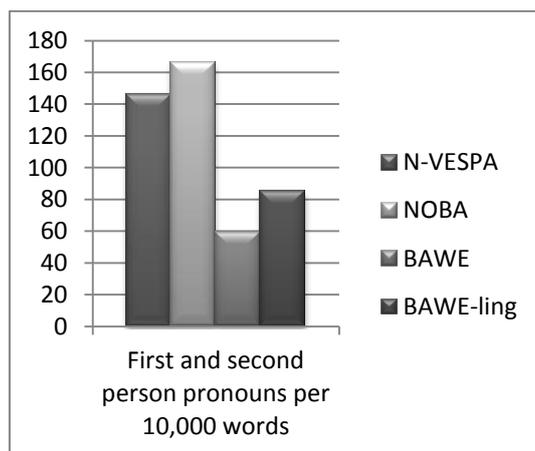
5.1.2. Academic writing in Norwegian and English

Figure 26 compares L2 English by Norwegian learners (N-VESPA) to L1 Norwegian (NOBA) and L1 English (BAWE and BAWE-ling) with respect to the frequency of 1st and 2nd person pronouns. The reason why both BAWE and BAWE-ling are represented is that, while BAWE-ling is the best comparable corpus to N-VESPA, the whole BAWE corpus is more comparable to the Norwegian L1 source NOBA, in terms of disciplines. In figure 26, the total frequency of 1st and 2nd person pronouns are compared across the four corpora. If these frequencies are, in fact, reliable, transfer from the L1 indeed seems like a plausible explanation for the overuse, because the frequency is significantly higher in the L1 Norwegian source than in both L1 English sources⁵¹. In addition, the frequency in the L2 English source is also significantly higher than the L1 English sources⁵². This is a perfect example of how the frequencies should be distributed, if transfer-induced overuse was going to be detected. However, also in this case, it is important to look at the frequencies for the individual pronoun groups as well, to see whether transfer seems to be the reason for all of the overuse, or just some of it.

⁵¹ 99.99 % certain.

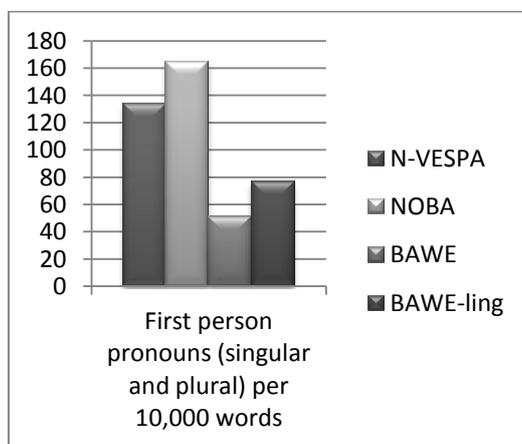
⁵² 99.99 % certain.

Figure 26: 1st and 2nd person pronouns per 10,000 words in L2 English, L1 Norwegian and L1 English (academic writing)



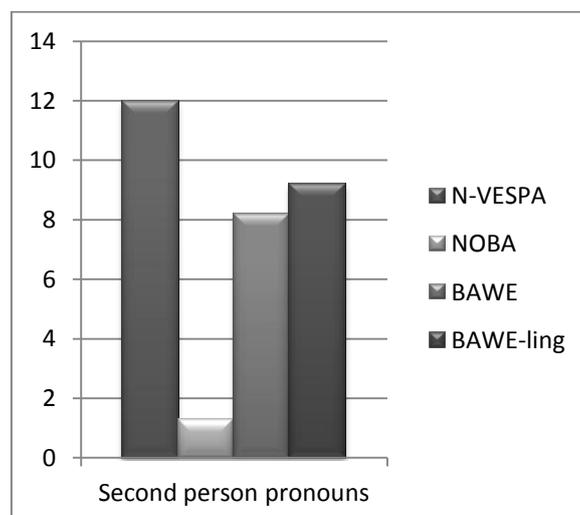
Since the NOBA corpus of Norwegian L1 writing includes several texts which have been written by more than one student, the distribution of the singular and plural form is likely to have been affected by this. Therefore, in this comparison, I have chosen to look at 1st person singular and plural at the same time. As shown in figure 27, the distribution of 1st person pronouns in the four corpora is similar to the distribution we saw in the former figure. This is related to the fact that 1st person pronouns are a lot more frequent than 2nd person pronouns in all the corpora, which means that they constitute the major part of the frequencies in the former comparison. Thus, L1 transfer still seems a possible explanation for the learner overuse, when looking at 1st person pronouns.

Figure 27: 1st person pronouns (singular and plural) per 10,000 words in L2 English, L1 Norwegian and L1 English (academic writing)



When it comes to 2nd person pronouns, the findings are different. 2nd person pronouns were found to have a very low frequency in the Norwegian L1 source NOBA. Since this was also the case in NOESS, the low frequency of 2nd person pronouns may in fact be characteristic of the Norwegian language. A possible explanation to the low frequency of 2nd person pronouns in Norwegian L1 writing and the high frequency in English L2 writing by Norwegian learners, might be related to the generic use of *you*. In Norwegian the generic pronouns *man* and *en*, which are similar to the generic pronoun *one* in English, are quite common – probably more common than the English *one*, because they are not considered too formal. If the learners know that it is not always appropriate to use the generic *one*, they might choose to use the generic *you* instead, where they would have used *man* or *en* in Norwegian. In this way, there might have been some L1 transfer going on, which does not show in this comparison. Unfortunately, this suggestion has to remain a speculation in the present study, as there was no time to investigate this possibility further.

Figure 28: 2nd person pronouns per 10,000 words in L2 English, L1 Norwegian and L1 English (academic writing)



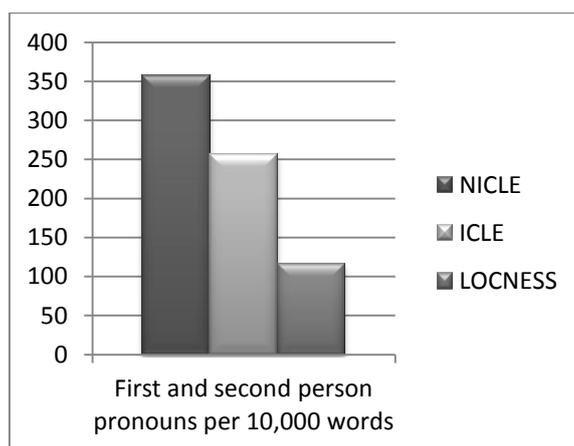
5.1.3. Norwegian learners and learners with other L1s

It proved difficult to determine whether pronouns are more accepted in Norwegian student writing than in English student writing, and whether this influences Norwegian learners' writing in English. However, another approach to the investigation of the L1 transfer effect led to more confident conclusions, namely the comparison of Norwegian learners and learners

with other L1s. This approach was also inspired by Gilquin's article (2008) about the DEE transfer model. Comparing different learner varieties of English is one of the steps related to detection of L1 transfer in Gilquin's transfer model. The comparison may be defined as a contrastive interlanguage analysis – interlanguage being the same as learner language.

The easiest way to find out whether Norwegian learners use pronouns more frequently than other learners is to compare them to a large group of learners with different L1 backgrounds. To my knowledge, the ICLE corpus is the largest available corpus of learner English, and it is also directly comparable to one of the sources of Norwegian learner writing in the present study, namely NICLE, which is a sub-corpus of ICLE. As opposed to the comparisons in the two previous sections, which were slightly confusing, the comparisons in this part of the study show refreshingly clear tendencies. Norwegian learners use more 1st and 2nd person pronouns per 10,000 words than ICLE learners in general, both when we look at the combined frequencies (figure 29) and when we look at the different pronoun groups⁵³ (figures 30-32). Additionally, we find overuse of the different pronoun groups in both NICLE and ICLE. This means that overuse in Norwegian learner writing might partly be caused by general language learning problems shared by different learner groups, and partly by problems specific to the Norwegian learner group, such as L1 transfer, for example.

Figure 29: 1st and 2nd person pronouns in NICLE, ICLE and LOCNESS



⁵³ All the differences in the ICLE/NICLE/LOCNESS comparisons were found to be significant with 99,99 % certainty.

Figure 30: 1st person sg pronouns in NICLE, ICLE and LOCNESS

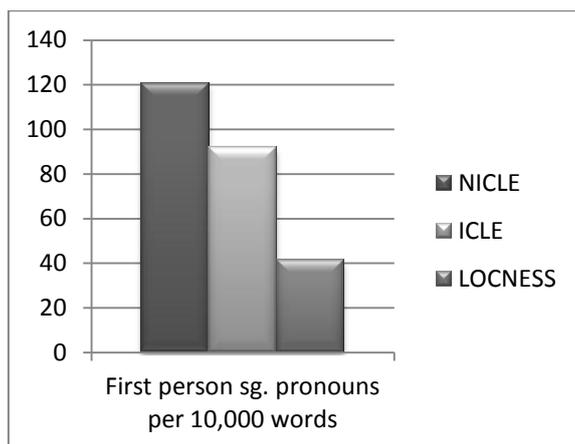


Figure 31: 1st person pl pronouns in NICLE, ICLE and LOCNESS

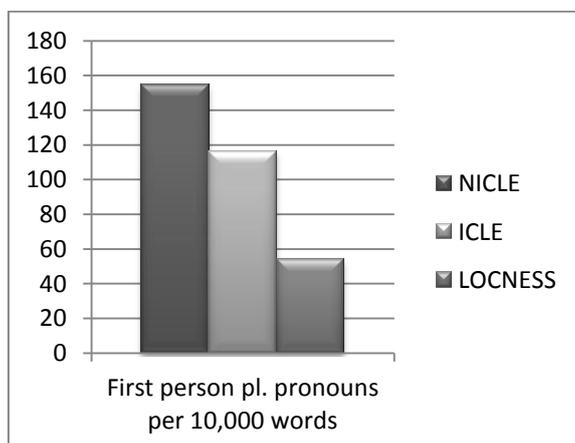
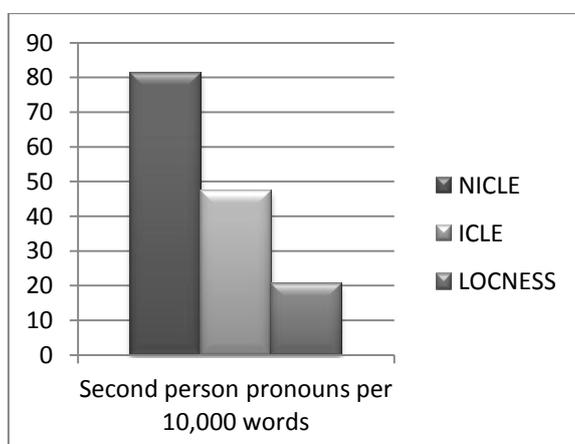


Figure 32: 2nd person pronouns in NICLE, ICLE and LOCNESS

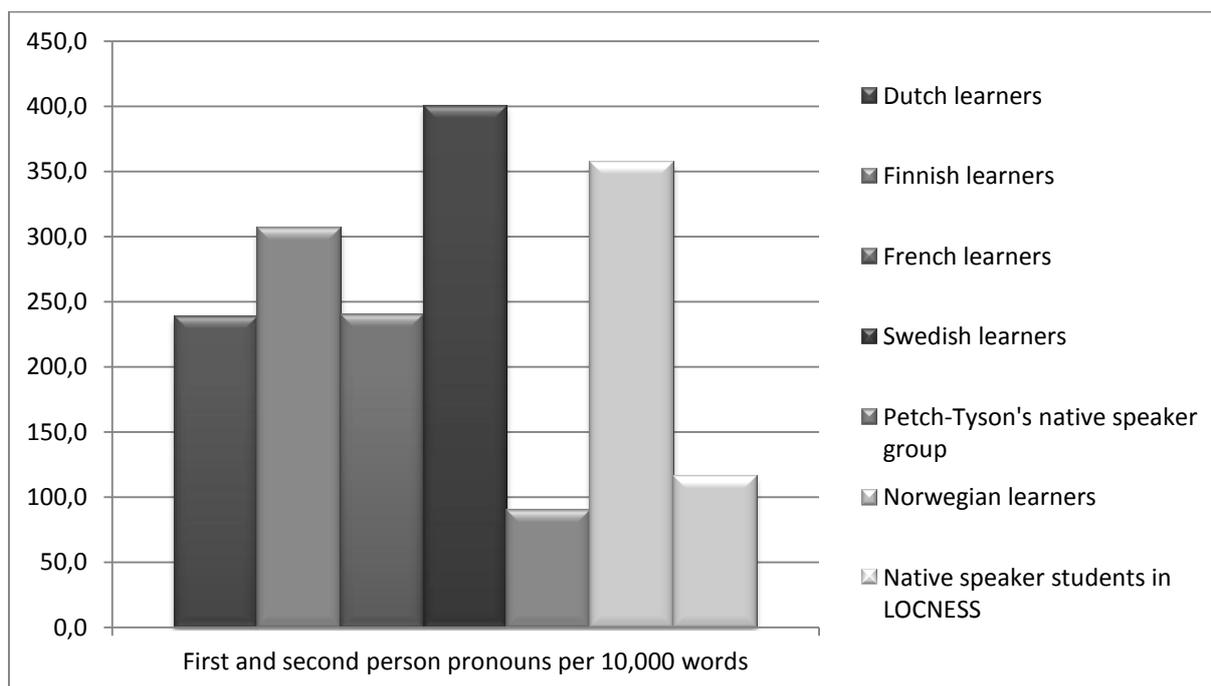


Although it is useful to look at all the learners in ICLE to search for general learner problems, it is essential that we also look at the variation between the learner groups. Instead of performing additional corpus investigations to find pronoun frequencies from different learner groups, I have chosen to compare my own findings from NICLE to Petch-Tyson's (1998) findings from the Dutch, Finnish, French and Swedish sub-corpora of ICLE, which were also referred to in chapter 2. Figure 33 shows the total frequency of 1st and 2nd person pronouns per 10,000 words in each corpus. The Swedish learners have the highest frequency of these pronouns, followed by the Norwegian learners⁵⁴, which is interesting with respect to the potential transfer effect, since Swedish and Norwegian are quite similar languages. Since the Finnish learner group has the third highest frequency⁵⁵, culture may also be of influence, as the Scandinavian countries are likely to share some cultural values and norms – perhaps even academic values and norms. The Dutch and French learners in Petch-Tyson are below the average ICLE frequency found in the present study, which was 255.5 pronouns (1st and 2nd person) per 10,000 words. In addition, the Dutch and French learner frequencies are significantly lower than the Swedish, Norwegian and Finnish learner frequencies. However, all the learners in these studies overuse 1st and 2nd person pronouns compared to native speakers, which supports the hypothesis that overuse of pronouns is a general learner problem which for some learner groups might be intensified due to transfer from the L1. The overuse might also be intensified because of cultural differences. According to Petch-Tyson, studies have shown that “(...) there is cultural variability in the levels of (acceptability of) interpersonal involvement in discourse (...)” (1998, 107).

⁵⁴ The difference between the Swedish and Norwegian learners was found to be significant.

⁵⁵ The difference between the Norwegian and Finnish learners was also found to be significant.

Figure 33: 1st and 2nd person pronouns in the writing of learners with different L1s, partly based on table 8.1 in Petch-Tyson (1998)



The distribution of the different pronoun groups in the different corpora is presented in table 30. Here we can see that the Finnish learners have the highest overuse of 1st person singular pronouns, while the Swedish learners have the highest overuse of 1st person pl pronouns. Norwegian learners have the highest overuse of 2nd person pronouns, closely followed by Dutch learners, who in contrast to the other learner groups have quite an even distribution of the different pronouns. The most important tendency to note from this table is the general tendency of all learners to overuse all the pronoun groups represented here, both when compared to Petch-Tyson's native speaker group and my own – LOCNESS.

Table 30: The distribution of the different pronoun groups in learner writing and native speaker writing, partly based on table 8.1 in Petch-Tyson (1998)

Pronouns	Petch-Tyson (1998)					The present study	
	Dutch ICLE	Finnish ICLE	French ICLE	Swedish ICLE	US native speakers	NICLE	LOCNESS
First person singular	70.7	105.3	62.7	88.1	30.9	120.8	41.3
First person plural	87.5	134.1	133.5	266.9	44.8	154.9	54.4
Second person	80.8	66.9	44.3	44.6	14.1	81.4	20.5
Total	239.0	306.3	240.4	399.6	89.8	357.1	116.2

The conclusion from these comparisons must be that it seems likely that transfer from Norwegian might have caused some of the overuse of 1st and 2nd person pronouns. This might be due to differences between the Norwegian and English languages when it comes to expressing opinions and referring to personal experiences, or due to differences in terms of the general acceptance of interpersonality in argumentative and academic writing.

5.1.4. Final remarks on the L1 transfer issue

Comparisons of original texts in L1 Norwegian and L1 English showed that transfer from the L1 might possibly be a reason for the overuse of W/R visibility features in L2 English writing by Norwegian learners. This seemed particularly plausible with respect to the overuse of 1st person pl pronouns in the argumentative genre and 1st person pronouns in general in the academic genre. Comparisons of Norwegian learners and other learner groups also support the L1 transfer hypothesis. Further research on the L1 transfer effect is needed in order to explain the details of what is being transferred from either the Norwegian language or the Norwegian culture. This would require good corpora of L1 Norwegian writing, which ideally would include both L1 student writing and L1 professional writing in different genres and disciplines, since the use of W/R visibility features, may vary according to genre, discipline and academic level.

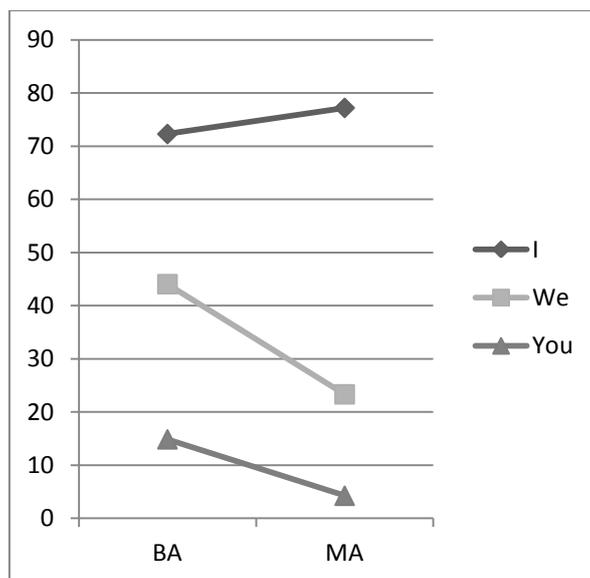
5.2. The effect of academic experience

The writers in N-VESPA use 1st and 2nd person pronouns less frequently than the writers in NICLE, but it is difficult to determine the exact reasons for this, since many variables distinguish the two corpora. One reason might be that the N-VESPA writers have more academic experience than the NICLE writers. Since the N-VESPA writers all have English courses as part of their education, the increase in academic experience may also imply an increase in English proficiency in their case. In order to investigate the effect of academic experience on learners' use of pronouns, I have decided to compare two student groups within N-VESPA – bachelor students and master students. Hopefully, the only factor dividing these two groups is the educational level (and probably the age). Note that, due to limitations in

time, I have only studied the frequency of the pronouns *I*, *we* and *you* in this corpus investigation.

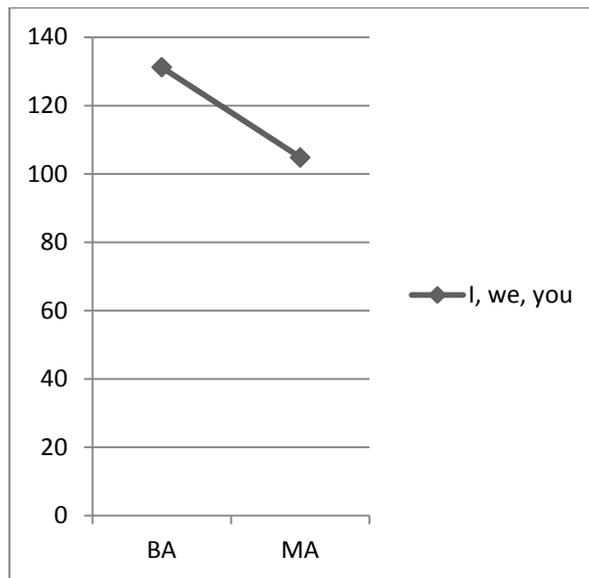
Figure 34 shows the development in the use of *I*, *we* and *you* from bachelor level to master level in the L2 writing by Norwegian learners in N-VESPA. The statistical tests showed that the increase in the use of *I* was not significant, whereas the decrease in the use of *we* and *you* was significant with 99.99 % certainty. The decrease in the use of *I*, *we* and *you* combined, which is shown in figure 35, was also found to be statistically significant⁵⁶. This means that academic differences between NICLE and N-VESPA might partly have caused the differences in pronoun usage. It also means that academic experience, combined with the age factor and the hopefully increasing level of proficiency in English, appears to affect Norwegian learners in a positive way – that is, if the aim is to become more native-like when it comes to pronoun use.

Figure 34: *I*, *we* and *you*, separately, in the writing of bachelor and master students in N-VESPA (per 10,000 words)



⁵⁶ 99.99 % certain.

Figure 35: *I, we and you, combined, in the writing of bachelor and master students in N-VESPA (per 10,000 words)*



It might be interesting to see if there are similar tendencies in native speaker student writing as well. If so, this would imply that students become less personally involved in their texts as they reach higher levels of education – independently of whether they write in their L1 or their L2. As the diagrams below illustrate, this appears to be the reality. Here, students from levels 1 and 2, as specified in the BAWE corpus, are compared to students from levels 3 and 4. The frequencies of the pronouns *I, we* and *you* all seem to decrease as the students reach levels 3 and 4. The decrease was found to be statistically significant for the pronouns *I* and *we*⁵⁷, but not for *you*. However, the total decrease in the use of these pronouns was significant⁵⁸.

⁵⁷ 99.99 % certain.

⁵⁸ 99.99 % certain.

Figure 36: *I, we and you, separately, in BAWE-ling at different levels (per 10,000 words)*

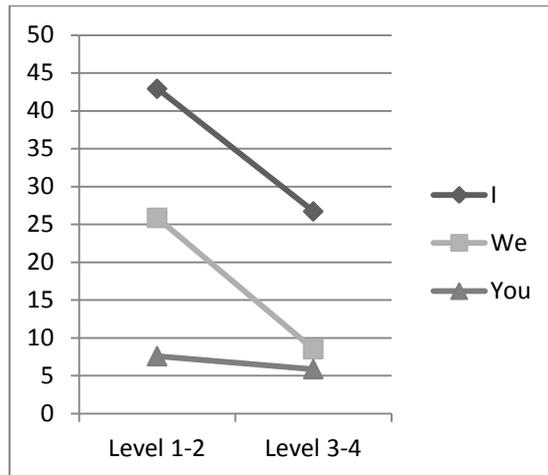
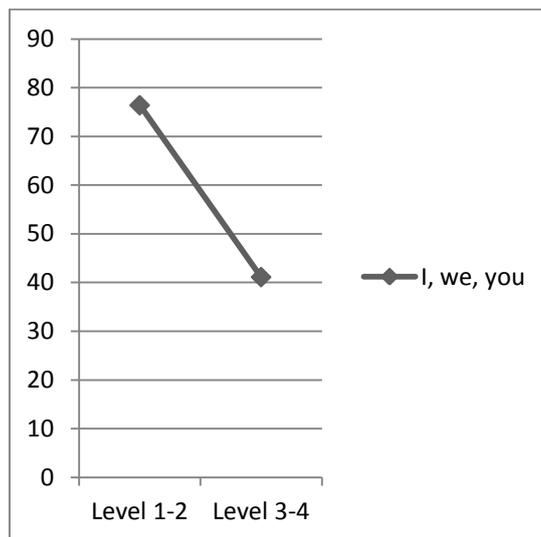


Figure 37: *: I, we and you, combined, in BAWE-ling at different levels (per 10,000 words)*



What this investigation can tell us is that, as students advance academically, their writing becomes less personal, whether English is their L1 or their L2. This might have to do with the target language they meet in their academic literature. We might thus expect the academic literature in the linguistics discipline to be less personally involved than the student writing in BAWE-ling. Unfortunately, I have not been able to find a good corpus of professional writing in the linguistics discipline. The British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), which are well-known native speaker corpora, do not seem to offer the opportunity to distinguish between disciplines in their academic

sections. If a professional linguistics corpus does exist, or will be collected, comparing students and professional academics within this discipline might be an interesting topic for future research.

5.3. Summing up the quantitative in-depth study

What this quantitative in-depth study of 1st and 2nd person pronouns can tell us about Norwegian learner writing is that, first of all, it seems likely that L1 transfer and general language learning problems could be sources of the overuse of pronouns. This is indicated by the general overuse tendency in learner writing, combined with the variable levels of overuse in the different learner groups. Comparisons of Norwegian and English L1 sources also show that, in some cases, L1 transfer seems likely. However, the Norwegian L1 corpora are too small to be completely reliable, so more research is needed to prove this hypothesis. The second part of the quantitative study of different parts of N-VESPA and BAWE-ling show that academic experience affects students in a way which makes their writing less personal. Both learners and native speaker students seem to move towards a more objective style.

6. Results and analysis part 3: The qualitative in-depth study

The aim of the qualitative in-depth study was to categorize the different uses of 1st and 2nd person pronouns in learner writing and native speaker writing in order to find out whether some uses were overused by learners while others were not. This was done by reading through a certain number of concordance lines for the pronouns in the four main corpora – NICLE, LOCNESS, N-VESPA and BAWE-ling. Due to the time-consuming nature of such readings, only the items *I*, *we* and *you* were included in the study, representing 1st person sg, 1st person pl and 2nd person pronouns. Also, instead of studying all the occurrences of *I*, *we* and *you*, only samples were studied, ranging from 100 to 200 concordance lines.

The reason why the samples are not of the same size is that I originally intended to study only N-VESPA and BAWE-ling qualitatively, because most of my reference literature on the use of personal pronouns considered the functions of these pronouns in academic writing. Therefore, I began by studying *I*, *we* and *you* in academic student writing only, in samples of 200 concordance lines per pronoun. However, *you* was only used 131 times in BAWE-ling, so to make the investigation more reader friendly I only investigated a sample of 131 in N-VESPA as well, making the findings directly comparable. When I later decided that it would be best to study NICLE and LOCNESS qualitatively as well, seeing as they are two of the main corpora in the present study, there was not enough time to study very large samples, so I decided on a size of 100 concordance lines per sample in this final study.

I tried to make sure that the samples were as random as possible. In NICLE, this was achieved by starting at the top of the list of concordance lines and analyzing only one item per NICLE writer, so that 100 writers were represented. The items for each individual writer were not sorted alphabetically, so I analyzed each writer's first usage of the investigated pronoun. In LOCNESS, the concordance lines were sorted alphabetically, so studying the first 100 lines might obscure the findings. Therefore, if a certain pronoun was used 900 times in the whole corpus, I analyzed every ninth concordance line to achieve a sample of 100. In N-VESPA, I chose to sort the concordance lines alphabetically by the fifth word to the right of the pronoun, which I found to be a quite random order. Additionally, in case this was not as random as I assumed, I analyzed a set of 50 concordance lines from four different areas on this list. The same approach was used in BAWE-ling – the only difference being that, here,

the concordance lines appeared alphabetically by the first word to the right of the item I had searched for.

The following sections will deal with the results from this qualitative investigation of the pronouns *I*, *we* and *you*. In each section, one pronoun is considered, in terms of its functions in learner and native speaker writing, either in the argumentative or in the academic genre. After presenting the relevant functions of the pronoun in question, examples of how these functions are expressed in the corpora will be listed and briefly discussed. Then, the frequencies of each function in the investigated samples are presented in tables, as are also the estimated total frequencies, as well as the estimated overuse/underuse in the learner corpus. In these tables, the frequency of irrelevant hits for each pronoun is also included – the reason being that this plays a role in the calculation of the estimated frequencies.

In each section, an analysis part will follow the results part. In the analysis, the functions are discussed in terms of how they are used in learner writing, compared to native speaker writing. The differences between the corpora are tested for statistical significance, which is referred to where relevant; however, as was also pointed out in chapter 3, it is important to remember that the significance of the differences is based on estimated frequencies, as if they were actual frequencies. Thus, the results are not as reliable as they would have been if we knew the exact raw frequencies of each function. Nevertheless, the samples are probably large enough to ensure a reasonable level of reliability, and the findings may at least serve as useful indicators of tendencies in the different corpora.

6.1. *I* in argumentative writing

6.1.1. Results

Norwegian learners and native speaker students were found to use *I* to perform four different functions in their argumentative writing, ranged below from the most frequent to the least frequent function in the learner corpus (NICLE).

- *I* as arguer
- *I* as organizer
- *I* as narrator
- Generic *I*

When the pronoun *I* is used about the writer as arguer, it is used to express the writer's opinions and typically co-occurs with verbs like *think*, *believe*, *feel* and *agree*. Norwegian learners and native speakers seem to use the arguing *I* to express similar meanings, such as evaluating a statement (ex. 1-2) and expressing their own thoughts (ex. 3-4). However, the learners' vocabulary is not as varied as the native speakers' vocabulary, so they tend to use the verb *think* very frequently, as we also saw in the quantitative study of patterns of subjective stance (section 4.2.2.). This is also illustrated in the examples.

1. **I think** this is only partly true. (NICLE)
2. Indeed, **I would agree** that the play is profoundly ambiguous (...) (LOCNESS)
3. **I think** there should be an equal amount (...) (NICLE)
4. Despite this, **I feel** that adoption should be made easier in order to (...) (LOCNESS)

When *I* refers to the writer as organizer in argumentative texts, it typically appears in sentences which give the reader information about what is going to be said in the text (ex. 5), what has been said (ex. 6) and what should be said (ex. 7).

5. In this essay **I'm going to** show you (...) (NICLE)
6. As **I have pointed out**, many criminals go free (...) (NICLE)
7. Now, maybe **I should mention** that this is a neighborhood where cars are (...) (LOCNESS)

Another function of *I* in argumentative texts is the narrative function, which is used to recount personal experiences. The writers have various reasons to bring up such experiences; for example, they want to convince the reader about the writer's authority in the current discussion (ex. 8), to compare the writer's childhood to the habits of children today (ex. 9), or to act as a witness to a situation (ex. 10).

8. **I** have worked at a mental institution for one year. (NICLE)
9. When **I** was a child we seldom watched Tv. (NICLE)
10. **I** have seen others cheat and get a better grade as a result (...) (LOCNESS)

The generic use of *I* is the least frequent of the different uses in argumentative writing. With only one hit in the NICLE sample and two in the LOCNESS sample, it is impossible to know how close to the reality the estimated frequencies for the whole corpora are. However, what we do know is that there are examples of generic use in both corpora – that is, sentences where *I* has been used to refer to people in general, instead of the more common *we*, *you* or *one*. The generic use of *I* is described by Kitagawa and Lehrer as a “safe choice” in some contexts, because it allows the writer to use himself as a role model (1990, 753). This might be the case in the following examples (11 and 12) from NICLE and LOCNESS.

11. Who is to decide what **I** should read in the newspaper and watch on the TV?
(NICLE)
12. (...) being female **I** have a right to express my opinions (...) (LOCNESS)

Table 31 shows the raw frequency of each function in the samples of 100 concordance lines from NICLE and LOCNESS. This table can only tell us about the proportion of each function within the samples, and is thus not of much value in itself. Instead, it is a useful source on which the estimated frequencies in the whole corpora are based.

Table 31: Functions of *I* in samples of 100 from NICLE and LOCNESS

Function	NICLE		LOCNESS	
	Raw frequency in sample of 100	% of sample	Raw frequency in sample of 100	% of sample
<i>I</i> as arguer	49	49	51	51
<i>I</i> as organizer	25	25	1	1
<i>I</i> as narrator	20	20	41	41
Generic <i>I</i>	1	1	2	2
Irrelevant	5	5	5	5
Total in sample	100	100	100	100

The estimated raw frequencies in NICLE and LOCNESS, which are given in table 32, are calculated by using the percentages from the above table. Since *I* appears as arguer in 49 out of 100 cases in the NICLE sample, we can estimate the total raw frequency in NICLE to be 49 % of the raw frequency of *I*. We know from the preliminary study (section 4.2.1.) that the raw frequency of *I* in NICLE is 1902; thus, the estimated raw frequency of *I* as arguer is 49 % of 1902, which is 932. Since these frequencies are only estimates, and do not necessarily represent the actual distribution of the different functions, the estimated raw frequencies have decimals.

The estimated normalized frequencies are calculated from the estimated raw frequencies and the total amount of words in the corpus, in the same way as actual normalized frequencies are calculated. Since the functions are normalized per 10,000 words, the estimated raw frequencies for each function is divided by the total word frequency in the corpus and then multiplied by 10,000. The estimated overuse values are based on the estimated frequencies for each function per 10,000 words in the same way as overuse values were calculated in the preliminary study – through dividing the normalized frequency in the learner corpus by the normalized frequency in the native speaker corpus.

Table 32: Functions of *I* – estimated frequencies in NICLE and LOCNESS and overuse in NICLE

Function	NICLE		LOCNESS		Estimated overuse in NICLE
	Estimated raw frequency	Estimated frequency per 10,000 words	Estimated raw frequency	Estimated frequency per 10,000 words	
<i>I</i> as arguer	931.98	44.30	498.78	15.38	2.88
<i>I</i> as organizer	475.50	22.60	9.78	0.30	74.97
<i>I</i> as narrator	380.40	18.08	400.98	12.36	1.46
Generic <i>I</i>	19.02	0.90	19.56	0.60	1.50
Irrelevant	95.10	4.52	48.90	1.51	
Total	1902.00	90.41	978.00	30.15	3.00

6.1.2. Analysis

What the estimated frequencies for NICLE and LOCNESS, in table 32, can tell us is that all the functions of *I* are overused in NICLE, but to varying extents. *I* as organizer is clearly the most overused function⁵⁹; since there was only one example of such usage in the LOCNESS sample, the normalized frequency of the function in the whole corpus was estimated to be only 0.30 per 10,000 words, while the NICLE frequency was estimated to be 22.60 per 10,000 words. Consequently, the estimated overuse value is 74.97. This overuse seems very high, and it is difficult to understand why Norwegian learners feel the need to guide the reader through the text as often as they apparently do, especially since these are quite short texts. At first glance, it seems like where native speakers would simply begin a discussion, learners first want to prepare their readers for the discussion. Also, where native speakers would write a concluding paragraph, learners would first sum up what has been said so far. However, a closer look at some NICLE and LOCNESS texts shows that this is not necessarily the case; native speakers also prepare their readers for what is to come, and sum up the discussion – the difference is that they do not refer to themselves while guiding their readers.

To illustrate how learners and native speakers differ when it comes to referring to content in their own texts, I have selected three final paragraphs in which the previous discussion is referred to. In example 1, a native speaker student refers specifically to the arguments which have been discussed, but does not mention himself as a writer, and in

⁵⁹ The overuse is statistically significant with 99.99 % certainty.

examples 2 and 3, Norwegian learners refer to what they as writers have done in the text. However, the learners refer to themselves in quite different ways; whereas writer 2 sums up all his deeds as a writer, writer 3 only briefly refers to something he has mentioned earlier. From a teacher's perspective, writer 3 seems more mature, and his writer reference does not have a negative effect on the text, while writer 2 seems less mature, and his repeated writer references disturb the reading of the text. This suggests that, although the overuse of *I* as organizer is quite remarkable in NICLE, the effect that this overuse has on the general impression of the texts, varies from text to text.

1. **All of these arguments are good, valid ones, but without any evidence they do not hold up well.** (...) Right now, many of us feel that suicide is horrid and completely, morally inexcusable, but eventually we must realize that suicide will never cease to exist and the best that we can do is try to understand it. (LOCNESS)
2. **I have stated** the facts, that the word anachronism and what it means. **Secondly I have written down** arguments for getting married. Finally, **I have written down** arguments against marriage. (NICLE)
3. **I mentioned** earlier that I don't think that the life-pattern of people today gives less room for dreams and imaginations. As we adapt to the society we take our dreams and fantasies with us, and this is perhaps the essential thing here: that we in our modern world, dominated by science technology and industrialisation, need our dreams and imaginations more than ever. (NICLE)

The arguing function, which is the most frequently occurring function of *I*, was also found to be significantly overused⁶⁰, with an overuse value of 2.88. According to the findings in Petch-Tyson (1998), the overuse of *I* as organizer and arguer, correlates well with the typical learner usage of *I* in Dutch, French and Swedish learner writing, as we saw in chapter 2. However, in addition to overusing *I* as organizer and arguer, the Dutch, French and Swedish learners were found to underuse *I* as narrator, which is not the case in Norwegian learner writing.

Norwegian learners in NICLE actually overuse the narrating function of *I*⁶¹. The only use of *I* which is not significantly overused is the generic *I*.

⁶⁰ 99.99 % certain.

⁶¹ 99.99 % certain.

6.2. *I* in academic writing

6.2.1. Results

Five different functions of *I* were found in the academic learner and native speaker corpora (N-VESPA and BAWE-ling). The three main functions are the academic functions of the writer in the text, namely the writer as researcher, organizer⁶² and arguer. This correlates with Fløttum's (2006) findings in the study of scientific articles in the linguistics discipline. The generic use of *I* is quite rare, but was identified in both corpora. *I* as narrator, which is when the writer talks about personal experiences not related to any research, was only found in the native speaker corpus. The following functions are ranged by their frequency in the Norwegian learner corpus (N-VESPA).

- *I* as researcher
- *I* as organizer
- *I* as arguer
- Generic *I*
- *I* as narrator

When *I* is used about the writer as researcher, it typically co-occurs with verbs in the past tense. In the academic genre, as opposed to the argumentative genre, co-occurrences of *I* with verbs in the past tense typically refer to research methods. For example, the writer might want to tell the reader about the research material (ex. 1-2), the foundations for his/her methodological choices (ex. 3) or the actual research methods (ex. 4-5).

1. Then **I read through all the examples** to see if they only included what **I** was looking for (...) (N-VESPA)
2. **I also looked at written feedback** on her Undergraduate module assignment. (BAWE-ling)
3. All approaches or theories agree that motivation matters; **a basic premise I work from** is that motivation is the key. (BAWE-ling)

⁶² The organizing function of the writer is similar to Fløttum's (2006) term "reader guide".

4. After having decided what to investigate, **I ran the search** in Word Smith and gathered the occurrences and put them into File Maker Pro to make the categorization easier. (N-VESPA)
5. For Task 2 **I interviewed** school age children to try and find out their thoughts on their own use of language and see what they thought about the use of non-standard forms in school. (BAWE-ling)

The use of *I* as organizer relates the writer to the text in a direct way. As was also the case in argumentative writing, we have the organizer telling the reader what is going to be done in the text (ex. 6-7) and what has been done (ex. 8). In addition, a role of the organizer which is perhaps more common in academic writing is to explain what the text will focus on (ex. 9), and what will not be included in the text (ex. 10).

6. In this essay **I will consider** how the word pairs ‘blank’ in English and ‘blank’ in Norwegian and the word pairs ‘lykke’ and ‘luck’ corresponds to each other in terms of meaning and distribution. (N-VESPA)
7. In this paper **I will be looking at** the role of pronunciation in my professional situation. (BAWE-ling)
8. The ‘problem’ **I mentioned** in the collocation with place was sort of legitimized (...) (N-VESPA)
9. A large amount of data was collected during the course of the case study, and while attempting to maintain the impartiality necessary for such a study, for the purposes of this assignment **I will focus on** some of the key areas emerging from the interviews and the other contacts made during the course of the study. (BAWE-ling)
10. This situation can be seen in various teaching contexts around the world, **I will not comment on** these here but limit the scope to address those contexts where there is a choice. (BAWE-ling)

When the pronoun *I* is used about the writer as arguer, it often co-occurs with the verbs *think* and *believe* (ex. 11-12), but the learners in N-VESPA show more variation than the learners in NICLE and only slightly overuse the phrases *I think* and *I believe*. Additional arguing verbs that co-occur with *I* in N-VESPA and BAWE-ling are more academic verbs like *conclude* (ex. 13), *find* (ex. 14) and *hypothesise* (ex. 15).

11. **I think** that the issue of class and gender can be significantly linked together, particularly if you look at the role of prestige. (BAWE-ling)
12. **I believe** this could also explain the non-explicit use of the reader as Senser (...) (N-VESPA)
13. **I do conclude** that in the text used in this analysis, the most frequently used process type is the material. (N-VESPA)
14. **I find** the study of false friends to be very interesting and useful. (N-VESPA)
15. Therefore, **I hypothesise** that the subject of this investigation will be achieving an appropriate Brown's stage (stage 2), as illustrated by SALT. (BAWE-ling)

As was the case in the argumentative genre, the generic use of *I* is rare – but still present – in the academic genre, both in learner writing and in native speaker writing. Examples 16 and 17 show the type of usage which was categorized as generic. In example 16, it seems like the writer wants to talk about himself/herself as any reader, and in example 17 the writer appears to be talking about herself as an example of any woman. Note that these are my interpretations of what the writers appear to express, and that in some cases it is impossible to be certain of what the writer actually means.

16. As a reader, **I** do not find myself wondering what the text is trying to tell me. (N-VESPA)
17. **I** am likely to talk to a female friend and a female employer in a different style and surely therefore other factors aside from gender must affect speech such as power as O'barr and Atkins realised. (BAWE-ling)

In argumentative writing, *I* was found to be used quite frequently about the writer as narrator, but in academic writing this is not the case. This narrative function was only found in six of the 200 concordance lines of *I* in BAWE-ling, and it was not found at all in N-VESPA. In the cases where this function appears, the intention seems to be to relate personally to the research topic of the text, as in examples 18 and 19.

18. As a native speaker Masters student, **I** have also found at times a lack of explicitness in instructions and difficulty in always knowing what was wanted by the tutors (...) (BAWE-ling)
19. (...) but perhaps **I** have been able to deal with this more effectively through such clarification strategies as questioning the tutor (...) (BAWE-ling)

Table 33 shows the distribution of the different functions of *I* in the samples from N-VESPA and BAWE-ling.

Table 33: Functions of *I* in samples of 200 from N-VESPA and BAWE-ling

Function	N-VESPA		BAWE-ling	
	Raw frequency in sample of 200	% of sample	Raw frequency in sample of 200	% of sample
<i>I</i> as researcher	100	50	73	36.50
<i>I</i> as organizer	48	24	38	19
<i>I</i> as arguer	31	15.50	56	28
Generic <i>I</i>	3	1.50	1	0.50
<i>I</i> as narrator	0	0	6	3
Irrelevant	18	9	26	13
Total in sample	200	100	200	100

Table 34 shows the estimated frequencies of the different functions of *I* in N-VESPA and BAWE-ling, based on the findings from the samples in the previous table. The estimated overuse in N-VESPA is shown in the column to the right (values above 1 indicate overuse).

Table 34: Functions of *I* – estimated frequencies in N-VESPA and BAWE-ling and overuse in N-VESPA

Function	N-VESPA		BAWE-ling		Estimated overuse in N-VESPA
	Estimated raw frequency	Estimated frequency per 10,000 words	Estimated raw frequency	Estimated frequency per 10,000 words	
<i>I</i> as researcher	834.00	36.00	233.60	13.36	2.69
<i>I</i> as organizer	400.32	17.28	121.60	6.95	2.48
<i>I</i> as arguer	258.54	11.16	179.20	10.25	1.09
Generic <i>I</i>	25.02	1.08	3.20	0.18	5.90
<i>I</i> as narrator	0.00	0.00	19.20	1.10	0
Irrelevant	150.12	6.48	83.20	4.76	
Total	1668.00	72.00	640.00	36.60	1.97

6.2.2. Analysis

As shown in table 34, some of the functions of *I* are overused in N-VESPA, while others are not. *I* as researcher was found to be overused by 2.69, which is a relatively high overuse value in N-VESPA⁶³. *I* as organizer was overused to a similar extent⁶⁴, with the overuse value 2.48, which is also quite high, but lower than the overuse value of this function in NICLE. Still, such high levels of overuse of *I* as researcher and organizer in N-VESPA are quite noteworthy, as these functions of *I* are very frequent, constituting 74 % of all the occurrences of *I* in N-VESPA.

There is also a tendency towards overuse of generic *I* in N-VESPA, but because of the low frequencies in the original samples, we may not draw any confident conclusions about this estimated overuse, although the log-likelihood test finds the overuse based on the estimated raw frequencies to be statistically significant. The narrating function of *I* also has low frequencies in the samples, so the estimated frequencies may not be completely reliable; in any case, the narrating function is not applied at all in the N-VESPA sample, so there is no overuse. However, I would not necessarily refer to the lack of *I* as narrator as learner underuse either, since this function of *I* is not expected in academic writing. Finally, the arguing function of *I*, which was found to be overused in NICLE, is not significantly overused in N-VESPA.

6.3. *We* in argumentative writing

6.3.1. Results

Of the different uses of *we* in the argumentative genre, the generic use was found to be the most frequent by far. The other functions of *we*, which either include or exclude the reader, were labeled inclusive and exclusive *we*, in accordance with Herriman (2009). However, the inclusive and exclusive functions have not been sub-categorized into specific and authorial functions, as was the case in Herriman. The functions, ranged by their frequency in Norwegian learners' argumentative writing, are listed below.

⁶³ 99.99 % certain.

⁶⁴ 99.99 % certain.

- Generic *we*
- Exclusive *we*
- Inclusive *we*

Note that the pronoun *we*, when used exclusively or inclusively, may have different functions in the text, just like the pronoun *I*. As we shall see, the exclusive *we* may function as narrator, and the inclusive *we* may function as organizer, for example. Even though it might have been interesting to sort all the different cases of exclusive and inclusive *we* into these functional categories to be able to compare the findings from this investigation to the findings from the investigation of *I*, this has not been done in the present study, due to the low frequencies of the exclusive and inclusive *we*. After all, the exclusive *we* was only used 13 times out of 100 in the argumentative genre, and the inclusive *we* was only used 3 times. If these groups were then sorted into 3 or 4 different sub-categories (depending on the genre), the findings might not be reliable. Therefore, the sub-categories of the exclusive and inclusive *we* have not been applied, but are only referred to in the description of the examples.

In some cases, it was difficult to determine which main category a certain case of *we* belonged to. Hopefully, the examples given below will help explain how the categories are defined. The rule of thumb was that, if *we* could be referring to people in general, it would be categorized as generic. Examples 1 and 2 are typical examples of the generic *we* in NICLE and LOCNESS. Sometimes *we* was considered generic even when it did not refer to people in general, but to people in a certain country. Examples 3 and 4 are examples of sentences where *we* has a generic function even though the general reference only includes members of one nation. This function of generic *we* relates to what Kitagawa and Lehrer (1990) refer to as vague *we*, as mentioned in section 2.6.3.

1. These are all main headlines **we** often hear of in the news. (NICLE)
2. Where would **we** find the time to hand wash our clothes? (LOCNESS)
3. But how could these small troops help Norway when **we** would face a much larger army? (NICLE)
4. In general **we** are already closely linked with the rest of Europe. (LOCNESS)

In the NICLE and LOCNESS samples there are no uses of the exclusive *we* to refer to a single writer, and none of the texts are written by more than one author. This means that the only exclusive use of *we* to be found here is the exclusive group, which includes the writer and some other people. The other people may be other children from the writer's childhood (ex. 5), the writer's classmates (ex. 6) or the writer's family (ex. 7). This function of *we* is very similar to the narrating function of *I*.

5. **We** played soccer, sat on the swings, played word games, riding our bikes. (NICLE)
6. In ninth grade in my school in Norway, **we** had one week to try a profession (...)
(NICLE)
7. **We** each have particular nights that **we** must prepare and cook a meal. (LOCNESS)

The least frequent function of *we* in argumentative writing is the inclusive function. The inclusive function, as it is defined in the present study, refers to the writer and the reader but not to other people. Consequently, the sentences in which the inclusive *we* appear are typically focused on the text itself, by for example talking about how to begin the text (ex. 8), or what the text has shown (ex. 9). These sentences share the functions of the organizing writer, apart from the fact that the reader is included.

8. If **we start with** punishment, why should this continue to be a part of the society's penal code? (NICLE)
9. As **we have seen** the university degree is not in a direct manner a preparation for the real world. (NICLE)

Table 35 shows the findings from the NICLE and LOCNESS samples.

Table 35: Functions of *we* in samples of 100 from NICLE and LOCNESS

Function	NICLE		LOCNESS	
	Raw frequency in sample of 100	% of sample	Raw frequency in sample of 100	% of sample
Generic	81	81	88	88
Exclusive	13	13	5	5
Inclusive	3	3	6	6
Irrelevant ⁶⁵	3	3	1	1
Total in sample	100	100	100	100

Table 36 shows the estimated frequencies of the different functions in NICLE and LOCNESS, as well as the estimated overuse in NICLE.

Table 36: Functions of *we* – estimated frequencies in NICLE and LOCNESS and overuse in NICLE

Function	NICLE		LOCNESS		Estimated overuse in NICLE
	Estimated raw frequency	Estimated frequency per 10,000 words	Estimated raw frequency	Estimated frequency per 10,000 words	
Generic	1574.64	74.85	811.36	25.01	2.99
Exclusive	252.72	12.01	46.10	1.42	8.45
Inclusive	58.32	2.77	55.32	1.71	1.63
Irrelevant	58.32	2.77	9.22	0.28	
Total	1944.00	92.41	922.00	28.42	3.25

⁶⁵ Irrelevant examples of *we* are those which were found to appear in quotes or in discussions of the pronoun *we*. (In N-VESPA, these cases should have been tagged, but in some cases they are not.)

6.3.2. Analysis

As table 36 shows, the generic function of *we* is the most common function in argumentative writing by learners as well as native speakers. In addition, generic *we* is found to be overused by learners when we compare frequencies per 10,000 words in NICLE and LOCNESS.

Learners use the generic *we* 2.99 times more frequently than native speakers in these corpora. Exclusive *we* is also overused by learners, with an overuse value of 8.45, and inclusive *we* is slightly overused, with an overuse value of 1.63, which is not very high according to NICLE standards in this investigation. Still, the overuse was found to be statistically significant in all cases⁶⁶.

Since Herriman (2009) also looked at the different functions of *we*, it might be interesting to compare her findings to the findings in the present study. However, the only comparable findings concern the distribution of the different functions percentage-wise in each corpus/sample. Even though the proportions of the different functions are compared and discussed in Herriman's study, I am not sure that comparing proportions of functions is reasonable in the present study, because I cannot see how the writers' usage of one function depends on their usage of another. Instead I would have wanted to compare different normalized frequencies of the functions, or possibly different overuse values. However, what we can learn from a comparison of Norwegian learners in the present study and Swedish learners in Herriman is that the proportion of generic *we* is higher in Swedish learner writing, whereas the proportion of inclusive and exclusive *we* is higher in Norwegian learner writing.

6.4. *We* in academic writing

6.4.1. Results

In academic learner writing we find the same uses of *we* as in argumentative learner writing, but the distribution of the uses is different. The different uses are listed below, ranged according to their frequency in academic learner writing, as estimated by the sample from N-VESPA.

⁶⁶ With 99.99 % certainty about the overuse of the generic and the exclusive *we*, and 99 % certainty about the overuse of the inclusive *we*.

- Inclusive *we*
- Generic *we*
- Exclusive *we*

The least frequent function of *we* in argumentative learner writing, namely the inclusive *we*, is the most frequent function in academic learner writing. The usage of the inclusive *we* is quite similar to what we saw in the argumentative corpora, often relating specifically to something in the text. Many of the verbs which typically follow the inclusive *we* are text-related, like *see* (ex. 1), *find* (ex. 2), *look* (ex. 3) and *have* (ex. 4). In these cases the function of the inclusive *we* is similar to the organizing function of *I*. However, there are also some cases where the inclusive *we* is used in an attempt to get the reader to agree with the writer, as in example 5, where the verb *assume* is used. Here, the inclusive *we* is more similar to *I* as arguer.

1. Backchannels can also indicate that the speaker does not want to take the floor as **we saw** at lines 12 - 15. (BAWE-ling)
2. Last but not least we also have the indicator of modality in the Norwegian construction which **we find** in example 33. (N-VESPA)
3. Furthermore, if **we look** at Larsen-Freeman, (2000), where she cites over 11 different methods in use today, we can see that there are essential similarities or a "significant...overlap". (BAWE-ling)
4. Finally, **we have** the Themes "The next president and Congress", and "The US", contributing to the information given above regarding "changes", and suggesting the article is of political nature (changes has to be made, and the American government has something to do with this.) (N-VESPA)
5. For instance, line 4, B mumbles the end of her utterance and **we must assume** that she has not yet organised her point. (BAWE-ling)

When it comes to the generic use of *we* in academic writing, most of the generic uses were found to be quite clear. Examples 6 and 7 are cases where *we* definitely refers to people in general. Example 8, on the other hand, is more difficult to categorize. The question is whether the writer is referring to a certain group of teachers, or to teachers in general, or perhaps to people in general. In this case, I assume that the writer is talking about teachers in general in the way that some of the examples of generic *we* in NICLE and LOCNESS referred to

members of a nation in general – not contrasting a group of people with other groups of people but talking about people in general inside a certain sphere, like the vague usage of *we*, referred to in Kitagawa and Lehrer (1990). Therefore, this use of *we* has been categorized as generic.

6. Whenever **we** speak or write, **we** adapt our language to the context. (N-VESPA)
7. Connotation being the cultural meaning **we** give to a sign this robot connotes technological efficiency and pleasure. (BAWE-ling)
8. Having said that, I feel perhaps **we** have unrealistic expectations of where a student should be at the beginning of a Masters course. (BAWE-ling)

In Kuo (1999), many cases of exclusive *we* referring to the writer were found in single-authored scientific articles, and Kuo suggested that this usage implies an “(...) intention to reduce personal attributions” (Kuo 1999, 125). Such uses of exclusive *we* were not found in the argumentative corpora, but they do appear in the academic corpora. However, I am not sure whether the intention behind these uses of exclusive *we* really is to reduce personal attributions; instead, it seems like the writers are trying to include the readers, almost like writers do when they use the inclusive *we*, but doing this in situations where readers cannot really be included. In some cases, it is impossible to know whether the writer meant to use exclusive or inclusive *we*, so the categorization of *we* is necessarily dependent on the researcher’s interpretation. In the present study, *we* is labeled as an exclusive reference to the writer in the cases where *we* is related to verbs which typically can be related to the writer and not the reader, such as *comment* (ex. 9), *illustrate* (ex. 10), *focus on* (ex. 11) and *investigate* (ex. 12). In examples 1 to 3, the exclusive *we* has a similar function as the organizing *I*, whereas, in example 4, the exclusive *we* has the function of the researcher.

9. The last thematic pattern on which **we will comment** is the use of the unmarked Subject Theme. (N-VESPA)
10. To **illustrate** some false friends, **we** find some examples from NICLE (...) (N-VESPA)
11. Again, much research has been done in this field and it is this research **we will focus on**. (BAWE-ling)
12. This option is not available to us in this paper; **we will investigate** the NICLE corpus only. (N-VESPA)

In addition to referring to the writer, the exclusive *we* is used to refer to a group which the reader does not belong to, as was also the case in argumentative writing. In these cases *we* can refer to the writer and his classmates (ex. 13), the writer and his students (ex. 14), and the writer and his research object (ex. 15). Considering the functions of the exclusive *we*, when used about a group, these uses of *we* are similar to the use of *I* as researcher, although the first example may be closer to the narrating function of *I*.

13. (...) both texts are more enjoyable to read than the texts **we** had in our previous obligatory assignment. (N-VESPA)
14. For my Japanese students, in order to focus on these strategies, **we** might watch a video clip of native speakers and highlight strategies such as rephrasing, using fillers and fixed phrase to bide for time. (BAWE-ling)
15. She preferred not to be tested by me as part of the interview, so instead **we** looked at some of her written work for her undergraduate module, at the marks she had achieved during her academic study in Britain, and at her IELTS score. (BAWE-ling)

The distribution of inclusive, exclusive and generic *we* in the samples from N-VESPA and BAWE-ling is shown in table 37.

Table 37: Functions of *we* in samples of 200 from N-VESPA and BAWE-ling

Function	N-VESPA		BAWE-ling	
	Raw frequency in sample of 200	% of sample	Raw frequency in sample of 200	% of sample
Inclusive	107	53.50	38	19
Generic	73	36.50	106	53
Exclusive	20	10	4	2
Irrelevant	0	0	52	26
Total in sample	200	100	200	100

Table 38 shows the estimated frequencies of the inclusive, exclusive and generic *we* in N-VESPA and BAWE-ling, as well as the estimated overuse of the different functions of *we* in N-VESPA.

Table 38: Functions of *we* – estimated frequencies in N-VESPA and BAWE-ling and overuse in N-VESPA

Function	N-VESPA		BAWE-ling		Estimated overuse in N-VESPA
	Estimated raw frequency	Estimated frequency per 10,000 words	Estimated raw frequency	Estimated frequency per 10,000 words	
Inclusive	444.59	19.19	63.46	3.63	5.29
Generic	303.32	13.09	177.02	10.12	1.29
Exclusive	83.10	3.59	6.68	0.38	9.39
Irrelevant	0.00	0.00	86.84	4.97	
Total	831.00	35.87	334.00	19.10	1.88

6.4.2. Analysis

In N-VESPA, the inclusive function of *we* is the most frequent function, and it is also quite heavily overused, with an overuse value of 5.29⁶⁷. This means that the overuse is actually higher in N-VESPA than in NICLE, in this particular case. Generic *we* is also relatively frequent in N-VESPA, and although the overuse value is not very high, the overuse was found to be statistically significant⁶⁸. However, if we look at the findings from a different perspective, the proportion of inclusive and generic *we*, in table 37, shows that whenever *we* occurs in a text, it is more likely to be generic in BAWE-ling and inclusive in N-VESPA.

Exclusive *we* was also found to be significantly overused⁶⁹, and, in addition, this function of *we* has a very high overuse value, at 9.39. The exclusive function was also highly overused in NICLE, so it seems like overuse of exclusive *we* might be a general tendency in Norwegian learner writing, although the relative frequency of exclusive *we* is quite low compared to the other uses of *we*.

⁶⁷ 99.99 % certain.

⁶⁸ 99 % certain.

⁶⁹ 99.99 % certain.

6.5. *You* in argumentative writing

6.5.1. Results

The findings from this investigation show that writers of argumentative texts, both learners and native speaker students, mainly use *you* in two ways – 1) generically and 2) directly addressing the reader. Ranged according to the frequency in the learner corpus (NICLE), the short list of functions looks like this:

- Generic *you*
- Reader-addressing *you*

In most cases, distinguishing between the generic *you* and the reader-addressing *you* was quite easy. Prototypical generic uses are exemplified below.

1. A crime is a felony, which is when **you** do something that is against the law. (NICLE)
2. Life is about rewards. It is about learning where **you** are and how **you** can move yourself to the next level. (LOCNESS)

When *you* is used to address the reader directly, the writer often appears to be trying to relate to the reader on some level. In some cases the writer makes assumptions about the reader (ex. 3) and in other cases he tries to make the reader imagine a certain situation (ex. 4). Some writers give their readers some advice or warnings (ex. 5-6), more or less explicitly talking to the reader of the text. Usually, connecting with the reader implies an effort to make the reader agree with the writer; however, in example 7, the writer makes it clear that the reader may have a different opinion, but that he, the writer, does not care.

3. If **you** ask yourself what the most important thing in life is, **you** will probably say: family. (NICLE)
4. Sit back in your chair and try to picture this: **You** have a few spare minutes (...) (NICLE)
5. If **you** are not willing to recycle ordinary batteries, buy only rechargeable. (LOCNESS)
6. I don't know how to express how important it is that everyone listen and follow the rules while riding the go-carts at an amusement park, but hopefully if **you** are reading this, the point will come across to you. (LOCNESS)
7. Whatever **you** believe, I shall continue to enjoy my roast beef and Yorkshire pudding, thankyou. (LOCNESS)

Table 39 shows the distribution of generic *you* and reader-addressing *you*, as well as the amount of irrelevant hits in the NICLE and LOCNESS samples.

Table 39: Functions of *you* in samples of 100 from NICLE and LOCNESS

Function	NICLE		LOCNESS	
	Raw frequency in sample of 100	% of sample	Raw frequency in sample of 100	% of sample
Generic <i>you</i>	75	75	82	82
Reader-addressing <i>you</i>	21	21	14	14
Irrelevant ⁷⁰	4	4	4	4
Total sample	100	100	100	100

Table 40 shows the estimated frequencies for generic *you* and reader-addressing *you* in NICLE and LOCNESS, as well as the estimated overuse in NICLE.

⁷⁰ As was also the case in the study of the pronoun *we*, the irrelevant examples of *you* are those which appear in quotes or in discussions of the pronoun *you*.

Table 40: Functions of *you* – estimated frequencies in NICLE and LOCNESS and overuse in NICLE

Function	NICLE		LOCNESS		Estimated overuse in NICLE
	Estimated raw frequency	Estimated frequency per 10,000 words	Estimated raw frequency	Estimated frequency per 10,000 words	
Generic <i>you</i>	1009.50	47.99	442.80	13.65	3.52
Reader-addressing <i>you</i>	282.66	13.44	75.60	2.33	5.77
Irrelevant	53.84	2.56	21.60	0.67	
Total	1346.00	63.98	540.00	16.65	3.84

6.5.2. Analysis

Compared to native speakers in LOCNESS, Norwegian learners in NICLE overuse both the generic *you* and the reader-addressing *you*⁷¹. The generic function of *you* is the most common function in both learner and native speaker writing, and it has an overuse value of 3.52 in learner writing. The reader-addressing function of *you*, which is not as common, has a higher overuse value, at 5.77.

6.6. *You* in academic writing

6.6.1. Results

The functions of *you* are the same in academic writing as in argumentative writing, at least in the corpora investigated here. However, in the native speaker corpus (BAWE-ling), only the generic use of *you* was found. The short list below is ranged according to the frequency of the functions in the sample from the academic learner corpus (N-VESPA).

- Generic *you*
- Reader-addressing *you*

⁷¹ 99.99 % certain.

The generic *you* and the reader-addressing *you* in the academic corpora are not very different from the same functions in the argumentative corpora, so only one example of each function to illustrate the difference will suffice. Example 1 is a typical generic use of *you*, and example 2 is a reader-addressing *you* where the writer explicitly tells the reader that he will receive the results in the conclusion.

1. Does what **you** have learned from the target language influence **your** choices (intralingual errors)? (N-VESPA)
2. In my conclusion I will give **you** my results of the favoured correspondences and what they will tell about the meaning of *burde/bør*, both in Norwegian and English. (N-VESPA)

The distribution of generic *you*, reader-addressing *you* and irrelevant uses of *you* in the N-VESPA and BAWE-ling samples is shown in table 41. Note that the frequency of irrelevant hits in BAWE-ling is extraordinarily high. If this was accounted for in the quantitative studies, the overuse value of the pronoun *you* in N-VESPA would probably have been much higher. This shows the impact that the amount of irrelevant hits may have on the findings, and is a reason for spending more time reading through the concordance lines before noting the frequency of an item. If there was more time, this would definitely have been prioritized in the present study.

Table 41: Functions of *you* in samples of 131 from N-VESPA and BAWE-ling

Function	N-VESPA		BAWE-ling	
	Raw frequency in sample of 100	% of sample	Raw frequency in sample of 100	% of sample
Generic <i>you</i>	114	87.00	45	34.40
Reader-addressing <i>you</i>	10	7.60	0	0.00
Irrelevant	7	5.30	86	65.60
Total sample	131	100	131	100

Table 42 shows the estimated frequencies of the different functions of *you* in N-VESPA and BAWE-ling, and the estimated overuse of these functions in N-VESPA.

Table 42: Functions of *you* – estimated frequencies in N-VESPA and BAWE-ling and overuse in N-VESPA

Function	N-VESPA		BAWE-ling		Estimated overuse in N-VESPA
	Estimated raw frequency	Estimated frequency per 10,000 words	Estimated raw frequency	Estimated frequency per 10,000 words	
Generic <i>you</i>	222.72	9.61	45.06	2.58	3.73
Reader-addressing <i>you</i>	19.46	0.84	0.00	0.00	#
Irrelevant	13.57	0.59	85.94	4.92	
Total	255.74	11.04	131.00	7.49	1.47

6.6.2. Analysis

As was the case in NICLE and LOCNESS, the generic function of *you* is the most frequent function in both N-VESPA and BAWE-ling, and the reader-addressing function of *you* only constitutes a small portion of the total samples. Actually, the reader-addressing *you* was not found at all in the BAWE-ling sample, so the overuse value could not be calculated, as marked by the # sign. This does not imply that the reader-addressing *you* is not overused, because the findings suggest that the learners have used the reader-addressing *you* in a context where it is not usually used by learners, which is a sort of overuse. However, the overuse is not necessarily a very important feature of learner writing, as the frequency of this function of *you* per 10,000 words in N-VESPA is only estimated to be 0.84. The overuse value of the more frequent generic *you*, however, has a higher level of reliability. According to the comparison of generic *you* in N-VESPA and BAWE-ling, Norwegian learners overuse this function of *you* with an overuse value of 3.73⁷².

⁷² 99.99 % certain.

6.7. Summing up the qualitative in-depth study

We already knew that the pronouns *I*, *we* and *you* were overused in learner writing. Now we know that the learners and the native speakers mainly use these pronouns to perform the same functions, which is good news for the learners. However, since the frequencies of the pronouns are higher in learner writing, some of the functions are necessarily overused by the learners. Through the comparisons in this chapter, we saw that most of the functions were actually overused. In fact, in NICLE, the only function which was not significantly overused was the very infrequent generic function of *I*. In N-VESPA, the learners did not overuse *I* as arguer, *I* as narrator or generic *I*. Although most of the functions were actually overused by the learners, the extent to which they were overused varied a great deal.

The learners' use of *I* to perform different functions showed that, in the argumentative genre, the learners' main problem was the heavy overuse of *I* as organizer. In the academic genre, the organizing function of *I* was also overused, but to a much lesser extent. The overuse value in N-VESPA was only 2.48, compared to the overuse value 74.97 in NICLE. Considering the fact that only a sample of 100 concordance lines were studied in NICLE, the high overuse value might not be completely reliable. Nevertheless, the findings are probably not completely incorrect either. It seems likely that NICLE writers have difficulties organizing their texts without referring to themselves. Since the writers in N-VESPA have lower overuse values for this function, it is possible that they may have benefited from having more academic experience, and from having studied the way that professional writers in their own discipline organize their texts. However, the N-VESPA students are still far from similar to the native speaker students, and they need to use fewer 1st person sg pronouns in referring to the organization of their texts, as well as in referring to their research methods, if they want to be more native-like.

When it comes to the learners' use of the functions of *we*, the learners overuse the exclusive *we* to the greatest extent, in both the argumentative and the academic genre. However, it seems like the overuse might have been caused by two different types of exclusive usage in the two different genres. In the argumentative genre, the only use of exclusive *we* was the reference to an exclusive group. Therefore, the overuse was mostly related to the narrating function of the writer, where other people were included in the stories told by the writer. In the academic genre, however, the overuse might have been related to the

fact that the writers use the authorial *we*, which does not seem to be very common in native speaker student writing at this level. With respect to this function of *we*, it would have been interesting to see whether the learners were actually closer to the professional academic norm. This might also be interesting regarding the overuse of inclusive *we*, which was more significant in academic learner writing than in argumentative learner writing. Comparing the functions of pronouns in student and professional writing would, in any case, be an interesting topic for future research.

Generic *we* and generic *you* are overused by learners in both genres, which is very interesting because of the frequency of these functions of the pronouns. In NICLE, the majority of the uses of *we* and *you* were generic, and in N-VESPA the only function of *we/you* which is more frequent than the generic function is the inclusive function of *we*. This means that the general overuse of pronouns in learner writing is related to references to people in general. What this says about the general level of W/R visibility in learner writing, is that it may not be as high as originally expected, since the generic use of *we* and *you* are the least interpersonal uses. Kitagawa and Lehrer (1990) refer to the generic usage of pronouns as being either impersonal or vague, depending on whether they refer to people in general (impersonal) or to a group of people, belonging to a nation, for example (vague). However, according to Kitagawa and Lehrer, impersonal pronouns still “mirror their more normative ‘personal’ use” (1990, 752). In this way, the generic uses of pronouns may contribute to making texts more personal, but perhaps not to the same extent as the actual personal uses of the pronouns.

Even though the generic *you* may create a personal mood, the reader-addressing *you* is definitely more personal, since it directly involves the reader. Along with the inclusive *we*, the reader-addressing *you* is probably the most interpersonal function referred to in this chapter, since it makes the writer *and* the reader visible at the same time. As pointed out by Smith (1983), any reference to the reader also makes the writer visible, since the writer must be the person making the references. Reader-addressing *you* was found to be significantly overused in argumentative learner writing. This function of *you* was also overused in academic learner writing, but it was not possible to calculate any overuse value, since there were no cases of reader-addressing *you* in the academic native speaker corpus. However, the reader-addressing function of *you* was not by any means found to be as frequent in academic learner writing as it

was in argumentative learner writing. Again it seems as if the N-VESPA students might have benefited from their academic experience, or from other variables distinguishing the two corpora.

7. Conclusion

Having access to many corpora opens up many research possibilities. In the present study, the ultimate aim of the corpus comparisons was to be able to describe how and to what extent the writer and the reader are made visible in the writing of Norwegian learners of English. In short, Norwegian learners were found to apply numerous W/R visibility features to a greater extent than native speakers, and also more frequently than many other learner groups. In the following sections, I will sum up the findings which contributed to this main conclusion, and discuss some pedagogical implications of these findings. Finally, I will suggest some topics for future research.

7.1. Summing up the findings

The preliminary study in chapter 4 showed that the general level of W/R visibility in argumentative writing was higher in Norwegian learner writing than in native speaker student writing. In argumentative learner writing, overuse was detected in the frequencies of several features of W/R visibility. Both 1st and 2nd person pronouns were found to be overused, as were the subjective stance markers *I would say*, *I think*, *I don't think*, *I remember*, *I believe* and *I know*. In addition, the disjunct *maybe* and the emphatic particles *just* and *really* were overused, along with the items *here*, *now* and *this essay* referring to the situation of writing/reading. Finally, questions and exclamations were also found to be overused in argumentative learner writing.

In the academic genre, learners were also found to overuse some features of W/R visibility, some of which were the same as the overused features in the argumentative genre. These include 1st and 2nd person pronouns, the disjunct *maybe*, and *here* as reference to the situation of writing/reading. In addition, some items were overused exclusively in the academic genre, namely the subjective stance marker *I would argue* and the disjuncts *of course* and *perhaps*. However, the general overuse of W/R visibility features was found to be more marked in argumentative writing than in academic writing.

Possible reasons for the high degree of W/R visibility in the argumentative genre may be related to the genre itself, and the tasks, which invite the writers to write about their personal opinions, as pointed out by Recksi (2004). In addition, the writers in NICLE are

generally younger than the writers in N-VESPA, and they have less academic experience, which may have affected the results from these comparisons. In chapter 5, the potential effect of academic experience was investigated, and the results showed that it was likely that the level of academic experience would affect the learners' use of features of W/R visibility.

According to Virtanen (1998), the relevance of the tasks may be of influence as well. In the academic corpora, the students know who their audience is to a greater extent than in the argumentative corpora. In addition, the students in the academic corpora are likely to be more interested in the topic of discussion since the topics are related to their own academic discipline, whereas in the argumentative corpora, the writers are not necessarily interested in the topics they are asked to discuss. Finally, as pointed out in Ädel (2008), timing and access to secondary sources may influence the learners' use of W/R visibility features. However, since timing was shown to be the most influential factor related to the use of pronouns, and since neither of the Norwegian learner groups were timed, Ädel's setting issues are probably not the main cause of the differences between the two learner groups in the present study. It seems likely that the difference between the learners in NICLE and N-VESPA are caused by a number of different factors, because so many variables distinguish the two corpora.

Moving on to the reasons for the general overuse in Norwegian learner writing, L1 transfer was investigated as a potential explanation. Comparisons of Norwegian learners and learners with other L1s showed that while learners generally overuse W/R visibility features, the overuse in Norwegian learner writing tends to be quite heavy compared to the overuse in other learner groups. Since Swedish and Finnish learners also have been shown to have high levels of overuse (Petch-Tyson 1998) it was considered likely that the overuse might be caused by transfer of norms from the L1, and perhaps cultural norms regarding the acceptance of a more personal style in formal genres. Comparisons of Norwegian L1 and English L1 sources supported the idea of L1 transfer to some extent, but the comparisons also showed that the Norwegian L1 sources might not be completely reliable, as they were quite small and not very representative.

Finally, in chapter 6, different functions of the pronouns *I*, *we* and *you* were identified. Some functions were found to be overused by learners to a greater extent than others. This can tell us more about the level of W/R visibility in learner writing. For example, the generic uses of *we* and *you* were frequent in learner writing and overused by learners compared to native

speakers, but this usage implies a lower degree of W/R visibility than the more personal uses of pronouns which refer specifically to the writer and/or the reader, such as *I* as arguer, inclusive *we* and reader-addressing *you*. However, most of the directly writer/reader-referring uses were also overused by learners, so there is no doubt that the level of W/R visibility is higher in learner writing than in native speaker writing.

7.2. Pedagogical implications

As some of the examples in chapter 6 illustrate, referring to the writer or the reader does not always make a text too informal – it depends on the ways in which the writer makes such references, and also how often the writer chooses to make such references. Even though Norwegian learners were found to have a higher level of W/R visibility than what is expected in advanced, formal English writing, this does not mean that Norwegian EFL teachers should start teaching their students not to refer to themselves at all, or that the students should necessarily begin to aim at objectivity. However, it is important that, if the accepted level of W/R visibility seems to be higher in Norwegian argumentative and academic writing at university level than in English writing in the same contexts, the students should be made aware of this difference.

Hyland (2002a) suggests that the underuse of personal reference in Hong Kong learner writing might have been caused by differences in writing instruction at school. If Hong Kong students are instructed to be objective and to be invisible as writers, and therefore underuse personal reference in their English writing, the overuse in Norwegian learner writing might be due to opposite instruction norms in Norway. It may well be that Norwegian students are encouraged by their teachers to relate personally to the topics they discuss and to participate actively in the text, or at least that teachers do not punish students who choose to be personally involved in their writing. Since cultural and academic norms may be different from country to country, it is crucial that language learners are taught the differences between the norms in their own language and the norms in their L2. In that way, they can at least choose whether or not to conform to the L2 norms. It might be that, in some cases, such as when a text is to be published in non-English speaking countries which are similar to Norway, it would be better to follow the Norwegian norms of interpersonality, whereas, in cases where

the main reader group would be native speakers of English, it would be better to conform to the L2 norms.

EFL teaching may also benefit from teaching alternative ways of referring to the writer. Perhaps learners rely too heavily on the words that they know well, such as pronouns, which are learnt at an early stage, and which can be applied in many settings. Being confident about the use of passive constructions and *-ing* participle clauses, for example, may help learners decrease their frequency of pronouns. In general, EFL teaching might benefit from focusing on constructions which are non-existent or uncommon in the learners' L1.

The qualitative in-depth study of pronouns showed that the organizing function of *I* in argumentative writing was the most overused function in learner writing. Thus, it might be useful, in EFL teaching, to focus on alternative ways to structure a text. As seen in the native speaker corpus, native speakers also refer to content in their own essays, whether something is to be done or something has been mentioned previously, but they do not necessarily refer to themselves while doing so. Furthermore, learners may benefit from learning to express their opinions more subtly, and in more varied ways, for example by presenting and evaluating arguments without necessarily using subjective stance markers like *I think* and *I believe* all the time. Additionally, EFL teaching might focus on ways to tell a story and explain research procedures without showing too much personal involvement, possibly by using the passive form, as suggested earlier.

In terms of the overuse of the generic pronouns *we* and *you*, it is difficult to determine what causes the learner overuse of these pronouns. If the case is that the learners generalize too often, this might imply lack of interest in, or knowledge about, the topic of discussion. However, the case may also be that learners and native speakers generalize in different ways, depending on their vocabulary. Whereas native speakers may refer to people in general in various ways (see examples 1-4), learners may not have the same precise vocabulary and may, therefore, more often rely on generic pronouns. This is, of course, only a hypothesis, and more a suggestion for further research than anything. However, since studies have shown that learners underuse academic vocabulary (Granger and Paquot 2009, Paquot 2010), there is reason to believe that a more specific focus on academic vocabulary in EFL teaching is needed. At the same time, it is probably important to teach register awareness, and to make

sure students know some of the most important differences between formal and informal language, and between speech and writing.

1. Now, **people** know that stereotypes don't hold up. (LOCNESS)
2. In the United States the right of the **citizens** to keep and bear arms is guaranteed by the Constitution, but has been variously interpreted, especially for the wrong reasons. (LOCNESS)
3. **The general public** may need access by foot (environmental schemes) or usually by vehicle. (BAWE)
4. He may then be asserting that **human beings** have no choice other than to perceive reality through the information they gain through their senses. (BAWE).

In my opinion, as a student of the English language and a future EFL teacher, findings from corpus studies may indeed be useful to EFL teaching. Some studies argue that corpora should be applied directly in the classroom (James 1981, Granath 2009), which may, in some cases, be both possible and beneficial, although it would be quite time-consuming for the teacher, I expect. However, corpus-based findings may be applied in other ways, such as in the making of EFL teaching materials and in the education of EFL teachers. As the present study has tried to show, corpus studies can add to our knowledge about learners' skills and learners' needs, and point out specific learner problems. In this way, I hope that the findings from the present study may contribute to EFL teaching, if only as a small part of the existing research on W/R visibility in learner writing.

7.3. Looking ahead

When searching for answers to my original research questions, many new questions and interesting possibilities appeared, such as the possible differences between the generic pronouns used in Norwegian and English (mentioned in section 5.1.2.). I would also have liked to study the functions of the writer and the reader in more detail, for example by describing the different ways in which *I* is used as arguer in larger samples so as to get a better picture of the variety of phrases which are used in Norwegian learner writing and native speaker student writing. Qualitative studies of the other W/R visibility features are also needed, since such studies might reveal how, and to what extent, the different features make the writer and/or the reader visible.

Further studies may also benefit from using additional corpora. For example, it may be interesting to include a corpus of American student writing, in addition to British student writing. Such studies may be able to find out whether Norwegian learners are more influenced by American or British English. I would also recommend using corpora of native speaker professional writing in addition to native speaker student writing. I originally intended to do so, in the present study, but there was no time. Also, I did not have a corpus of professional writing in the linguistics discipline, which would have been preferable in comparisons with N-VESPA and BAWE-ling. Comparisons of students and professional writers would be useful in establishing to what extent native speaker students are valid control groups for learners, for example.

It would also be interesting to see further studies on the L1 transfer effect. As pointed out in section 5.1., new corpora are needed in order to investigate the differences between the Norwegian language and the English language – preferably large corpora of Norwegian L1 writing in the argumentative and academic genres, and from different disciplines. Furthermore, corpora of student writing in Norwegian are needed, to be able to compare the students' writing skills in their L1 and their L2. If possible, it would be most interesting to have corpora of school children's writing as well, to cover the different stages all the way from primary school to university education. In the present study (section 5.2.), investigations showed developmental tendencies from bachelor levels to master levels at university. With more wide-ranging corpora, such developmental tendencies could be explored more extensively, not only with regard to W/R visibility, but in a number of ways which might be useful for the field of EFL teaching.

However, awaiting the collection of such corpora, studies may still point to tendencies which could be investigated more in detail in the future. This is, in any case, the objective of the present study. Hopefully, the findings may even be of interest at the moment, as indications of Norwegian EFL learners' skills and needs.

Bibliography

- Altenberg, B. & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In S. Granger, *Learner English on Computer* (pp. 80-93). London: Longman.
- Arnaudet, M., & Barrett, M. (1984). *Approaches to Academic Reading and Writing*. Englewood Cliffs, NJ: Prentice Hall.
- Biber, D., Conrad, S., & Leech, G. (2002). *Longman Student Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.
- Coffin, C., & Mayor, B. (2004). Texturing Writer and Reader References in Novice Academic Writing. In D. Banks, & D. Banks, *Text and Texture. Systemic Functional Viewpoints on the Nature and Structure of Texts* (pp. 239-264). L'Harmattan.
- Coventry University. (2011, April 10). *British Academic Written English Corpus (BAWE)*. Retrieved October 4, 2011, from <http://wwwm.coventry.ac.uk/researchnet/BAWE/Pages/BAWE.aspx>
- Fløttum, K. (2006, January 5). En språklig analyse av norsk medisinsk prosa. *Tidsskrift for den norske legeförening*, pp. 65-68.
- Gilquin, G. (2008). Combining contrastive and interlanguage analysis to apprehend transfer: detection, explanation, evaluation. In G. Gilquin, S. Papp, & M. B. Díez-Bedmar, *Linking up contrastive and learner corpus research* (pp. 3-33). Amsterdam: Rodopi.
- Granath, S. (2009). Who Benefits From Learning How to Use Corpora? In K. Aijmer, *Corpora and Language Teaching* (pp. 47-66). Amsterdam: John Benjamins Publishing Company.
- Granger, S. (1998). The computer learner corpus: a versatile new source of data for SLA research. In S. Granger, *Learner English on Computer* (pp. 3-18). London: Longman.
- Granger, S., & Paquot, M. (2009). In search of a general academic vocabulary: A corpus-driven study. *Options and Practices of L.S.A.P practitioners*. Heraklion: University of Crete.

- Granger, S. & Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger, *Learner English on Computer* (pp. 119-131). London: Longman.
- Harwood, N. (2005). 'Nowhere has anyone attempted ... In this article I aim to do just that': A corpus-based study of self-promotional I and we in academic writing across four disciplines. *Journal of Pragmatics* 37, pp. 1207-1231.
- Hasselgård, H. (2009a). Temporal and Spatial Structuring in Norwegian Student Essays. In R. Bowen, M. Mobärg, & S. Ohlander, *Corpora and Discourse - and Stuff. Papers in Honour of Karin Aijmer* (pp. 93-104). Gothenburg: Acta Universitatis Gothoburgensis.
- Hasselgård, H. (2009b). Thematic choice and expressions of stance in English argumentative texts by Norwegian learners. In K. Aijmer, *Corpora and Language Teaching* (pp. 121-139). Amsterdam: John Benjamins Publishing Company.
- Hellekjær, G. O. (2009, October). Academic English reading proficiency at the university level: A Norwegian case study. *Reading in a foreign language* 21(2), pp. 198-222.
- Hellekjær, G. O. (2005). *The acid test : does upper secondary EFL instruction effectively prepare Norwegian students for the reading of English textbooks at colleges and universities?* Oslo: University of Oslo, Faculty of Arts, The Department of Teacher Education and School Development.
- Herriman, J. (2009). The use of *we* in argumentative essays by Swedish advanced learners of English. In R. Bowen, M. Mobärg, & S. Ohlander, *Corpora and Discourse - and Stuff. Papers in Honour of Karin Aijmer* (pp. 105-120). Gothenburg: University of Gothenburg.
- Holmes, J. (1999). *An Introduction to Sociolinguistics*. New York: Longman.
- Hyland, K. (2002a, February 5). Authority and invisibility: authorial identity in academic writing. *Journal of Pragmatics* 34 , pp. 1091-1112.
- Hyland, K. (2002b, October). Options of identity in academic writing. *ELT Journal* , pp. 351-358.

- Hyland, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies* , 173-192.
- James, C. (1981). *Contrastive Analysis. Applied linguistics and language study*. Harlow: Longman.
- Kitagawa, C., & Lehrer, A. (1990). Impersonal uses of personal pronouns. *Journal of Pragmatics* 14 , 739-759.
- Kuo, C.-H. (1999). The Use of Personal Pronouns: Role Relationships in Scientific Journal Articles. *English for Specific Purposes, Vol. 18, No. 2* , 121-138.
- Leech, G. (1998). Preface. In S. Granger, *Learner English on Computer*. London: Longman.
- Lorenz, G. (1999). Learning to cohere: causal links in native vs. non-native argumentative writing. In B. Bublitz, U. Lenk, & B. Ventola, *Coherence in Spoken and Written Discourse. How to Create it and How to Describe it* (pp. 55-75). Amsterdam: John Benjamins.
- McCrostie, J. (2008). Writer visibility in EFL learner academic writing: A corpus-based study. *ICAME Journal* , pp. 97-114.
- Meunier, F. (2000). A computer corpus linguistics approach to interlanguage grammar: noun phrase complexity in advanced learner writing. Unpublished PhD thesis (referred to in Paquot 2010). Université catholique de Louvain: Louvain-la-Neuve.
- Meyer, C. F. (2002). *English Corpus Linguistics. An Introduction*. New York: Cambridge University Press.
- Paquot, M. (2010). *Academic Vocabulary in Learner Writing*. London: Continuum.
- Paquot, M., & de Cock, S. (2011, August 23). *The Varieties of English for Specific Purposes dAtabase (VESPA) learner corpus*. Retrieved September 29, 2011, from Centre for English Corpus Linguistics: <http://www.uclouvain.be/en-cecl-vespa.html>
- Paquot, M., Hasselgård, H., & Ebeling, S. O. (forthcoming). Writer/reader visibility in learner writing across genres: A comparison of the French and Norwegian components of the ICLE and VESPA learner corpora.

- Petch-Tyson, S. (1998). Writer / reader visibility in EFL written discourse. In S. Granger, *Learner English on Computer* (pp. 107-118). London: Longman.
- Rayson, P. (n.d.). *Log-likelihood calculator*. Retrieved November 1, 2011, from <http://ucrel.lancs.ac.uk/llwizard.html>
- Rayson, P., & Garside, R. (2000, October). *Comparing Corpora using Frequency Profiling*. Retrieved November 1, 2011, from the Log-likelihood calculator website: http://www.comp.lancs.ac.uk/~paul/publications/rg_acl2000.pdf
- Recski, L. J. (2004, August). Expressing Standpoints in EFL Written Discourse. *Revista Virtual de Estudos da Linguagem - ReVEL*. V. 2, n. 3 .
- Smith, E. L. (1983, March). *Achieving Impact through the Interpersonal Component*. Retrieved September 11, 2011, from Education Resources Information Center: http://eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED240547&ERICExtSearch_SearchType_0=no&accno=ED240547
- Université catholique de Louvain, Centre for English Corpus Linguistics. (2010, April 5). *Centre for English Corpus Linguistics: LOCNESS*. Retrieved September 26, 2011, from <http://www.uclouvain.be/en-cecl-locness.html>
- Virtanen, T. (1998). Direct questions in argumentative student writing. In S. Granger, *Learner English on Computer* (pp. 94-106). London: Longman.
- Wales, K. (2011). *A Dictionary of Stylistics*. Harlow: Pearson Education Limited.
- Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. Amsterdam: John Benjamins.
- Ädel, A. (2008). Involvement features in writing: do time and interaction trump register awareness? In G. Gilquin, S. Papp, & M. B. Díez-Bedmar, *Linking up contrastive and learner corpus research* (pp. 35-53). Amsterdam: Rodopi.

Corpora used

ICLE (NICLE)

The ICLE corpus was investigated by using ICLE 2 on CD-ROM.

LOCNESS

The LOCNESS corpus was accessed through the “tekstlab” at the University of Oslo at <http://www.tekstlab.uio.no/cgi-bin/omc/LOCNESSsearch.cgi>

VESPA (N-VESPA)

The VESPA corpus is not yet published. The texts were studied in Wordsmith Tools.

BAWE (BAWE-ling)

The BAWE corpus⁷³ was accessed online at Coventry University’s website: <http://wwwm.coventry.ac.uk/researchnet/BAWE/Pages/SketchEngine.aspx>.

⁷³ The data in this study come from the British Academic Written English (BAWE) corpus, which was developed at the Universities of Warwick, Reading and Oxford Brookes under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics [previously called CELTE], Warwick), Paul Thompson (Department of Applied Linguistics, Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the ESRC (RES-000-23-0800)