

UNIVERSITY OF OSLO  
DEPARTMENT OF LITERATURE, AREA STUDIES AND EUROPEAN LANGUAGES

Who am *I*? and Who are *we*?  
in argumentative and academic writing

---

A corpus-based study of author identity in the writing of native  
speakers and Russian learners of English

Ekaterina Zolotova  
May 2014



Supervisor: Professor Hilde Hasselgård

## ABSTRACT

Many EFL learners find it challenging to construct their author identity in English L2 argumentative and academic writing. Learners may present brilliant content by means of inappropriate rhetorical features and fail to communicate their message. This study focuses on the construction of author identity in English L2 writing by Russian learners. The quantitative study investigates the overuse, underuse and misuse of first person pronouns in English L1 and L2 writing. Preliminary results suggest the overuse of first person pronouns and the dominance of the 'we' perspective, since the overuse rate is the highest here. The qualitative analysis supports this finding. The pronoun *we* is overused in high-risk functions, whereas the pronoun *I* is overused in low-risk functions. To account for the overuse, the study researches the possible reasons that could cause it, including the effect of L1 transfer, influence of speech, developmental and teaching-induced factors. All the factors, except for the influence of speech, are found to have an impact on learner writing. The final chapter presents a comparative analysis of L2 writing by learners with Slavic backgrounds, which reveals a common tendency to overuse the 'we' perspective. The study contributes to the series of investigations of various learner varieties on the material of the ICLE and VESPA corpora.

## **ACKNOWLEDGEMENTS**

First of all, I wish to express my sincere gratitude to Professor Hilde Hasselgård for her expert, valuable and sincere guidance in the process of writing this master thesis. I appreciate that Hilde always promptly answered my questions and was willing to meet me, as a good piece of advice, that comes in time is twice as helpful. I also thank Hilde for carefully reading my work and giving constructive comments on it.

I want to express my genuine gratitude to Corpus Linguistics Group and all the people who work with compiling corpora. I will also take this opportunity to record my wholehearted thanks to the contributors of my collection of academic writing, who shared their bachelor and master assignments with me. It made it possible for me to conduct a more profound qualitative study.

I am grateful to my husband for him trying to grasp all the subtleties of my topic, while working hard on his own master thesis in computer science. It was great to share our worries and difficulties, but also breakthroughs and exciting results. My father provided me with great moral support and practical advice throughout the study, which was of immense importance to me.

## TABLE OF CONTENTS

ABSTRACT .....	1
ACKNOWLEDGEMENTS .....	2
TABLE OF CONTENTS .....	3
LIST OF TABLES .....	8
LIST OF FIGURES .....	10
LIST OF MENTIONED CORPORA .....	11
LIST OF ABBREVIATIONS .....	13
1. INTRODUCTION .....	14
1.1 Aim and scope .....	14
1.2 Research questions.....	15
1.3 Hypotheses.....	17
1.4 Thesis outline .....	18
1.5 Terminology.....	18
2. THEORETICAL BACKGROUND OF THE STUDY .....	20
2.1 Notion of nativeness .....	20
2.2 Involvement features in learner writing.....	22
2.3 Author identity in argumentative writing .....	24
2.3.1 Author identity in NS argumentative writing.....	25
2.3.2 Author identity in NNS argumentative writing.....	27
2.3.3 Challenges in argumentative writing.....	30
2.3.4 Concluding remarks on author identity in argumentative writing.....	31
2.4 Author identity in academic writing .....	33
2.4.1 Author identity in NS academic writing.....	35
2.4.2 Author identity in NNS academic writing.....	37
2.4.3 Author identity within the scope of academic disciplines.....	40
2.4.4 Author identity on various academic levels .....	41

2.4.5	Concluding remarks on author identity in academic writing .....	43
2.5	Anglo-American and Slavic rhetoric .....	44
2.6	Reasons for overuse, underuse and misuse of first person pronouns .....	46
2.6.1	Influence of speech.....	46
2.6.2	L1 transfer .....	47
2.6.3	Teaching-induced factors .....	47
2.6.4	Developmental factors.....	48
2.6.5	Extralinguistic factors.....	48
2.7	Functions of the first person pronouns <i>I</i> and <i>we</i> .....	49
2.7.1	Functions of <i>I</i> in argumentative writing .....	49
2.7.2	Functions of <i>we</i> in argumentative writing .....	50
2.7.3	Functions of <i>I</i> in academic writing.....	50
2.7.4	Functions of <i>we</i> in academic writing.....	51
2.8	Research objective .....	52
3.	METHOD AND MATERIAL.....	54
3.1	Methodological background .....	54
3.2	Material.....	56
3.2.1	NS corpus of English argumentative writing .....	57
3.2.2	NNS corpus of English argumentative writing.....	57
3.2.3	NS corpus of English academic writing .....	59
3.2.4	NNS corpus of English academic writing .....	60
3.2.5	NS corpus of Russian argumentative writing.....	61
3.2.6	NS corpus of Russian academic writing.....	63
3.2.7	Other corpora.....	64
3.3	Concluding remarks on method and material .....	67
4.	RESULTS AND ANALYSIS: QUANTITATIVE STUDY OF AUTHOR IDENTITY IN ARGUMENTATIVE AND ACADEMIC WRITING .....	69
4.1	Author identity in argumentative writing .....	71

4.1.1	The first person singular pronoun <i>I</i> .....	72
4.1.2	The first person singular pronoun <i>ME</i> .....	74
4.1.3	The first person singular pronoun <i>MY</i> .....	77
4.1.4	The first person singular pronoun <i>MINE</i> .....	78
4.1.5	The first person plural pronoun <i>WE</i> .....	79
4.1.6	The first person plural pronoun <i>US</i> .....	81
4.1.7	The first person plural pronoun <i>OUR</i> .....	83
4.1.8	The first person plural pronoun <i>OURS</i> .....	84
4.1.9	Concluding remarks on author identity in argumentative writing .....	86
4.2	Author identity in academic writing .....	88
4.2.1	The first person singular pronoun <i>I</i> .....	88
4.2.2	The first person singular pronoun <i>ME</i> .....	91
4.2.3	The first person singular pronoun <i>MY</i> .....	91
4.2.4	The first person singular pronoun <i>MINE</i> .....	93
4.2.5	The first person plural pronoun <i>WE</i> .....	93
4.2.6	The first person plural pronoun <i>US</i> .....	94
4.2.7	The first person plural pronoun <i>OUR</i> .....	95
4.2.8	The first person plural pronoun <i>OURS</i> .....	95
4.2.9	Concluding remarks on author identity in academic writing .....	96
5.	RESULTS AND ANALYSIS: QUALITITAIVE STUDY OF AUTHOR IDENTITY IN ARGUMENTATIVE AND ACADEMIC WRITING .....	98
5.1	The pronoun <i>I</i> in argumentative writing .....	99
5.1.1	Results .....	99
5.1.2	Analysis .....	102
5.2	The pronoun <i>WE</i> in argumentative writing .....	104
5.2.1	Results .....	104
5.2.2	Analysis .....	106
5.3	The pronoun <i>I</i> in academic writing .....	107

5.3.1	Results .....	107
5.3.2	Analysis .....	110
5.4	The pronoun <i>WE</i> in academic writing .....	110
5.4.1	Results .....	110
5.4.2	Analysis .....	113
5.5	Concluding remarks on the qualitative analysis of author identity.....	114
6.	RESULTS AND ANALYSIS: REASONS FOR OVERUSE AND UNDERUSE OF AUTHORIAL PRONOUNS IN ENGLISH L2 WRITING .....	116
6.1	The effect of L1 transfer .....	116
6.1.1	The effect of L1 transfer in argumentative writing .....	117
6.1.2	The effect of L1 transfer in academic writing .....	121
6.2	Developmental factors .....	123
6.3	Teaching-induced factors.....	125
6.4	The influence of speech .....	127
6.5	Concluding remarks on the reasons for overuse and underuse of authorial pronouns in English L2 writing .....	128
7.	RESULTS AND ANALYSIS: AUTHOR IDENTITY IN L2 WRITING OF LEARNERS WITH SLAVIC L1 BACKGROUND.....	130
7.1	Argumentative writing of Slavic learners .....	130
7.2	Academic writing of Slavic learners.....	132
7.1	Concluding remarks on author identity in Slavic learner writing.....	134
8.	CONCLUSION .....	135
8.1	Summary of the findings.....	135
8.2	Pedagogical implications .....	138
8.3	Looking ahead.....	141
	BIBLIOGRAPHY .....	143
	CORPORA USED.....	150
	APPENDIX 1. The frequencies of the 2- to 4-gram clusters with the pronoun <i>I</i> in LOCNESS and ICLE-RU .....	151

APPENDIX 2. The frequencies of the 2- to 4-gram clusters with the pronoun <i>I</i> in BAWE-LING and VESPA-RU.....	152
APPENDIX 3. The collocations of the pronoun <i>me</i> in BAWE-LING and VESPA-RU (collocation window span: 0 right – 1 left).....	153



## LIST OF TABLES

Table 1. Features of involvement in learner writing (examples from ICLE-RU) .....	22
Table 2. Personal pronouns in French and Norwegian learner corpora compared to NS corpus according to Table 2 in Paquot, Hasselgård and Ebeling (2013) .....	28
Table 3. Discourse functions of self-mention in sample of 10 research articles .....	36
Table 4. Discourse functions of self-mention in NS vs. Hong Kong student writing .....	38
Table 5. ‘ <i>I/we</i> ’ occurrences in the corpora of research articles in linguistics. ....	45
Table 6. A concise overview of the key characteristic factors of the main corpora.....	68
Table 7. The use of the pronoun <i>I</i> in LOCNESS and ICLE-RU .....	72
Table 8. Recurrent word-combinations in LOCNESS and ICLE-RU .....	72
Table 9. The distribution of <i>I think, I feel, I believe</i> and <i>I would like to/I’d like to</i> in LOCNESS and ICLE-RU.....	73
Table 10. Patterns of subjective stance in ICLE-RU and LOCNESS .....	74
Table 11. The pronoun <i>I</i> in LOCNESS and ICLE-RU .....	74
Table 12. The top fifteen 3- and 4-gram clusters with the pronoun <i>me</i> in LOCNESS.....	75
Table 13. The top fifteen 3- and 4-gram clusters with the pronoun <i>me</i> in ICLE-RU.....	75
Table 14. The pronoun <i>my</i> in LOCNESS and ICLE-RU .....	77
Table 15. The pronoun <i>mine</i> in LOCNESS and ICLE-RU .....	78
Table 16. The pronoun <i>we</i> in LOCNESS and ICLE-RU.....	79
Table 17. The use of tenses in the clusters with the pronoun <i>we</i> + verb <i>see</i> .....	80
Table 18. The use of tenses in the clusters with the pronoun <i>we</i> + verb <i>say</i> .....	80
Table 19. The pronoun <i>us</i> in LOCNESS and ICLE-RU.....	81
Table 20. 2- and 3-gram clusters with the pronoun <i>us</i> in LOCNESS and ICLE-RU .....	81
Table 21. First person plural imperatives in ICLE-RU and CRUAW .....	82
Table 22. The verb <i>give</i> + the pronoun <i>us</i> in ICLE-RU and CRUAW .....	83
Table 23. The pronoun <i>our</i> in LOCNESS and ICLE-RU.....	83
Table 24. The 2-gram cluster <i>in our and в наш (in our)</i> in LOCNESS, ICLE-RU and CRUAW.....	84
Table 25. The pronoun <i>ours</i> in LOCNESS and ICLE-RU .....	85
Table 26. The first person pronouns in LOCNESS and ICLE-RU .....	86
Table 27. The pronoun <i>I</i> in BAWE-LING and VESPA-RU .....	89
Table 28. The top ten recurrent word-combinations in LOCNESS and ICLE-RU .....	89
Table 29. The pronoun <i>me</i> in BAWE-LING and VESPA-RU .....	91
Table 30. The pronoun <i>my</i> in BAWE-LING and VESPA-RU.....	92

Table 31. The collocations of the pronoun <i>me</i> in BAWE-LING and VESPA-RU .....	92
Table 32. The pronoun <i>we</i> in BAWE-LING and VESPA-RU .....	93
Table 33. The pronoun <i>we</i> + the verb <i>find</i> in BAWE-LING and VESPA-RU .....	94
Table 34. The pronoun <i>us</i> in BAWE-LING and VESPA-RU .....	94
Table 35. The pronoun <i>our</i> in BAWE-LING and VESPA-RU.....	95
Table 36. The first person pronouns in BAWE-LING and VESPA-RU.....	96
Table 37. The functions of the pronouns <i>I</i> and я (=I) in the sample of up to 300 instances from L1 English, L2 English and L1 Russian argumentative writing.....	99
Table 38. The functions of the pronouns <i>I</i> and я (=I) in L1 English, L2 English and L1 Russian argumentative writing (estimated frequencies) .....	100
Table 39. The functions of the pronouns <i>we</i> and мы (=we) in the sample of up to 300 instances from L1 English, L2 English and L1 Russian argumentative writing.....	104
Table 40. The functions of the pronouns <i>we</i> and мы (=we) in L1 English, L2 English and L1 Russian argumentative writing (estimated frequencies).....	104
Table 41. The functions of the pronouns <i>I</i> in the sample of up to 300 instances from L1 and L2 English academic writing. ....	107
Table 42. The functions of the pronouns <i>I</i> in L1 and L2 English academic writing (estimated frequencies) .....	107
Table 43. The functions of the pronouns <i>we</i> and мы (=we) in the sample of up to 300 instances from L1 English, L2 English and L1 Russian academic writing.....	111
Table 44. The functions of the pronouns <i>we</i> and мы (=we) in L1 English, L2 English and L1 Russian academic writing (estimated frequencies) .....	111
Table 45. Generic/inclusive/exclusive <i>we</i> and мы (=we) in the sample of up to 300 instances from L1 English, L2 English and L1 Russian academic writing.....	114
Table 46. Generic/inclusive/exclusive <i>we</i> and мы (=we) in L1 English, L2 English and L1 Russian academic writing (estimated frequencies) .....	114
Table 47. 2- to 4-gram clusters with the pronoun <i>I</i> in LOCNESS .....	151
Table 48. 2- to 4-gram clusters with the pronoun <i>I</i> in ICLE-RU .....	151
Table 49. The top ten 2- to 4-gram clusters with the pronoun <i>I</i> in BAWE-LING. ....	152
Table 50. The top ten 2- to 4-gram clusters with the pronoun <i>I</i> in VESPA-RU .....	152
Table 51. The collocations of the pronoun <i>me</i> in BAWE-LING.....	153
Table 52. The collocations of the pronoun <i>me</i> in VESPA-RU.....	153

## LIST OF FIGURES

Figure 1. Interpersonal involvement in learner and NS writing.....	23
Figure 2. Features of W/R visibility in NS argumentative writing. ....	25
Figure 3. Personal pronouns in NNS argumentative writing.....	27
Figure 4. Average frequency of first person pronouns per research paper.....	40
Figure 5. Relative frequency per million of <i>I think</i> in academic writing, novice native student writing, learner writing and speech .....	42
Figure 6. The use of first person pronouns across the corpora.....	70
Figure 7. The use of the first person pronouns in NS argumentative and academic writing. ....	70
Figure 8. The use of the first person pronouns in NNS argumentative and academic writing.....	71
Figure 9. The sequence <i>as for me</i> across corpora per 10,000 words.....	76
Figure 10. The transfer of L1 verbs with a semantic meaning 'give' into L2.....	83
Figure 11. The construction ADJ + N + <i>of/like/as OURS</i> across the corpora .....	86
Figure 12. L1 transfer of frequency, function, register and form into the interlanguage.....	87
Figure 13. Historical perspective on the use of the pronoun <i>we</i> in Russian L1 writing in RNC.....	117
Figure 14. The 'I' perspective in CRUAW, ICLE-RU and LOCNESS .....	118
Figure 15. The 'we' perspective in CRUAW, ICLE-RU and LOCNESS.....	119
Figure 16. The 'I' and the 'we' perspective from 1994 until 1998 in ICLE-RU .....	119
Figure 17. The 'we' perspective in L1 Russian, L1 and L2 English argumentative writing .....	120
Figure 18. The 'I' perspective in L1 Russian, L1 and L2 English argumentative writing.....	121
Figure 19. The 'I' perspective in L1 Russian, L1 and L2 English academic writing.....	122
Figure 20. The 'we' perspective in L1 Russian, L1 and L2 English academic writing .....	122
Figure 21. The 'I' perspective in the master and bachelor L1 Russian, L2 and L1 English academic writing .....	124
Figure 22. The 'we' perspective in the master and bachelor L1 Russian, L2 and L1 English academic writing .....	124
Figure 23. The use of the 'I' perspective in NS and Slavic learner argumentative writing .....	131
Figure 24. The use of the 'we' perspective in NS and Slavic learner argumentative writing.....	131
Figure 25. The use of the 'I' perspective in NS and Slavic learner academic writing .....	133
Figure 26. The use of the 'we' perspective in NS and Slavic learner academic writing.....	133

## **LIST OF MENTIONED CORPORA**

BAWE – British Academic Written English corpus

BAWE-LING – British Academic Written English corpus – linguistic component

BNC – the British National Corpus

CRUAW – Collection of RUssian Argumentative Writing

CRUMBA – Collection of RUssian Master and Bachelor Assignments

CELAW – Collection of English L2 Argumentative Writing

CODIF – Corpus de Dissertations Françaises

ICLE - the International Corpus of Learner English

ICLE-BL– the Bulgarian component of the International Corpus of Learner English

ICLE-CZ– the Czech component of the International Corpus of Learner English

ICLE-DU – the Dutch component of the International Corpus of Learner English

ICLE-FR– the French component of the International Corpus of Learner English

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

ICLE-PL– the Polish component of the International Corpus of Learner English

ICLE-RU– the Russian component of the International Corpus of Learner English

ICLE-SP– the Spanish component of the International Corpus of Learner English

ICLE-SW – the Swedish component of the International Corpus of Learner English

LOCNESS – the Louvain Corpus of Native Essays

London-Lund Corpus

RNC – Russian National Corpus

USE – Uppsala Student Essay

VESPA – the Varieties of English for Specific Purposes database

VESPA-CZ – the Czech component of the Varieties of English for Specific Purposes database

VESPA-NO – the Norwegian component of the Varieties of English for Specific Purposes database

VESPA-PL – the Polish component of the Varieties of English for Specific Purposes database

VESPA-RU – the Russian component of the Varieties of English for Specific Purposes database

## **LIST OF ABBREVIATIONS**

AWL – Academic Word List

CA – Contrastive Analysis

CAE – Certificate in advanced English

CIA – Contrastive Interlanguage Analysis

EFL – English as a foreign language

ELF – English as a Lingua Franca

ESL – English as a second language

ICM – Integrated Contrastive Model

IELTS – International English Language Testing System

NNS – non-native speaker

NS – native speaker

TOEFL – Test of English as a Foreign Language

W/R visibility – writer/reader visibility

## 1. INTRODUCTION

Argumentative as well as academic writing carries an ideational message and expresses an authorial stance towards it. To fulfil these tasks successfully, writers should arrange their discourse in conformity with socially defined rules. Each culture has its own understanding of how author identity should be portrayed in the text. In Anglo-American rhetoric, authors are expected to show commitment to their arguments and opinions. The absence of an explicit author in the text can be interpreted by native speakers as a reluctance to take an independent stance or a desire to downplay their role in the study (Hyland 2002a). Probably the most direct expression of authorial self is the use of first person pronouns. For a learner writer, however, new writing conventions can present a significant challenge, as they can be very different from the practices in their L1. Learners' ignorance about such rhetorical differences may present them in a negative light compared to their native peers.

Previous studies have shown that learners overuse writer/reader (W/R) visibility features including first person pronouns (*cf. e.g.* Petch-Tyson 1998; Paquot, Hasselgård and Ebeling, 2013). The overuse of involvement features has been registered in the writing of various learner groups, and seems to be characteristic of novice writers in general. However, these studies have mainly comprised learners with Western European backgrounds. Hyland's study (2002a) of Hong Kong student writing reports underuse of authorial reference, especially when students express commitment to an argument. The present study focuses on identifying such problem areas for Russian learners of English. In the study by Čmejrková (1996), Slavic rhetoric is discussed in terms of a reader-responsible language (Hinds 1987), where the author supplies the reader with information to think about. Russian learners may find it challenging to adopt Anglo-American individualistic rhetoric, which implies an overt presence of the pronoun *I*. However, overuse, underuse or misuse do not always result from the difference in rhetoric. Some findings may indicate that Russian learners are similar to their counterparts with other L1 backgrounds, and thus confirm that some problems are general for most learner varieties.

### 1.1 Aim and scope

The main aim of this study is to contribute to the existing knowledge about W/R visibility, and explore in detail the notion of author identity in English L2 writing. Previous studies have analysed this feature in the writing of learners with various L1 backgrounds. This study focuses on how Russian learners use first person pronouns to construct author identity in English L2 writing. The definition of pronoun groups varies from study to study. The personal pronouns *I*,

*me, we* and *us* are typically included in this group. This study also considers the possessive pronouns (*mine* and *ours*) and possessive determiners (*my* and *our*) as a part of first person pronouns. In this study, I explore the use of first person pronouns as they present the most overt way of expressing author identity in the text. On the one hand, author identity in the form of first person pronouns is a rather narrow subject; on the other hand, it allows for an in-depth qualitative study of the most overt method to emphasize authorial presence in the text. Qualitative studies are able to reveal the functional differences between rhetoric identities and give possible reasons for overuse, underuse or misuse.

The main research question of this study can be posed in the following way: **How do Russian learners construct their author identity in English L2 writing?** The next section presents a more detailed overview of the research questions I aim to tackle in this master thesis.

## 1.2 Research questions

In order to answer the main research question of this master thesis, I distinguish four sub-questions. All of them are interconnected and investigate author identity in English L2 writing by Russian learners. Some of these four questions have supplementary sub-questions, which specify the direction of the investigation.

### 1. How visible is the author in the writing of Russian learners of English compared to native speaker writing?

- At what rate do Russian learners overuse or underuse first person pronouns in English L2 writing compared to native speakers?
- Do Russian learners misuse first person pronouns?
- What can potentially provoke the overuse, underuse or misuse of first person pronouns in learner writing?

My preliminary study in Chapter 4 aims at answering these research questions. The two main Sections 4.1 and 4.2 deal with argumentative and academic writing respectively. Each of these sections has in its turn eight sub-sections dedicated to one of the first person pronouns: *I, me, my, mine* and *we, us, our, ours*. To answer the question of how visible the author is in the writing of Russian learners of English, I investigate every pronoun in the relevant corpus, register the results and calculate frequencies per 10,000 words. The same study is conducted to identify the visibility of the author in the native speaker (NS) writing. The relative frequencies are used to calculate the overuse or underuse of a pronoun in learner writing against NS writing. The



function *Clusters* in the *WordSmith Tool 5.0.0.334* provides material for making assumptions about the reasons for the overuse/underuse and for identifying misuses. Further investigations are often required to test the assumptions. They can involve a study of L1 material or a comparison of n-gram frequencies in learner and NS writing.

## **2. Are there any functional discrepancies between Russian learners and native speakers as regards their authorial identity?**

- What functions do first person pronouns fulfil in the L2 writing and in NS writing?
- Do Russian learners overuse, underuse or use the pronouns in the same way as native speakers?

Chapter 5 of the master thesis deals with these questions. It presents a qualitative analysis of first person pronouns, which is based on the quantitative data. First, I identify the functions of the pronouns in the learner and NS writing, then analyse the extracted sample according to the designed classification and, finally, judge about the overuse/underuse. In some sections, I also elaborate on the reasons for the deviation in learner writing.

## **3. What may be the reasons for the learners' overuse, underuse and misuse in English L2 writing?**

- Does learners' L1 affect the way they construct their author identity in the English L2 writing?
- To what extent may the overuse, underuse and misuse be due to developmental factors and academic experience?
- To what extent may the deviation be caused by the teaching-induced factors?
- Can the writing by Russian learners experience the influence of speech?

This research question is addressed in Chapter 6, which investigates four factors that could lead to the overuse, underuse and misuse of first person pronouns in learner writing. These factors were identified by previous studies that attempted to find an explanation of the overuse of involvement features in learner writing. This study examines the importance of the factors in the case of Russian learners. Sections 6.1.1 and 6.1.2 discuss the effect of L1 transfer in argumentative and academic writing. Section 6.1.3 evaluates the significance of the academic

experience for a more native-like construction of author identity. Teaching-induced factors and the influence of speech are analysed in Sections 6.1.4 and 6.1.5 respectively.

#### **4. To what extent may the learners' Slavic L1 background determine author identity in the English L2 writing?**

Chapter 7 presents a small-scale study of English L2 writing collected from students with different Slavic L1 backgrounds. The study aims to find out if the similarities in the rhetoric of the Slavic languages find its reflection in the learners' L2 writing. The small scale of the study is due to a very limited amount of available material.

### **1.3 Hypotheses**

Based on previous research, my main hypothesis is that Russian learners overuse first person pronouns compared to NS writers. Argumentative essays are expected to show a higher level of involvement than academic assignments. To my knowledge, there are no studies of W/R features in the English L2 writing of Slavic learners. It is thus difficult to hypothesize about the rate of the overuse.

Considering the reasons for the overuse, my main hypothesis relies on the results of Paquot's study (2008). It illustrates that learners' interlanguage often reflects their L1. The effect of L1 transfer can be observed in form, function, register and frequency of L2 units. My hypothesis is therefore that Russian learners overuse the pronoun *we* to a higher degree than other learner groups. The reason is that Russian, similarly to other Slavic languages, is a reader-responsible language, where the author supplies the reader with data and is not likely to take an independent stance. Especially in academic genre, the use of the pronoun *I* is considered very direct and spoken-like. The pronoun *we*, on the other hand, is more subtle and more appropriate for academic discourse. Hyland (2002a) discusses the influence of similar L1 academic conventions in his study of English L2 writing of learners with a Hong Kong L1 background, and concludes that they are likely to avoid the pronoun *I* and use a less assertive *we* instead.

Another hypothesis derives from Gilquin and Paquot's study (2008), which illustrates that learners are drilled to use a number of multi-word units. The use of the units is often promoted by teaching materials. Some multi-word units include first person pronouns (*as for me*, *I think*, *to my mind* etc.) and can lead to the unconscious overuse of the pronouns.

Learners with a Slavic L1 background will probably have similar rates of overuse/underuse in English L2 writing. The writing conventions of Slavic languages share a number of common features, which differ significantly from what is a norm in Anglo-American rhetoric. Due to the effect of L1 transfer, English L2 writing will probably reflect Slavic rhetoric.

#### **1.4 Thesis outline**

The present master thesis consists of eight chapters. The first chapter is introductory and aims to provide a brief overview of the research questions, hypothesis, terminology, aims and scope of the study. Chapter 2 builds a theoretical background of the study and introduces the findings of the previous research. This helps to put my own investigation in the context of what has already been done. The main topics covered by the chapter include the W/R visibility in learner and NS writing, the influence of academic experience on the way students construct their author identity, the reasons for the overuse, underuse and misuse of personal pronouns by learners and the functions of personal pronouns in argumentative and academic discourse. The next chapter gives a detailed description of the methods and material used to conduct my own investigation. The main material is the corpora of learner and NS writing, which is discussed in terms of its reliability and comparability. Chapters 4 to 6 focus on presenting and interpreting the results of my research. Preliminary results are considered in Chapter 4, and an in-depth qualitative study is presented in Chapter 5. Chapter 6 analyses the possible reasons for the overuse, underuse and misuse that are registered in the preliminary study. In Chapter 7, I aim to compare English L2 writing of learners with various Slavic backgrounds and find out if there are any similarities between them. The final Chapter 8 summarizes the findings of my research, discusses pedagogical implications and proposes the directions for the further research.

#### **1.5 Terminology**

Through investigating the use of first person pronouns, I aim to find out how learners and native speakers construct their author identity in argumentative and academic writing. In different parts of the master thesis, various terms are adopted to refer to the subject of the research: writer/reader visibility, the 'I' and the 'we' perspectives, and author identity.

The term *writer/reader (W/R) visibility* places emphasis on the quantitative analysis of personal pronouns in learner and NS writing. It focuses primarily on how visible or invisible the writer is in discourse. The notion of overuse and underuse is central for this connotative

meaning. The term is widely used in quantitative studies of learner writing (Petch-Tyson 1998; Paquot, Hasselgård and Ebeling 2013).

The term *author identity* emphasizes the qualitative component and investigates how personal pronouns function in the writing of learners compared to native speakers. The notion of identity is central for this connotative meaning. The term is used in Hyland's study (2002a), which analyses the functions of first person pronouns in learner and NS writing.

The '*I*' and the '*we*' *perspectives* are terms I adopt from Vassileva's study (1998). The '*I*' perspective implies the use of the pronouns *I*, *me*, *my* or *mine*, whereas the '*we*' perspective builds up around the pronouns *we*, *us*, *our* or *ours*. These terms highlight the difference between the individualistic and the collective mind setting of the author. Analysing author identity from this angle reveals interesting results, which are discussed in Sections 6.1.1 and 6.1.2 of this master thesis.

The term author identity is central for this master thesis, as it puts emphasis on the qualitative side of the study, but, at the same time, comprehends the preliminary quantitative study and the investigation on the reasons for the overuse, underuse and misuse. Therefore, the term author identity is applied in the title and analysis of the master thesis.

## **2. THEORETICAL BACKGROUND OF THE STUDY**

This chapter presents a theoretical overview of studies that deal with W/R visibility and author identity in learner writing. The first section focuses on explaining why NS writing can be considered a norm to compare L2 writing to and why learners tend to aspire to native speakers. The next few sections introduce the research on authorial presence in argumentative and academic writing. The findings from these studies are later compared to the results of my own research. Author identity in academic discourse is discussed in terms of the influence of such factors as discipline and academic experience. Since my study focuses on how Russian learners construct their author identity in English L2 writing, I find it important to dedicate Section 1.8 to the differences between the Russian and Anglo-American rhetoric. The last two sections deal with the reasons for the overuse or underuse of first person pronouns and the functions of *I* and *we* distinguished by different researches. The findings of these studies build the grounds for my in-depth qualitative study of author identity.

### **2.1 Notion of nativeness**

The main aim of this section is to account for the use of the NS corpora as a control tool for identifying deviation in Russian learner writing. This study relies on the assumption that nativeness is a key feature of a text used to identify the norm of the (standard) English language use. The notion of nativeness is described by Pawley and Syder (1983) as the ability of the native speaker to convey a meaning by a phrase that is not only grammatical, but also natural and idiomatic. In order for a text to be considered native-like, it should also construct author identity according to certain rhetoric rules. However, learners may produce texts that are non-nativelike or even highly marked. Although such texts can be fully grammatical, they would be a ‘foreignism’ to a native speaker. Thus, a native-like selection is only a limited number of syntactic and lexical bundles and features used with a definite frequency in certain contexts. This assumption is supported by numerous corpus-based investigations (Gilquin and Paquot 2007; Paquot, Hasselgård and Ebeling 2013; Aijmer 2001; Mauranen 1996; Hyland 2002c). In the mentioned studies, NS corpora are used as a control instrument to estimate the overuse, underuse or misuse in the learner writing.

This approach to English as a foreign language (EFL) and English as a second language (ESL) is challenged by advocates of English as a Lingua Franca (ELF). They believe that English is shaped equally much by non-native speakers as it is by native speakers (Seidlhofer 2005). The features that do not lead to misunderstanding can be dropped in the language learning

process, and teachers can focus on communication strategies that are crucial in ELF settings. These minor nuances, as seen from the point of view of an ELF advocate, are excessive, and their mastering requires substantial language experience (*ibid*). However, numerous investigations (Granger 1998; Aijmer 2002; Rundell and Granger 2007; Hasselgård 2009; Neff *et al.*, 2007) provide the evidence that due to limited lexical resources, learners may produce unnatural, repetitive and somewhat incoherent texts. The authors of these texts may be accused of faulty thinking, which would not be due to the wrong concept they portray, but rather due to language- and culture-specific discourse strategies they apply (Mauranen 1996).

Furthermore, authors who aim at performing successfully at an international level are to comply with the dominant Anglo-American rhetoric, as the Anglo-American community is leading the publishing market in the world and defines the criteria of what is scholarly appropriate for publishing (*ibid.*). Following this rhetoric implies producing naturally flowing and logical texts. One of the key features of Anglo-American rhetoric is studied in this master thesis and concerns the way writers construct author identity in the text. Failing to comply with the rhetoric may lead to misunderstanding from the reader's side. If, as suggested by the advocates of the ELF, NS model is not to be aimed at, it would hardly be possible to obtain common discourse grounds to logically communicate the message between various rhetoric practices.

This argument, however, should not discourage learners from being active language users, and their knowledge of English should not be considered as deficient or not 'flawless' enough to be native-like (Zaytseva 2011), but rather initiate them to perfect their foreign-language competence. Nativeness is hard to achieve for those with little exposure to English, who learn the language through textbooks, as the majority of the NS clauses and sentences are memorized, and only a minority are entirely novel creations (Pawley and Syder 1983). In this sense, research on nativeness provides learners with valuable insight into how they can further develop their language skills. NS corpora, in its turn, constitute a control group against which researchers can compare the results from learner corpora. The present thesis is a moderate contribution to the field of EFL, which adopts the notion of nativeness to identify the deviation in the use of first person pronouns in L2 writing of Russian learners. A more profound overview of what has already been done in this field and what features characterize learner writing in general is presented in the following sections.

## 2.2 Involvement features in learner writing

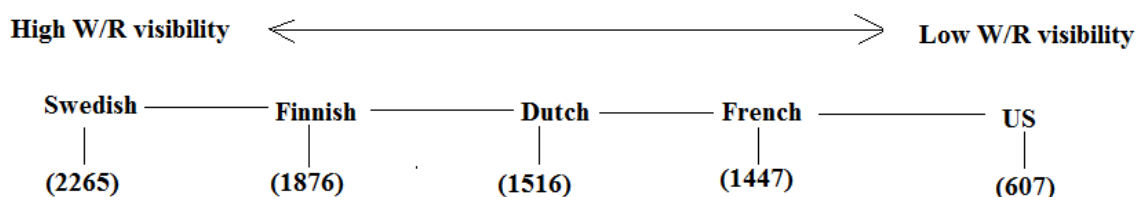
Authorial presence in the text can be signalled through a range of features. The most explicit one, *i.e.* first person pronouns, is investigated in detail in this master thesis. A number of studies have explored the use of first person pronouns and other involvement features in learner and NS writing. As research results show (Petch-Tyson 1998; Aijmer 2001; Gilquin and Paquot 2007; Paquot, Hasselgård and Ebeling 2013), learner writing is typically characterized by a high level of involvement if compared to corresponding NS literate tradition. The following involvement features are often overused by learners: (*cf.* Table 1)

**Table 1.** Features of involvement in learner writing (examples from ICLE-RU)

first person singular pronouns	<i>I, me, my, mine</i>
first person plural pronouns	<i>we, us, our, ours</i>
second person pronouns	<i>you, your, yours</i>
Epistemic adverbs	<i>certainly, maybe, of course, probably, perhaps</i>
Discourse markers	<i>well, you see, you know</i>
Emphatic particles	<i>just, really</i>
Questions	<i>e.g. What missions should the prison system fulfil? (ICLE-RU)</i>
Exclamations	<i>e.g. Then come and follow me! (ICLE-RU)</i>
Reference to situation of writing/reading	<i>here, now, this essay</i>

First person singular signals explicitly the presence of the author in a text, whereas second person pronouns introduce the reader into narration. First person plural unites the author and the reader or other people using pronouns *we, us, our, ours*. Previous investigations have illustrated the overuse of the pronouns in learner writing. Petch-Tyson (1998) concludes that learners are more prominent in writing than native speakers, which could be due to cultural difference in persuasive strategies, teaching instructions or the nature of the task. The figure below shows that Scandinavian writers are more overtly present in texts than their Dutch and French counterparts (*cf.* Figure 1).

**Figure 1.** Interpersonal involvement in learner and NS writing according to Figure 8.3 in Petch-Tyson (1998)



Individual difference has not been studied by Petch-Tyson (*ibid.*). They are later investigated in detail in works by, for example, Granger and Rayson (1998), who studied profiling in French learner texts and concluded that French learners overuse first and second personal pronouns and underuse prepositions. Prepositions are typical for nominal tendency in literate tradition. Paquot, Hasselgård and Ebeling (2013) and Fossan (2011) analyse data from Norwegian learners and find out that they demonstrate a higher level of involvement than other learner varieties. The reason for this is likely to be the L1 transfer along with general language problems. L1 Norwegian writing is characterized by a high level of involvement, which together with frequent exposure to spoken language (above all through TV), results in the overt presence of the author. Aijmer (2001) introduces data from Swedish learners, who overuse *I think* in order to make their argument sound more persuasive. The overuse is explained in terms of the influence of discussion and debates as well as the influence of the Swedish rhetorical tradition, which supports the results from the two previous investigations. Similar results for *I think* are registered in Spanish (Neff *et al.* 2007). To my knowledge, the use of involvement features have not yet been studied in Slavic learner writing, which makes it difficult to assume how visible Russian learners are in English L2 writing.

An involved style in writing is witnessed in English language learning. The findings are interpreted by Altenberg (1997) as poor register awareness. However, the intercorpus comparison by Ädel (2008) reveals that an overtly involved style is primarily caused by task settings (time available) and intertextuality (access to secondary sources). The work investigates the level of involvement in the time limited and unlimited sections of ICLE-SW (The Swedish part of the *International Corpus of Learner English*) and *Uppsala Student Essay* (USE) corpus of untimed essays written by Swedish students, who had access to secondary sources. The results reveal that less involvement features are registered in untimed essays and even less, when students have access to other sources.



Altenberg's (1997) study raises awareness about the overuse of involvement features by learners and aims at developing pedagogical instructions, which can help students to achieve stylistically correct and target-like texts. Ädel (2008), in her turn, stresses the importance of the right settings in addition to better instructions and more experience with register. Timed writing tasks with no secondary sources to consult can hardly be compared to professional or NS writing. Therefore, the overuse of involvement features by learners is a complex issue, which has to be addressed multilaterally.

### **2.3 Author identity in argumentative writing**

A part of the investigation presented in this thesis is dedicated to the comparison of author identity in argumentative NS and NNS writing. A short overview of works on argumentative learner writing is given in this chapter to create a background and basis for the comparison of my own results in the practical part of this thesis. Argumentative writing has not been equally much explored as academic writing, and fewer works report on how learners can construct author identity and support their stand using appropriate reasoning techniques without overusing personal pronouns. This problem, however, is one of the key issues for the advanced learners of English who take international certification tests as IELTS (*International English Language Testing System*), TOEFL (*Test of English as a Foreign Language*), CAE (*Certificate in advanced English*) and others. Such tests are not an aim on its own; however, they attempt to reflect the needs of an advanced learner in a real world, where good argumentative skills are held in high esteem and are a must in the academic world.

Author identity in argumentative writing has previously been researched in learners whose L1 backgrounds are Swedish (Aijmer 2001), Dutch, Finnish, French and Swedish (Petch-Tyson 1998), Norwegian and French (Paquot, Hasselgård and Ebeling 2013) and Norwegian (Fossan 2011). This thesis has as its aim to introduce new data to this research field by looking at a different variety of learner argumentative writing – learners with Slavic L1 background. The frequencies in learner writing are compared with results from a compatible NS corpus and data from the analysed L1 texts. Following the example of the above-mentioned studies, this work relies on material extracted from the relevant parts of ICLE (Russian, Polish, Czech and Bulgarian) for the student learner writing and from *The Louvain Corpus of Native English Essays* (LOCNESS) for the NS writing. Therefore, the results from the Russian and Polish parts of ICLE can be compared to the corresponding results from other parts of the corpus and

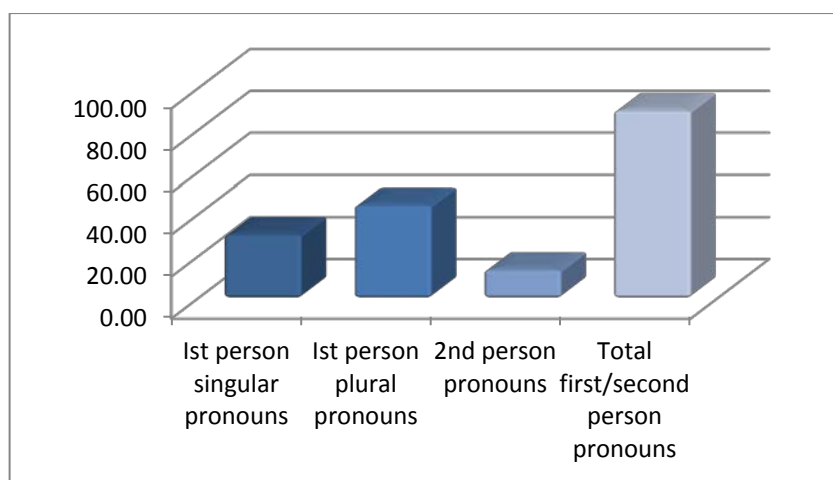
contribute to achieving a more complete picture over learner argumentative writing as a whole, and Slavic learner writing in particular.

### 2.3.1 Author identity in NS argumentative writing

This section presents an overview of the previous research on features used by native speakers to express author identity in argumentative writing. Describing these features is a key factor for analysing learner writing, as terms ‘overuse’, ‘underuse’ or ‘misuse’ are dependent on the norms in the NS community.

Author identity in NS argumentative writing is studied by Petch-Tyson (1998), who uses it to research W/R visibility in four learner varieties. The results for native English-speaking students are based on the data from the US part of LOCNESS.<sup>1</sup> The analysis includes features that are most suitable for automatic retrieval. Figure 2 shows the results of the investigation for native speakers.

**Figure 2.** Features of W/R visibility in NS argumentative writing according to Table 8.1 in Petch-Tyson (1998). Frequency per 10,000 words.



As one could expect, pronouns are the most explicit realisation of author identity in a text. It seems interesting that half of the pronouns are first person **plural** (*we, us, our, ours*): 45 out of 90 instances of pronouns per 10,000 words. First person **singular** (*I, me, my, mine*) comes second: 31 instances per 10,000 words. Second person pronouns (*you, your, yours*) are least often: 14 instances per 10,000 words. Therefore, an interesting feature of NS argumentative writing is that it introduces the reader and other people into the discourse. This way, the writer

<sup>1</sup> More information on LOCNESS is available in Section 3.2.1 of this master thesis or at the official webpage of the *Université Catholique de Louvain* <https://www.uclouvain.be/en-cecl-locness.html>

appeals to the reader and tries to convert the opponent to his or her side. It functions as an extra instrument of argumentation. Concordances data give a more profound insight into how native speakers use personal pronouns to construct author identity in the text. Petch-Tyson (*ibid.*) points out that in the US corpus practically half of occurrences of *I* are with verbs in the past tense, and many of them form chains, which describe personal experiences.

*I think* in argumentative writing is analysed by Aijmer (2001). She reports that *I think* is typical in conversations to signal uncertainty, which is supported by the data from the London-Lund Corpus, where it occurs 51 times per 10,000 words (Aijmer 1997). Simon-Vandenberg (2000) reports even higher frequencies in political interviews and television debates. The work is also interesting as it points out that *I think* implements various functions. For example in political discourse, it has a function of ‘making a point’ or ‘adopting a stance’ and expresses the speaker’s authority.

John (2012) claims that writers can express their author identity through attentions to linguistic features other than personal pronouns. The research is first of all based on academic discourse, but the results can also be valid for argumentative writing. John focuses her investigation on citation. Writer’s presence in the text can be expressed through the evaluation of the cited references. For example, the verbs *find* and *suggest* are reporting verbs, and author’s evaluation of information is not obvious. The verb *point out* signals agreement between the author and the cited source. It closes down the possibility of disagreeing with the source. The verb *claim*, on the other hand, is a verb of arguing, and it implies the difference in the views of the author and the source. These linguistic features can be successfully adopted into argumentative writing for expressing an opinion and supporting or refuting an argument.

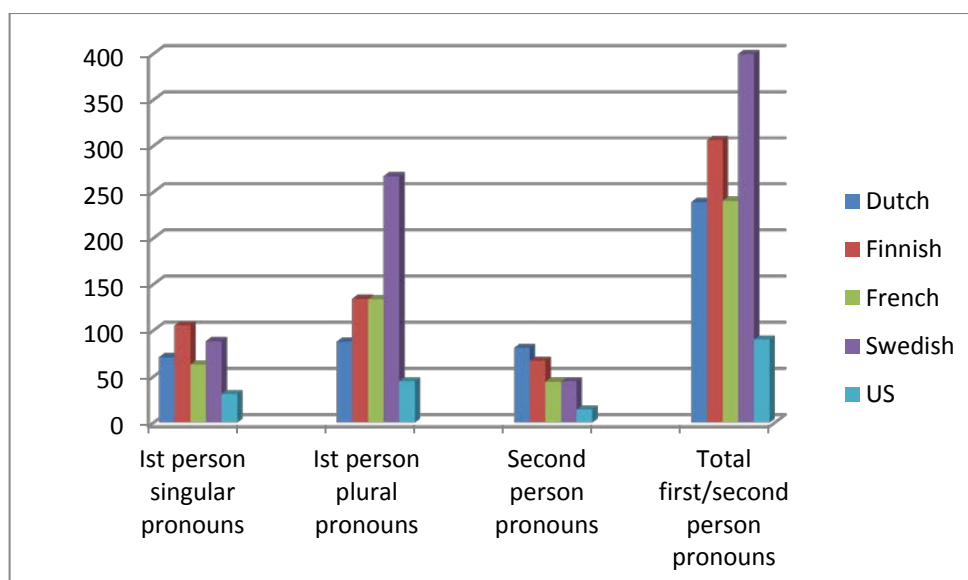
This section focused on the previous research on author identity in native English-speaking argumentative writing. All the results with the exception of the research by John (2012) are extracted from the LOCNESS corpus, which is the most complete available source of comparable argumentative essays. The features described above will be treated as a presumed norm for the comparison of other learner varieties. I would like to stress that the norm is not static and depends on many criteria such as level of education (bachelor, master or professional), discipline (humanities, social or technical sciences), type of writing (argumentative vs. academic) and others. The LOCNESS corpus contains numerous essays from NS writers and is a reliable source of results that can be compared to results from ICLE, which contains learner essays.

### 2.3.2 Author identity in NNS argumentative writing

This section reviews previous research on author identity in learner argumentative writing. The findings discussed in this section are used to evaluate the overuse rate registered in Russian learner writing (*cf.* Chapter 4). The majority of the studies on learner argumentative writing is carried out within the framework of the ICLE corpus.<sup>2</sup> ICLE consists of argumentative essays from different learner groups, which are considered comparable to the texts from LOCNESS.

Learner argumentative writing is typically characterized by a higher level of involvement in comparison to NS writing. The work by Petch-Tyson (1998) includes data for English language learners from four different language and cultural backgrounds: Dutch, Finnish, French and Swedish. The overuse of involvement features is registered in all of them (*cf.* Figure 3).

**Figure 3.** Personal pronouns in NNS argumentative writing according to Table 8.1 in Petch-Tyson (1998). Frequency per 10,000 words.



According to Figure 3, first person plural (*we, we’x, us, our* and *ours*) is the most frequently used personal pronoun both among learner varieties and in NS writing. Scandinavian learner writing is characterized by a higher level of involvement than Dutch and French, especially in case of the first person singular and plural.

To gain a more profound insight, one needs to look at the concordances. The concordances of the pronoun *I* in NNS writing show that learners use it for organizing purposes rather than for realising author identity in the text. They tend to apply it for ‘saying something about the writer functioning within the text or what the writer thinks’ (Petch-Tyson 1998: 114).

<sup>2</sup> More information on ICLE is available in Section 3.2.2 or at the official webpage of the *Université Catholique de Louvain* <http://www.uclouvain.be/en-cecl-icle.html>

The only exception is Finnish learner group, where *I*, like in NS writing, describes personal experience or feelings. Another striking difference of the NNS writing is the chains of involvement features in the beginning of the sentence:

*To conclude, I think I have made it clear that we ...* (Petch-Tyson 1998: 115)

*I mean that according to my opinion there...* (*ibid.*)

*As I am an optimist, I do not think that things...* (*ibid.*)

Such frequent occurrence of the pronoun *I* puts much focus on this feature. In addition, it is often stressed by either topicalisation or emphatic particles. Native-speaker concordances usually have one pre-‘topic’ feature. All in all, Petch-Tyson (*ibid.*) concludes that learners are more overtly present in writing than native speakers and their presence is visible through various features.

Aijmer (2001, 2002) draws similar conclusions when investigating the use of *I think* and the use of modal elements in the Swedish component of ICLE. The major conclusion in both cases is that learners overuse modal elements and the phrase *I think* in order ‘to influence the reader’s beliefs and attitudes’ (Aijmer 2002: 65). Instead of applying native-like argumentation strategies, learners resort to involvement features, including *I think* and modal elements, to make their argument sound more persuasive. Just as pointed out by Petch-Tyson (1998), *I think* is overused by learners in the sentence-initial position and often in chain expressions as *I think we should*. Granger (1998) proves the same findings to be true for French-speaking learners: 72 instances of *I think* per 50,000 words in learner corpus against three examples in NS texts. French learners are also shown to overuse first and second person pronouns in argumentative writing in the study by Paquot, Hasselgård and Ebeling (2013). Besides, the research supports the conclusion by Petch-Tyson (1998) that Scandinavian learner writing is characterized by a more involved style (*cf.* Table 2):

**Table 2.** Personal pronouns in French and Norwegian learner corpora compared to NS corpus according to Table 2 in Paquot, Hasselgård and Ebeling (2013). Frequency per 100,000 words.

	ICLE-FR	ICLE-NO	LOCNESS
1 <sup>st</sup> person	1898	2550	791
2 <sup>nd</sup> person	344	797	227

Ringbom (1998 as cited in Paquot, Hasselgård and Ebeling 2013: 5) points out that the high frequency of the first person pronoun *I* is partly caused by the expression *I think* which is especially frequent among learners in argumentative writing.

An in-depth study of subjective stance markers was conducted by Hasselgård (2009), who points out that *I* often co-occurs with verbs of mental and verbal processes such as *think*, *believe*, *say*, *tell* and others. Data for the Norwegian learner variety suggests that expressions of belief and probability (*I think*, *I believe*, *I would say* and *I don't think*) are strongly overused by learners.

Another study that focuses on the Norwegian learner variety is the master thesis by Fossan (2011). She investigates in detail the use of involvement features in Norwegian learner writing. The qualitative in-depth study shows that the overuse of the pronoun *I* is caused by L1 transfer. However, this work also explains that *I* is heavily overused as organizer, and the overuse of the pronoun *we* is often related to the narrating function of the author, where s/he wants to include other people who were a part of a story told. The majority of uses of the pronouns *we* and *you* in ICLE-NO are found to be generic, *i.e.* referring to people in general. This implies that the level of involvement in learner writing is not as high as it was expected, as the generic use is the least interpersonal use of the pronouns.

As already mentioned in the previous section, John (2012) points out that learners can realize their author identity by using reporting verbs to express their opinion on a citation. The study reveals that learners tend to misuse reporting verbs: they use them based on how often the verb has been used before, rather than author's intended attitude towards the citation. Here lies the potential for a better pedagogical explanation of how one can construct one's learner identity and how one can reduce the overuse of personal pronouns.

Ädel (2008) provides an important clarification about the reasons for the overuse of involvement features. She examines if the access to secondary sources and exam conditions influence the level of involvement. It was previously suggested by Altenberg (1997) that explicit presence of NNS writers in discourse resulted from poor register awareness. Ädel (2008) argues that this is not the only factor influencing the performance of students. She points out that students need to be given the right setting in the form of access to other texts for input and no timing. This may lead to other challenging issues as authenticity and plagiarism. A perfect solution that would account for both of the problems has yet to be found.

### 2.3.3 Challenges in argumentative writing

Argumentative writing has never been considered easy, even when writers use their mother tongue. Based on the previous research, I would like to summarize the problems of native speakers when writing argumentative essays. This could shed some light on the problems that learners may experience when writing essays, as they also learn language gradually, though in a different environment.

McCann (1989) suggests that the main problem for schoolchildren is that they are good at stating claims and propositions, but weak at using warrants and recognizing the possibility of an opposing opinion and responding to it. A similar problem can be recognized in learner writing, as they construct their author identity in texts by overstressing their personal opinion. As pointed out by Aijmer (2001), learners overuse *I think* to make their statement sound more persuasive. A more appropriate approach would be to use warrants to support one's opinion. A similarity between learners and children here could be that they both have a limited language experience and vocabulary. For schoolchildren, it is L1 language experience and vocabulary, whereas for learners – L2. Therefore, an involved style in learner writing could be caused not only by, for example, poor register awareness, but also by the limitations of their vocabulary and language experience. The involved style of learner writing can as well be influenced by extralinguistic factors such as the interest of the writer in the topic of discussion.

A different factor that can influence the quality of argumentative writing is how much attention is paid to spelling and grammar – so called *formal revisions*. Crowhurst (1983) points out that eleventh grade students had more experience and felt less need to make formal revisions. This can partly be relevant to learner writing. If essays are written during exams, students' focus may be concentrated on producing a grammatically and orthographically correct piece of work. Students may therefore pay less attention to the 'nativeness' of their text. Under the pressure of being graded, learners can 'play safe' and choose secure techniques in front of revealing their true possibilities and risking making a mistake.

The main problem in NS argumentative writing is not in presenting a point, but in supporting it beyond general statements and personal likes and dislikes (Applebee *et al.* 1986). Great influence on the ability to argument for or against a point can be caused by the preceding discussion of the topic in class. Students can gain from gathering together to discuss a topic and carefully sorting the relevant information. This factor can also be of importance for learner writing. A discussion on the topic can initiate new ideas, help structure the essay and focus on argumentation. Under the pressure of the time limitations and the absence of reference sources,

non-native speakers may not be able to put their thoughts together and organize their essay properly. Ädel (2008) showed that timing and no access to secondary sources provoke a more involved learner style. Students could also benefit from a short discussion before writing an essay. It could initiate new ideas and help students structure their arguments. A further investigation would be required to make any statements.

Young students traditionally have more problems writing persuasive discourse, which involves argument, than narrative, descriptive or expository tasks. Stein and Glenn (1979) point out that narrative is the basis of our daily communication and thus is learned without any extra effort. Argument, on the other hand, cannot be learned in the course of daily interchange, though '[G]iven the presumed importance of argument, it is unclear why children and adolescents display so little skill in its use' (McCann 1989: 63). In a broad sense, argumentative writing is important as an essential tool for the political, social and ethical development of society, whereas in a narrow sense – as a key part of students collage career (*ibid*).

In this section, I wanted to stress that argumentative writing is a complicated multifaceted challenge for every writer. Producing an essay implies not just good language skills, but a profound understanding of argumentative techniques. To argument successfully, one has to grasp the topic. At this stage, a discussion can be of much help. Secondly, one should feel free to express oneself in the language one writes in, including the grammar and spelling aspects of it. The settings of an exam can make learners focus on the formal correctness of the text, and not the nativeness and style issues, even when they are aware of them. Finally, the limitations of vocabulary can prevent learners from using warrants, which is a key point of every argumentative essay. Instead, they appeal to explicit presence of the author in the text as the argumentation techniques.

#### **2.3.4 Concluding remarks on author identity in argumentative writing**

The nature of argumentative writing implies a more apparent presence of the author in the text; especially, when tasks are presented in the form of a question, initiating students to express their personal 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?**



- Some people say that in our modern world, dominated by science technology and industrialisation, there is no longer a place for dreaming and imagination. **What is your opinion?**
- 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 topics above were used in the compilation of the ICLE corpus. Such tasks induce students to take a stance. Recski (2004) argues that it is hard to avoid personal references, when they are explicitly requested in the task. A student should present the ‘for’ and ‘against’ arguments, then support personal opinion and come to a conclusion. Argumentative writing implies higher involvement level than, for example, academic writing. However, native speakers tend to construct author identity using more diverse means and not to rely heavily on the personal pronouns and other involvement features.

The review of the previous research on learner argumentative writing reveals that learners express their author identity majorly through the use of personal pronouns and have a rather involved style. The results for learner groups vary. Scandinavian learner writing stands out as the most involved one, and the possible reason for it is the effect of L1 transfer. However, in the majority of cases *I* is used as an organizer, and *we* and *you* in its generic sense. This implies that these personal pronouns are not involved in shaping author identity in the text, and the overall involvement level is lower than expected. Author identity can also be realised through the use of correct reference verbs in citing (John 2012): *claim, suggest, point out, state* and others. Learners underuse and misuse the verbs and, in this sense, let out of view an important instrument for realising author identity and may also fail to correctly communicate their viewpoint.

Altenberg (1997) suggests that poor register awareness is the reason for the explicit presence of NNS writers in discourse. Ädel (2008), on the other hand, argues that learners are aware of register, but are influenced by exam settings such as timing and no access to secondary sources.

My own investigation focuses primarily on Russian learner writing, but also carries out a small-scale study of the three Slavic learner varieties: Polish, Czech and Bulgarian. These varieties have not been explored from this perspective before. Thus, my investigation will contribute to the field of second language acquisition and give a wider perspective on the issue of high level of involvement amongst learners.

## 2.4 Author identity in academic writing

Academic writing is fundamentally different from the general type of argumentative writing outlined above. If the main purpose of the latter is to express one's personal opinion and to argue for it, then the purpose of the former is to introduce facts and objective results of a research. The settings of academic writing are typically different from those of argumentative: academic writing is usually an untimed activity with a free access to secondary sources. Taking into account Ädel's findings (2008), one should expect a lower level of W/R visibility in academic writing. Besides, the tasks here do not directly prompt writers to express their personal opinion, which, according to Recski (2004), influences the writers of argumentative writing. The (dis)interest in the topic of the discourse should not have any influence on the level of W/R visibility. Students usually write academic papers in the disciplines they major in, which one can expect to be a conscious decision.

The use of personal pronouns in academic discourse is a controversial issue in many studies. Some claim that personal pronouns are a key feature for creating a successful academic discourse (Hyland 2002a, 2002b; Tang and John 1999; Kuo 1999; Ivanić and Camps 2001). Others argue that such discourse should be objective and follow the conventions of Anglo-American rhetoric, which traditionally presupposes that interlocutors are not explicitly visible in written discourse (Spencer and Arbon 1996; Gong and Dragga 1995; Arnaudet and Barrett 1984).

A traditional view of the problem implies that academic writing should be objective and there is no place for personal pronouns. Personal pronouns signal interaction and involvement with the audience when less attention is paid to the organisation of discourse. The speaker and the listener share the same environment and knowledge, and misunderstandings can be discussed and eliminated. Written discourse, on the contrary, is expected to be thought through and logical. Writers, relying heavily on personal experience in argumentation, can fail to successfully defend their academic position. The following citations from textbooks are presented by Hyland (2002b) to illustrate this viewpoint:

*The total paper is considered to be the work of the writer. You don't have to say 'I think' or 'My opinion is' in the paper. [...] Traditional formal writing does not use I and we in the body of the paper.*

*(Spencer and Arbon 1996: 26 as cited in Hyland 2002b: 351)*

*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 as cited in Hyland 2002b: 351)*

A search on the Internet reveals that this viewpoint is in part shared by some universities. For example, an online Writing Center at the College of Arts & Sciences (Chapel Hill, US) explains that 'first person' and 'personal experience' may sound like the same thing to students, but they can be used in two different ways in writing.<sup>3</sup> A student may choose to use 'I' or 'we', but not refer to any personal opinion or introduce his/her own findings. On the other hand, a student may introduce an experience to illustrate a point without using 'I' or 'we'. In this way, the use of personal pronouns and the use of personal experience are two separate questions, which may be more or less appropriate in one or another academic project.

Though the avoidance of 'I' has long been a part of the academic tradition, it is important to mention that the situation has been changing lately. Nowadays, some academics consider the use of personal pronouns to be acceptable and even crucial in creating a clear and interesting academic paper. This view is expressed by Hyland (2002a, 2002b), who claims that 'academic writing is not just about conveying an ideational "content", it is also about the representation of self' (Hyland 2002a: 1091) and that 'writers gain credibility by projecting an identity invested with individual authority, displaying confidence in their evaluations and commitment to their ideas' (*ibid.*). Hyland (2002b) underlines that academic writing should not be regarded as a single unity, but as a continuum of subject-specific entities. Disciplinary conventions differ. Humanities and social sciences tolerate a greater degree of involvement than technical disciplines. The reason may be that arguments in the former ones are less measurable and clear-cut than in the latter ones. In sciences, writers make a greater use of plural forms because of the preference for joint authorship. Disciplines vary and one approach to all of them would result in an uncreative and unsuccessful writing. The influence of academic disciplines and academic experience on W/R visibility and author identity is discussed in detail in Sections 2.4.3 and 2.4.4 respectively.

The overuse/underuse of personal pronouns in learner writing does not always result from the conventions of academic writing. The overuse may result from L2 students finding it easier to use active construction with agent in subject position, as a desire to avoid these

---

<sup>3</sup> More information is available at the official webpage of Writing Center at the College of Arts & Sciences (Chapel Hill, US) <https://writingcenter.unc.edu/handouts/should-i-use-i/>

constructions may lead to grammatical or lexical mistakes. On the other hand, students may underuse personal pronouns as they may believe that 'I' and 'we' are not appropriate in academic discourse. Cultural features may also cause underuse. As shown by Hyland (*ibid.*), Hong-Kong students do not feel comfortable with the strong writer authority 'I' and 'we' imply. The next section gives a brief account of the way native speakers use personal pronouns to construct their author identity in academic texts. The results are used in Section 2.4.2 as a yardstick against which the writing of L2 writers can be compared, and deviation can be identified. Such comparison does not imply any negative connotation. As pointed out in Section 2.1, it is aimed at working out pedagogical implications to create learner awareness and to help L2 students successfully communicate their message within the framework of Anglo-American rhetoric.

#### **2.4.1 Author identity in NS academic writing**

Previous corpus-based studies reveal that NS academic writing is characterized by a specific type of vocabulary (Paquot 2010; Biber *et al.* 1999; Coxhead 2000) and a number of grammatical features (Biber 2006; Kertz and Haas 2009). One of these features is typically discussed in connection with author identity, namely first person pronouns. Hyland presents an in-depth study of first person pronouns in his paper 'Options of Identity in Academic writing' (2002b). In the study, the results from L2 writers are compared to expert writing. However, it is not clear from the study if expert writers are native speakers or L2 professional writers. Expert writers are reported to use first person pronouns 'when they are presenting their claims and bottom-line results, and intrude into the text to clearly link themselves with their main contribution (*ibid.*: 353). The use of fewer personal pronouns indicates a less personal style and strengthens the effect of objectivity by means of letting the results of the research speak for themselves. Exploring rhetorical functions of first person pronouns helps to identify the points where writers choose to make themselves visible. Professional academic writers use personal references in the most essential parts of academic discourse: to present arguments, claims and results (*cf.* Table 3):

**Table 3.** Discourse functions of self-mention in sample of 10 research articles (%), according to Table 5 in Hyland (2002a).

Function	Total	
	Raw	%
Explaining a procedure	400	38
Stating results or claims	273	26
Elaborating an argument	220	21
Stating a goal/purpose	158	15
Expressing self-benefits	0	0
Total	1051	-
Percentage	-	100

Almost half of the cases of self-mentioning are used to present results or claims or to elaborate an argument. These are the most high-risk functions, which involve reasoning and making a knowledge claim. It is therefore logical that authorial identity would be most visible here, as in this way, authors express commitment to their results and arguments. Writers typically use personal pronouns in thematic position and present new ideas and concepts in the rheme. Thus, the writer emphasises the important information and performs as the source of the statement. This allows the author to seek agreement for the statement and leaves the reader with a clear understanding of the writer's position. Consider the following examples from (Hyland 2002a: 1093):

*I* will show that a convincing reply is available to the minimalist. (Philosophy)

*We* shall prove, however, that this is not the case. (Physics)

First person pronoun is a powerful way of expressing author identity, which is, according to Hyland (*ibid.*) a key to successful academic writing.

The study by Paquot, Hasselgård and Ebeling (2013) reports that there is significant decrease in the use of W/R visibility from argumentative to academic writing. The analysis of *I*-clusters (*ibid.*) reveals that in argumentative writing, *I* is used with verbs of expressing opinions, whereas in academic writing, it is used to refer to the procedures of conducting linguistic analysis and writing a paper.

Some interesting features of NS writing are presented in the study on the phraseology of English academic writing by Howarth (1996). He points out that academic register is highly conventional, and less than 1% of collocations are deviant. The reason for this is that such writing is supposed to be clear, and minimum attention should be drawn to any individual features or the style of the author. This finding hints in the direction of English academic writing

concentrating on the clarity of thought, meanwhile pushing the personality of the individual writer into the background.

A different way of expressing author identity is discussed in the study by John (2012). It deals with citation sequences in literature reviews of dissertations. Reviewing external sources and placing one's academic work in the context of other research writing, helps authors to make their argument sound convincing. The results from John's (2012) paper have already been touched upon in the section on argumentative writing; however, they are even more relevant in this part of the master thesis. Author's presence in a text can be signalled by verbs with evaluative component such as *point out* and *claim* or verbs with neutral connotations such as *suggest* and *find* accompanied by adverbs like *strongly*, *critically*, *aptly* and others. These verbs are used to express author's evaluation of the cited source. The verb *find* provides no evaluation about citation, while *suggest* makes the author merely visible. The verb *claim*, in contrast to the last two verbs, bears rather strong evaluative information and implies disagreement between the author and the cited writer. As for *point out*, it indicates that the author supports the opinion expressed in the citation. This way of expressing W/R visibility is natural for native speakers of English, but can be a great challenge for learners. Non-native speakers need to keep in mind the connotations of such verbs and make good use of them in order to express their author identity successfully.

For many L1 undergraduates, there are many challenges related to constructing their author identity in academic discourse. They have to adopt a new identity, which is be able to select, evaluate, report and conclude. This new 'I' may be very different from what students are used to in their day-to-day life.

#### **2.4.2 Author identity in NNS academic writing**

While academic writing can be challenging for L1 students, it is a great deal more so for language learners. L2 students come from cultures that may be significantly different from the English one. Due to this, learners may experience problems in conveying their author identity. Anglo-American rhetoric implies that writers position themselves in relation to other research results and look for a way to introduce new arguments and findings. Learners' L1 rhetoric may differ considerably from Anglo-American academic conventions. For example, cultures with a more collective mind may find it challenging to express their individualistic identity, which can result in underuse of personal pronouns. Deviations in learner writing may also result from the use of personal pronouns in non-nativelike functions.

The results from the study by Paquot, Hasselgård and Ebeling (2013) show that both French and Norwegian learners strongly overuse first and second person pronouns, but at a lower rate than in argumentative writing. This means that L2 students are aware of register differences and use a more formal style in linguistic assignments than in argumentative essays. Just like in NS writing, *I*-clusters in learner writing indicate genre variation: the clusters from ICLE express opinions and from VESPA refer to analytical processes (*ibid*).

While the results above report the overuse of personal pronouns in French and Norwegian L2 writing, Hyland (2002a) found underuse of these features in the academic writing of Hong Kong students. He explains that cultural norms are a complicating factor, and L2 writers from certain backgrounds may be reluctant to promote their author identity. This study also reports on the functions of first person pronouns in learner writing (*cf.* Table 4). These functions are adopted in my qualitative study of the pronouns *I* and *we* in L2 writing of Russian students.

**Table 4.** Discourse functions of self-mention in NS vs. Hong Kong student writing (%), according to Tables 4 and 5 in Hyland (2002a).

Function	Hong Kong student writing		NS writing	
	Raw	%	Raw	%
Explaining a procedure	199	31	400	38
Stating results or claims	103	16	273	26
Elaborating an argument	49	8	220	21
Stating a goal/purpose	228	36	158	15
Expressing self-benefits	58	9	0	0
Totals	637	-	1051	-
Percentage	-	100	-	100

The function of *expressing self-benefits* is absent in professional NS writing, but constitutes 9% of the L2 student writing. Self-mentioning here is least risky and does not require the writer to step outside the familiar student identity. The majority of the uses occur in the conclusion of the paper, where students explain what they personally gained from the assignment. *Stating a purpose* is another metatextual function, which implies little risk. It is used to set aims and explain the direction of research. Over one third of all uses in student writing had this function in contrast to only 15% in NS writing. The function *explaining a procedure* is slightly more often used by NS professionals than by learners (38% vs. 31% respectively). Both groups recognise the importance of clearly stating their methodological approach and explaining research procedures. This is a completely justified use of self-reference, though it implies a low level of personal exposure. The two high-risk functions are *elaborating an argument* and *stating results/claims*. Both are underused by L2 students. Few learners collocate personal pronouns

with cognitive verbs like *think*, *believe* or *assume*. L2 students choose to stay neutral and not to express either agreement or disagreement with a position. Grammatical structures favoured by learners indicate their desire to avoid the ‘potentially problematic role of writer-as-thinker’ (*ibid.*: 1103). Hyland points out that such ‘tentativeness and reluctance to display an authoritative persona among Asian writers may, in part, be a product of a culturally and socially constructed view of self which makes assertion difficult’ (*ibid.*: 1111). It will therefore be important for me to keep in mind the Russian cultural views of self when investigating the corresponding L2 learner writing.

Other interesting features of learner academic writing are presented in the study by Gilquin and Paquot (2007). This large-scale project on rhetorical functions (*e.g.* exemplification, reformulation, concession or the expression of personal opinion) in academic writing reports the overuse of many lexical and grammatical features typical of spoken language: first and second person pronoun (as already discussed above), or short Germanic adverbs (*also*, *only*, *so*, *very* and others). It also reports the underuse of formal writing features, for example, nominalization. What could be of particular interest to the present research is that all learners, irrespective of their L1 background, make themselves explicitly visible through the overuse of a number of expressions, the so-called ‘lexical teddy-bears’ (Hasselgren 1994) : *to my mind*, *from my point of view*, *it seems to me* – to express opinions, and *I would like to/want/am going to talk about* – to introduce a new topic. Gilquin and Paquot (2007) underline the harmful influence of lists of expressions presented in textbooks with no further comments on how to use them.

As discussed by Granger and Paquot (2009), EFL students underuse a range of verbs belonging to *Academic Word List* (AWL), which was compiled to help students meet academic requirements. Among the underused verbs are such as *achieve*, *contribute*, *demonstrate*, *establish*, *examine*, *identify*, *provide*, *report*, *suggest* and *support*. The verb *analyse*, on the other hand, is overused. Though Hyland and Tse (2007) question the existence of Academic Vocabulary, I believe it can be a helpful instrument for the present investigation. Examining some of the key verbs from AWL can provide us with an idea of how academic/unacademic learner writing is.

It seems that learner academic writing is strongly influenced by the background students come from. In some cultures authors are not supposed to promote their individual self, whereas in others, writers are typically rather visible. L2 authors often introduce the conventions of academic writing from their mother tongue into English. A successful academic discourse in L2 implies that learners move away from what they are used to in their culture and try to function



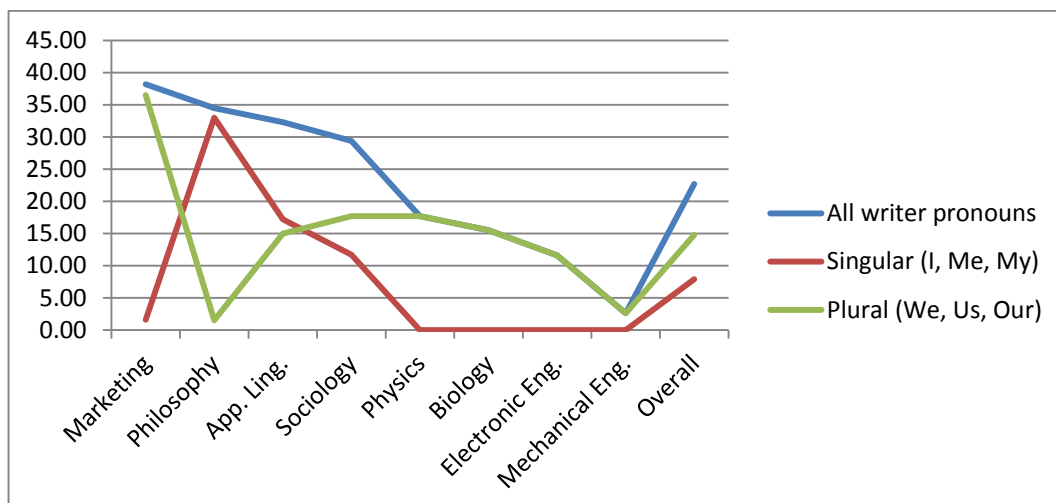
within the framework of EFL conventions. Learners should be made aware of where first person pronouns are appropriate in academic discourse and what other instruments are available for them to express their authorial identity.

### 2.4.3 Author identity within the scope of academic disciplines

Hyland and Tse (2007) deny the existence of a general *academic vocabulary* and call it a chimera. They believe that ‘all disciplines shape words for their own uses’ (*ibid.*: 240) and thus insist on investigating the *academic vocabulary* separately for each of the disciplines. Though my view over the subject is not equally radical, I still find the interdisciplinary variation significant for the present master thesis, as my study of academic writing relies on the linguistics component of BAWE and VESPA.

Gardner and Holmes (2009) describe how challenging it can be for students to get a constructive piece of advice on disciplinary writing conventions from a university writing centre or EAP programme tutor. The major challenge is that tutors rely on disciplinary norms they are familiar with. Therefore, talking to three different tutors may yield three different results. Hyland (2002b) reports that author identity is typically downplayed in engineering and hard sciences and more visible in humanities and social sciences. Naturally, the majority (75%) of all author pronouns occur in the latter disciplines. Hyland (*ibid.*) claims it is due to results in these research areas being less precisely measurable and clear-cut. Writer in engineering and hard sciences, on the other hand, are less visible in discourse due to their desire to sound more objective (*cf.* Figure 4):

**Figure 4.** Average frequency of first person pronouns per research paper, according to Table 1 in Hyland (2002b)



The frequencies in the Hyland's (*ibid.*) study are given per paper, which is slightly vague, as there is no information on the length of the papers. However, according Figure 4, authors in social sciences and humanities make more use of both singular (*I, me, my*) and plural (*we, us, our*) first person pronouns, whereas writers in engineering and hard sciences extensively use only plural forms due to frequent joint authorship. The findings for social sciences and humanities are most relevant for my own investigation, which relies on academic assignments in linguistics. I can therefore hypothesize that the present investigation will reveal a slightly higher use of first person pronouns than what can be found in engineering and hard sciences.

This chapter can be summed up by some remarks from Ivanič (1998). She points out that author identity is not something new or individual, but established through the discourse the writer adopts. Learners, therefore, do not learn to be creative, but learn to apply the discourse that already exists – creatively. The notion of discourse community is one of the key concepts in a study of author identity, because each student adopts an identity depending on the community they come in contact with. Discourse communities introduce a social element into the construction of academic identity, as it is through the identification with, the membership of, the practices and values that a person's identity is built up. According to Fløttum *et al.* (2006b), the process of academic socialisation involves communicating the knowledge using the rhetoric of a discipline. In the light of the globalisation process in the academic world, disciplines acquire recognisable identities that are cross-cultural (Becher and Trowler 2001).

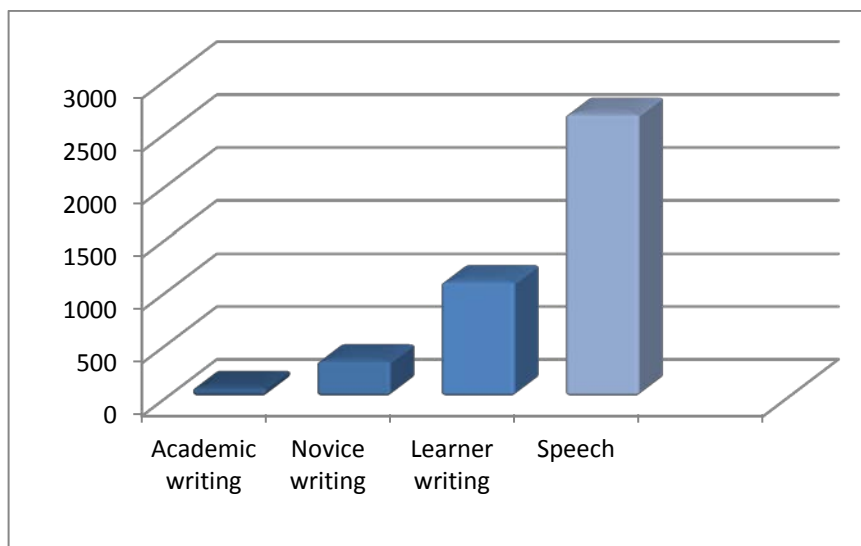
#### **2.4.4 Author identity on various academic levels**

The VESPA corpus opens new possibilities for investigation in the field of learner writing. The fact that VESPA contains texts produced by BA and MA students, presents researchers with a possibility to study how developmental factors influence the level of involvement in academic writing. My own study of Russian learner writing discusses the impact of this factor in Chapter 6, together with other potential reasons for overuse/underuse. Previously, developmental factors were investigated by Gilquin and Paquot (2008) and Fossan (2011).

The first study compares speech with the writing of learners, native students (novice writers) and experts. As it had been published before the VESPA corpus was compiled, the study makes use of the corpora available at the time: ICLE for learner writing, LOCNESS for NS novice writing, the academic and spoken components of the *British National Corpus* (BNC) for the expert writing and speech. It is important to keep in mind, that ICLE and LOCNESS are two corpora of argumentative writing (see Chapter 3 for further information on the corpora). As

discussed in Sections 2.3 and 2.4, this implies a more involved style in comparison to academic discourse. The relative frequencies of *I think* in the four corpora are presented below (cf. Figure 5):

**Figure 5.** Relative frequency per million word of *I think* in academic writing, novice native student writing, learner writing and speech according to Figure 10 in Gilquin and Paquot (2008).



The figure above shows that NNS and novice NS writing takes an intermediate position between academic writing and speech. The overuse of the spoken-like unit *I think* is more marked in learner writing than in novice student writing. However, one should mention that the spoken nature of writing is not simply typical of learner writing, but is a stage on the way towards expert writing. In Chapter 4 of this thesis, I conduct a similar study of first person pronouns with a focus on Russian learner writing. My results reveal the same tendency as in the work by Gilquin and Paquot (*ibid.*) (cf. Figure 6).

Fossan (2011) investigates the influence of developmental factors in the writing of BA and MA based on BAWE-LING and the Norwegian component of VESPA. The pronouns *I*, *we* and *you* are studied first separately and then combined in the writing of Norwegian bachelor and master students. The results suggest that the overall level of involvements decreases with the increase of academic experience. However, the use of the pronoun *I* increases in Norwegian student writing, though the increase is not statistically significant. These results indicate that academic experience together with increasing proficiency in English and age factor lead to a more native-like performance with regards to the use of personal pronouns. Similar results are registered in NS student writing. Students become less personally involved in their writing in the process of gaining academic competence.

Both studies presented above conclude that developmental factors influence the level of involvement in learner and in NS writing. However, it is not clear if there are any changes in the functions of personal pronouns in MA writing compared to BA writing. Such results would indicate if students become more aware of the role of personal pronouns in academic discourse.

#### **2.4.5 Concluding remarks on author identity in academic writing**

Academic discourse implies a lower level of involvement than what is characteristic of argumentative writing. The decrease in the use of personal pronouns is registered both in NS and in learner writing, which means that L2 students are aware of register differences and adopt a more formal style in academic discourse than in argumentative essays (Paquot, Hasslegård, Ebeling 2013). However, compared to L1 students, their L2 peers continue to overuse or underuse personal pronouns in academic assignments depending on their L1 background. At the same time, research reports the underuse of formal writing features, for example, nominalization and a range of verbs belonging to AWL (Gilquin and Paquot 2007, Granger and Paquot 2009).

The functional analysis of first person pronouns reveals that learners overuse low-risk functions and underuse high-risk functions compared to NS writers. Typical low-risk functions include stating a purpose, explaining a procedure and expressing self-benefits, whereas high-risk functions are elaborating an argument and stating results/claims. To function successfully within the framework of L2 academic conventions, learners should be made aware of how first person pronouns function in Anglo-American academic discourse, and what alternative means they can use to express their authorial self.

The use of first person pronouns depends on disciplinary conventions. The most relevant for the present study are the findings for social sciences and humanities, where authors tend to make a greater use of both singular (*I, me, my*) and plural (*we, us, our*) first person pronouns. Writers in engineering and hard sciences extensively use only plural forms, which is due to frequent joint authorship. Fossan (2011) and Gilquin and Paquot (2008) report the significant influence of developmental factors. The increase in academic experience leads to decrease in the level of W/R visibility. The academic writing of Russian learners is investigated in Section 4.2 of this master thesis. The impact of developmental factors is analysed in Section 6.3, and the functional analysis of the pronouns *I* and *we* is presented in Sections 5.3 and 5.4 respectively.

## 2.5 Anglo-American and Slavic rhetoric

Since the writer's background can in part be the reason for downplaying or overtly presenting author identity in writing, I find it important to give a brief comparison of Anglo-American and Slavic rhetoric. In Chapter 5 and 6, I refer to the concepts discussed in this section in order to account for the findings of my research.

The study by Hyland (2002a) illustrates that while Anglo-American rhetoric implies a deliberate use of authorial identity to communicate author's ideas, other cultures may purposefully downplay their authorial presence in discourse. For example, Asian writers may in part be reluctant to promote their author identity due to culturally and socially induced factors. This is reported to be the case with Hong Kong students in Hyland's (*ibid.*) study. Other cultures, on the contrary, adopt a more spoken-like style in L2 academic writing as a result of transfer of L1 academic conventions. A good example would be the studies of Norwegian by Paquot, Hasselgård and Ebeling (2013) and Fossan (2011) or the studies of Swedish by Ädel (2008) and Aijmer (2001).

Anglo-American rhetoric implies author responsibility and is commented by Bakhtin (1986) in the following way: '*I*-authors are positioned by their discursal choices as single authors, as independently responsible for the contents of their writing' (*ibid.*: 301). In English, the use of *I* is preferable in high-risk functions: Elaborating an argument or stating results and claims. Self-mentioning is a central part of Anglo-American rhetoric, which focuses on individualistic approach and downgrades group solidarity. A different feature typical of English academic writing is that it ought to be clear and not overtly complex. According to Day and Gastel (2006), a scientific text is not a literary work, and thus scientists should avoid metaphors or any other means of altering the style of the paper. These academic conventions may appear unnatural to the students from non-western backgrounds. If learners do not receive instructions on how to function within the framework of English academic discourse, they risk either adopting L1 conventions into L2 texts or following their introspection about the interlanguage.

Russia takes an intermediate position between the individual Western world and the collective Asian mentality. Ohta (1991) and Scollon (1994) report that the use of the pronoun *I* is almost unacceptable in the Asian tradition as it promotes individual identity. These conclusions go alongside with the results from the study by Vassileva (1998), who investigates the '*I*' and '*we*' perspectives in research articles from five academic discourses, two of them being English and Russian. She claims that the '*we*' perspective predominates academic discourse in the two Slavic languages she investigates: Russian and Bulgarian (*cf.* Table 5)

**Table 5.** *I/we* occurrences in the corpora of research articles in linguistics according to Table 1 in Vassileva (1998).

	<b>English</b>	<b>German</b>	<b>French</b>	<b>Russian</b>	<b>Bulgarian</b>
<b><i>I</i></b>	69%	47%	40%	0,5%	6%
<b><i>we</i></b>	31%	53%	60%	95,5%	94%
<b>occurrences</b>	526	227	153	203	300

According to the study, Russian and Bulgarian experience the influence of the English tradition in a much smaller scale than, for example, German or French. Vassileva (*ibid.*) explains the dominant role of the ‘we’ perspective in Russian academic discourse as a result of ‘the long-standing and powerful influence of communist ideology’ (*ibid.*: 181), which ‘aims at suppressing the individual in favour of the community’ (*ibid.*). I suppose that this explanation would be the most common one among Russians themselves if they were asked to comment on the predominance of the ‘we’ perspective.

A different study contrasts English to another Slavic variety – Czech. Čmejrková (1996) suggests that Czech academic texts correspond in many ways to German conceptions and in a lesser degree to English academic texts. Czech is a reader-responsible language (Hinds 1987), where emphasis is on supplying the reader with knowledge, theory and stimulus to think (*ibid.*). Here, readers are supposed to read between the lines and to make an effort to understand the content. English is a writer-responsible language (Hinds 1987), where the writer is supposed to make text clear and easy to read. In Anglo-Saxon tradition, writing academic texts is a part of the general theory of writing, which has a rather strong emphasis on the interactive nature of any writing process. Czech structuralistic stylistics opposes scientific functional style to the other writing styles and insists that scholarly writing is aimed at a distant or unknown addressee. These conclusions by Čmejrková (1996) derive from Clyne’s (1987) comparison of German and English, and Hinds’s (1987) classification of languages into reader- or writer-responsible. A high level of formality, non-interactive nature and distanced addressee allows me to hypothesise that Czech and, if generalized, other Slavic languages could potentially have a lower level of involvement than English in academic discourse. Currently, this remains only a hypothesis, which is investigated in the practical part of the present master thesis.

Previous research reveals that Slavic languages belong to a different rhetoric than the Anglo-American one. As pointed out by Maurinen (1996) and Vassileva (1998), it is important to increase general awareness of global academic rhetoric, which results from specific and culture-induced strategies. At the same time, authors who aspire to succeed as the members of the global academic community have to develop their English discourse competence. My

research is a modest contribution to this field and reveals some features that Russian learners need to be made aware of in order to communicate their message more successfully in English.

## **2.6 Reasons for overuse, underuse and misuse of first person pronouns**

The majority of EFL studies intend to work out solutions, which can help learners to tackle the problems they encounter. To find a solution, one has to understand the roots of the problem. Therefore, the present section considers the possible explanations of the overuse, underuse and misuse of first person pronouns in learner writing. Gilquin and Paquot (2007) provide a profound overview of the potential reasons. Each of these reasons is discussed in the sections below. The influence of these factors on Russian learners is studied in Chapter 6 of this master thesis.

### **2.6.1 Influence of speech**

A number of studies conclude that learner writing demonstrates a variety of spoken-like features. Altenberg (1997) found the overuse of various involvement features in Swedish, Crawford (2005) in German, Spanish and Bulgarian, Narita and Sugiura (2006) in Japanese, Neff *et al.* (2007) in Spanish, Paquot (2008) in French, Hasselgård (2009) and Fossan (2011) in Norwegian. The reason could be that, EFL students first learn to communicate orally and only at later stages, they practise in writing. The influence of speech could be of significant influence in countries, which are exposed to American or other English-speaking media: movies, radio or internet. Norway and Sweden are good examples in this respect. In these countries, people may gain a good command of English through watching TV, as many programmes and movies here are not dubbed, but only provided with subtitles.

The influence of speech can, however, hardly be true for the majority of EFL students, who learn English through instruction. In such cases, spoken input is limited to conversations in class and listening exercises. In Russia, for example, there is no broadcast or sale of movies or programmes in English. One can of course gain access to such materials through the Internet, but English is still far from being a part of an everyday life. The majority of Russians learn English through language courses, which do not reflect the complexity and density of NS speech. I would, however, not completely deny the influence of register awareness on L2 writing, and investigate this reason for overuse/underuse in Section 6.4. I believe that students may benefit from extra training in the use of domains.

### 2.6.2 L1 transfer

The influence of the mother tongue is recognized by Gilquin and Paquot (2008) as one of the main factors that forms learner's interlanguage. Learners knowledge about an L1 word is usually accompanied by such information as the word's collocations, sentence position, stylistic and register characteristics, the frequency of use and discourse functions. This type of L1 knowledge is especially strong among adult learners. The same information for the English counterpart may be absent or weak in the mental lexicon of a learner. Therefore, the learner is likely to borrow the L1 lexico-grammatical and stylistical preferences into L2 (Paquot 2010).

Gilquin and Paquot (2008) illustrate L1 transfer with examples from French. French makes great use of first person plural imperatives, which possibly results in the overuse of *let's/let us* in their L2 student essays:

(1) ***Prenons l'exemple des*** sorciers ou des magiciens au Moyen Age (CODIF)

*'Let us* take the example of wizards or magicians in the Middle Ages'

(Gilquin and Paquot 2008: 54)

The researchers (*ibid.*) conclude that transfer cannot be the only reason for the overuse, because the imperative form is also overused by other learner varieties (*e.g.* Dutch), which do not typically use first person plural imperatives in academic writing.

### 2.6.3 Teaching-induced factors

In the process of learning a language, students are at various points presented with lists of words and phrases that they can use to express their opinion or to structure a text. They include such as *in my opinion; personally, I think/believe; to my mind; firstly, I would like to; let us start by considering the facts; in conclusion, I can say that* and others. Gilquin and Paquot (2008) point out that these units provoke learners to use first person pronouns without considering the function they use them in. Such lists make learners think that all phrases are equally frequent and stylistically appropriate in any context. To reduce the overuse and misuse of first person pronouns, each phrase should come with detailed information on how native speakers use it and what function it performs. The influence of teaching-induced factors is explored in Section 6.3.



#### **2.6.4 Developmental factors**

As previously mentioned, learners may have problems in writing similar to those of novice writers in English-speaking countries. A major challenge is not in presenting an argument, something people master at the early stage of learning a language, but in supporting it beyond general statements and personal likes and dislikes. The more students practise the language, the easier it becomes to present claims and less personal involvement is registered in students' essays. Therefore, high level of involvement is the result of learners being novice writers and lacking knowledge about formal academic style. Developmental factors influence the writing of native speakers as well as learners. Gilquin and Paquot (2008) compare learner writing with novice NS writing and report that 'while native students do better than non-native students with respect to lexico-grammar and phraseology, they share learners' problem with register to a certain extent, overusing items which are more typical of speech than writing' (*ibid.*: 56). The spoken-like nature of writing is not exclusively a learners' problem, but the problem of all writers at an early stage of developing their expert writing skills. Developmental factors are considered in Section 6.2 of this master thesis.

#### **2.6.5 Extralinguistic factors**

Learners often have to write essays on the topics that do not appeal to them and that they have little knowledge about. When learners have no facts to rely on, they use their authority to persuade the reader and to argument for or against a statement. Virtanen (1998) discusses this factor together with other extralinguistic factors in the following way:

*'First, if students are faced with a topic lacking in personal interest, they often have problems inventing what to write about. The writing process is also affected by the unnatural setting, the ill-defined or fictional audience and the teaching/testing purpose of the task' (ibid.: 94).*

The study by Ädel (2008) contributes to the understanding of how extralinguistic factors can influence learners' performance: when students are not limited by time and have access to secondary sources, they show lower level of involvement. Students may not have enough time to adopt a more formal style of writing and not have model texts to aspire to.

Finally, one can often hear that Anglo-American rhetoric is more liberal in comparison to other rhetorics. Students may have their own view of what 'liberal' means and may purposefully introduce a more spoken-like style. Careful and considerate instructions from teachers are required to correct students' view over the conventions of English academic writing.

To sum up, I think it is hardly possible to single out one explanation for the overuse or underuse of a feature in learner writing. It is always a mixture of various factors. However, one or another factor may be dominant for a feature in a learner variety, which can yield important results and help learners to communicate their message.

## 2.7 Functions of the first person pronouns *I* and *we*

A quantitative analysis presents interesting data about the overuse and underuse of first person pronouns. However, it is not enough to know that EFL students should reduce or increase the use of these pronouns. One has to inform them how *I* and *we* function in the NS writing. This is where qualitative analysis plays a key role. Comparing the functions of first person pronouns in NS and learner writings reveals the challenging for L2 writers areas and provides for a more competent piece of advice on how to achieve native-like performance. This section discusses existing classifications of first person pronouns, which are adopted in the qualitative study of L2 writing of Russian learners in Chapter 5.

Previous studies of argumentative and academic writing indicate that author pronouns fulfil different functions in these genres. Therefore, I will consider the functions of the pronouns *I* and *we* in argumentative and in academic writing separately.

### 2.7.1 Functions of *I* in argumentative writing

One of the main differences in the use of *I* in NS and learner corpora is pointed out in the study by Petch-Tyson (1998). NS students typically use *I* to **introduce their personal experience** into the text. Such uses are often signalled by verbs in the past tense:

(2) I honestly ran out of my room.

(3) I read the above quote to ask.

(*ibid.*: 113)

Learners, with the exception of Finnish students, tend to underuse this function. Instead, they overuse *I* in the function of '**either saying something about the writer functioning within the text or what the writer thinks**' (*ibid.*: 111, 114). These two non-nativelike uses of *I* are typically found in the surrounding of a number of NNS features: chains of interpersonal features, end placement of *I think* and *I guess*, and emphasisers of the writer's opinion. Hasselgård's (2009) findings for ICLE-NO support the conclusion that the frequent use of self-reference is a striking feature of learners' essays. Hasselgård (*ibid.*) specifies that *I*, as the source of an opinion,

takes thematic position. In ICLE-NO, the pronoun often collocates with verbs expressing mental and verbal processes (for example, ‘think’ or ‘say’).

Two additional sub-functions result from the analysis of *I think* by Aijmer (2001) – the function of **overstatement** or **reinforcing a statement**. This function is typical of learner writing, though even in NS writing one can run across the overstatement phrase *personally I think*. The second function of *I think* is to **signal uncertainty or to weaken the claim** made by the speaker. In this function, *I think* usually comes in the non-initial position.

The overview of the previous studies allows me to distinguish three main functions and two sub-functions of the first person pronoun *I* in argumentative writing:

- recounting personal experience;
- organizing the text;
- expressing personal opinion: to reinforce a statement or to weaken a claim.

### 2.7.2 Functions of *we* in argumentative writing

Neither English nor Russian distinguishes between exclusive and inclusive uses of the pronoun *we*, except for (partially) the imperative *let us* (Harwood 2005). This lack of semantic diversity ‘can help to create the sense of newsworthiness and novelty’, though can be challenging for analysing (*ibid.*: 365). Herriman (2009) investigates how the pronoun *we* functions in argumentative writing. She points out that the pronoun *we* may or may not imply the presence of the reader or writer and, therefore, distinguishes between the **inclusive *we*** and **exclusive *we***. Each of these, in its turn, subdivide into inclusive specific *we*, inclusive authorial *we* and **generic *we***, and exclusive specific *we* and exclusive authorial *we*. Comparing argumentative writing of Swedish learners to NS students, shows that the majority of uses in both cases are generic: 95% and 91% respectively. Herriman (2009) concludes that student writers apply generalisations in their argumentation. Professional writers, however, present their arguments using personal experiences and, therefore, make greater use of exclusive *we* (39% of occurrences). To analyse the writing of Russian learners, this study distinguishes between inclusive, exclusive and generic *we*, but leaves out the sub-division into specific and authorial.

### 2.7.3 Functions of *I* in academic writing

Fløttum *et.al.* (2006a) discuss author roles in academic disciplines in connection with verbs. The first role is **the role of a writer**, which is signalled by verbs presenting an idea or research (*describe, illustrate, present, summarise*), or verbs structuring the text (*begin by, focus on,*

(re)turn to). The author takes on this role throughout the whole text, but more or less explicitly depending on the priority of other roles. The second role is **the role of the researcher**, where the pronoun *I* is followed by verbs as *analyse, assume, consider, examine, find* and *study*. The third role is **the arguer role**, typically indicated by verbs *argue, dispute, reject* or *claim*. The last role is **the evaluator role** presented by verbal units as *feel, be content to, be sceptical about, be struck by, find something + evaluative adjective*. The results show that in NS academic writing, the researcher function is the most frequent and the arguer – the least frequent.

Different functions of the pronoun *I* in academic texts are distinguished by Hyland (2002a). Three of these functions are considered as high-risk functions: **elaborating an argument, stating results/claims** and, partly high-risk, **explaining a procedure**, and two low-risk functions: **stating a purpose** and **expressing self-benefits**. Professional writers typically use *I* in high-risk functions to present their own arguments and to show commitment to their arguments. Hong Kong L2 students, on the other hand, avoid responsibility and prefer not to use *I* in high-risk functions. Hyland's classification of functions does not directly correlate with the one by Fløttum *et.al.* (2006a). I think Fløttum *et.al.* (*ibid.*) takes a more general view over the functions of author in the text, whereas Hyland (2002a) considers in detail the role of author as a researcher. It is hardly possible to choose a more appropriate system for my own research, as it seems that these two classifications complement each other. However, since I focus on the functions of *I* in academic writing, Hyland's (*ibid.*) classification may better meet the requirements of my study.

#### 2.7.4 Functions of *we* in academic writing

The pronoun *we* in academic writing can be **exclusive, inclusive** or **generic**. According to Fløttum *et.al.* (2006a), in the exclusive use, *i.e.* where *we* refers to the author(s) of the article, the author roles are the same as discussed in Section 2.7.3 for *I* in academic writing:

- the writer role
- the researcher role
- the arguer role
- the evaluator role

The inclusive use, *i.e.* where *we* refers to both the writer and the reader, is not common with verbs presenting an idea or research: *we illustrate* or *we present*. The nature of these verbs implies that the reader is a receiver of the information. Inclusive use is, on the other hand, common in text-structuring function and in researcher role. According to Kuo (1999) the main aim of inclusive *we* is to seek agreement and cooperation. Hyland (2002a) studies the use of *I*

and *we* in academic writing and comments on why students opt for *we* instead of *I*: ‘Several students mentioned the collaboratively conducted research, which contributed to the individual reports, but underling many responses was a clear desire to reduce attributions to self’ (*ibid.*: 1108).

The generic function of *we* is included under inclusive by Harwood (2005). This function is least discussed by both Harwood (*ibid.*) and Fløttum *et.al.* (2006a) due to low authorial presence. Generic use gives least information about the writer, but rather stands for a large group of people.

To study *we* in academic writing of Russian learners, I, in addition to Fløttum *et.al.*'s (*ibid.*) classification, adopt the functions from Hyland's study (2002a) of author identity. The pronoun *we* is analysed based on the same classification as the pronoun *I* (*cf.* Section 2.7.3): **elaborating an argument, stating results/claims, explaining a procedure, stating a purpose and expressing self-benefits.**

## 2.8 Research objective

This chapter has provided an overview of previous research on author identity in argumentative and academic writing. Each study covers a certain research area. Some studies compare several learner varieties quantitatively or investigate one of them in different genres, whereas other studies choose to undertake a qualitative analysis, to focus on the influence of academic experience on the level of W/R visibility or to investigate the reasons for the deviations. Thus, each study looks at author identity from a new angle and elaborates the findings of the previous research.

The problem statement of this study involves investigating author identity in the learner writing of Russian students. The study consists of preliminary quantitative and functional qualitative analysis of first person pronouns in NS and learner writing. The findings from this study are compared to the results from previous research, which allows me to observe the correlation between the overuse/underuse of Russian learners and other learner groups. The functional analysis of first person pronouns builds on the classifications by Petch-Tyson (1998), Herriman (2009) and Hyland (2002a). A small-scale investigation of Slavic L2 writing contrasts Anglo-American and Slavic rhetoric. It is supposed to reveal if learners with Slavic L1 background encounter similar problems in English L2 writing. Pedagogical implications of the

study can lead to a more effective instruction of Russian/Slavic learners on how to construct author identity in English L2 writing.

### 3. METHOD AND MATERIAL

This chapter presents the methodology and the material I use to conduct my investigation. The methodology section explains what procedures and approaches are adopted in the study to obtain reliable results. The material section presents an overview of the corpora used in the investigation and explains the criteria used to collect the L1 texts.

#### 3.1 Methodological background

The present master thesis is a corpus-based study, which uses ‘corpus data in order to explore a theory or hypothesis [...] in order to validate it, refute it or refine it’ (McEnery and Hardie 2012: 6). The hypothesis of this study is that Russian learners portray their author identity differently from native speakers of English. For my investigation, I adopt a quantitative approach to corpora for identifying the overuse/underuse of author identity features. This allows me to highlight the key differences in the writing of Russian learners and native speakers, and, furthermore, work out pedagogical implications on how to achieve nativeness. The quantitative findings are elaborated further in Chapter 5, which presents the qualitative results of my study. Some parts of the research rely on a corpus-driven approach, where the Sinclairian ‘trust the text’ takes over from the automated search tools.<sup>4</sup> However, the main part of the investigation is based on the results extracted with the help of *Wordsmith Tools 5.0.0.334* and *AntConc 3.2.4w*. The former tool is applied for all research of the English language corpora, whereas the latter is used for investigating the Russian Cyrillic texts.

This study focuses on the analysis of the ‘I’ and the ‘we’ perspectives, namely the first person pronouns *I, me, my, mine* and *we, us, our, ours*. The possessive pronouns (*mine* and *ours*) and possessive determiners (*my* and *our*) are also considered to be a part of first person pronouns. I opt for this narrow subject for my investigation, and exclude other involvement features that are typically overused in English L2 writing. My initial aim was to study a set of involvement features such as disjuncts, emphatic particles, reference to the situation of writing/reading etc. However, a small-scale preliminary study uncovered the contrast of the individual Anglo-American rhetoric and the collective-minded Russian writing. This difference in author identity and its influence on L2 writing constitute a fruitful and interesting sphere for a contrastive analysis. Therefore, this study focuses primarily on first person pronouns and their role in constructing author identity.

---

<sup>4</sup> Corpus-driven is understood here in the terms of Tognini-Bonelli (2001): The main difference from a corpus-based lies in the ‘degree to which empirical data from a corpus is relied on’ (McEnery and Hardie 2012: 151).

The results of this master thesis exclude the instances of the pronouns that do not express author identity in the text. For example, it can be quotes or pronouns used due to the topic of the paper:

- (4) For example, it could be useful to compare frequency of *I/we* + shall , you + shall , etc. in Brown vs LOB. (VESPA-RU)
- (5) My mother answered: '*I* think it would be your own choice whether to make a military career or not'. (ICLE-RU)

Two of the corpora in this study – BAWE and VESPA – use a markup, which follows the guidelines of TEI.<sup>5</sup> It makes it possible to exclude direct quotes and examples that are not a part of student writing (Paquot, Hasslegård, Ebeling 2013: 379). Due to the markup of the elements that 'have been explicitly marked as such by the students' (*ibid.*), occasional mistakes are impossible to avoid. One of such cases is demonstrated in example (4) above. The use of the pronouns *I* and *we* is due to the topic of the assignment and does not make the writing more involved. The amount of such mistakes was rather significant, and I was afraid that they could influence the results of my study. Therefore, I manually analysed every hit of first person pronouns and excluded the cases that were not relevant for this study.

In the quantitative study, I analyse every first person pronoun in a separate section, whereas in Chapter 6 on the reasons for the variations, I divide the pronouns into two groups: the first person singular and the first person plural, namely the 'I' and the 'we' perspective. In my qualitative study, a functional analysis is carried out only for the pronouns *I* and *we*. The reason is that these pronouns express author identity more distinctly than other pronouns, and due to time limitations, I could not study all eight pronouns in both argumentative and academic writing. In the qualitative study (Chapter 5), the results of the functional analysis are based on a sample of up to 300 instances. Unfortunately, VESPA-RU and Russian L1 corpora do not contain enough hits, and, therefore, I analysed every available instance. The majority of English language corpora account for significantly more than 300 hits. To make the sample random, I extracted every fifth sample in ICLE-RU, every third sample in LOCNESS and every second sample in BAWE-LING.

In the vein of Paquot (2008), the present study relies on the *Explanatory Version* of the *Integrated Contrastive Model (ICM)* by Granger (1996). The model combines *Contrastive Analysis (CA)* and *Contrastive Interlanguage Analysis (CIA)*. The diagnostic approach from CIA

---

<sup>5</sup> More detailed information on the markup of the BAWE corpus is available at the official webpage of the University of Reading <http://www.reading.ac.uk/internal/appling/bawe/BAWE.documentation.pdf>



to CA reveals the L1-interlanguage performance similarities, and, therefore, displays the possible effect of L1 transfer. The approach is adopted in Sections 6.1.1, 6.1.2 and partly throughout Chapters 4 and 5 of this master thesis. To investigate this effect, the L1 Russian corpora are compiled and are discussed in Sections 3.2.5 and 3.2.6. However, the main focus of this study is to compare the native speaker and learner writing. For this purpose, I adopt CIA, which highlights the overuse/underuse and possible learners' errors. Another type of CIA is adopted in Chapter 7 to contrast the interlanguages of the same language, *i.e.* the English of Russian, Polish, Czech and Bulgarian learners. It places Russian learner writing in the perspective of learner writing of students with other Slavic L1 backgrounds.

A substantial part of this study relies on the frequency variation between the NS and learner corpora. Before interpreting these variations, I need to make sure that they are statistically significant. The evaluation of statistical significance is based on the test run in an online log-likelihood calculator.<sup>6</sup> In my study, the overuse is statistically significant if the result is at least 95% certain. For the level of  $p < 0.05$ , the critical value equals 3.84. The calculation of statistical significance is of critical importance for this study, as it allows me to be confident about my findings.

### **3.2 Material**

Since this is a corpus-based and partly corpus-driven investigation, the main material is the corpora. The English language corpora of argumentative and academic writing constitute the main part of the material. For my research, I use the following corpora: the *Louvain Corpus of Native English Essays* (LOCNESS), the *International Corpus of Learner English* (ICLE), the *British Academic Written English* (BAWE) and the *Varieties of English for Specific Purposes dAtabase* (VESPA). The above-mentioned corpora are well-compiled broad collections of student writing, which give many exciting possibilities for a researcher. However, to investigate the effect of L1 transfer, I need Russian language corpora that are comparable to the ones mentioned above. No such corpora are currently available in Russian. Thus, I had to collect argumentative and academic student writing in Russian by myself. The procedures connected to the collection of the material and the characteristics of the texts are presented in Sections 3.2.5 and 3.2.6.

---

<sup>6</sup> The log-likelihood calculator can be accessed at the webpage <http://ucrel.lancs.ac.uk/llwizard.html>

### 3.2.1 NS corpus of English argumentative writing

Many studies of learner writing (cf. Section 2.3.1: Petch-Tyson 1998; Paquot, Hasselgård and Ebeling 2013) use LOCNESS to investigate NS argumentative writing. The corpus contains 290 student essays and a total of 322,444 words. It consists of essays written by native speakers: British pupils' A-level essays, British and American university students' essays. The main topics of the corpus include *television, death penalty, computers, euthanasia, money is the root of all evil, etc.*<sup>7</sup> LOCNESS is usually used as a control corpus for the argumentative writing of the ICLE sub-corpora, and is directly comparable to it (Granger 1998).

A possible disadvantage of LOCNESS is its diversity.<sup>8</sup> The first factor is age. Though the majority of the essays are written by students aged between 18 and 20 years old, still a number of students are from 28 to 57 years old. Another key factor concerns the influence of task settings and intertextuality. Section 2.2 presents the findings by Ädel (2008), who argues that timing and access to secondary sources influence the use of first person pronouns. In this regard, LOCNESS has a significant variation. It includes timed, untimed and not rigidly timed (+/- 1 hour) essays. Essays differ in access to reference tools and the exam/not exam settings. Besides, the corpus consists of argumentative essays, literary essays and answers to exam questions. It makes it slightly less comparable to ICLE, which only contains the former category.

Notwithstanding its diversity, LOCNESS is definitely a valid control corpus for ICLE. In both corpora, the majority of the essays are argumentative, and they share similar topics. The authors are approximately of the same age and have similar level of education. Another advantage is that authors are partly British and partly American. It neutralizes the possible differences in the conventions in these countries.

### 3.2.2 NNS corpus of English argumentative writing

The Russian NNS argumentative writing is investigated on the basis of the Russian sub-corpus of the ICLE corpus, which is considered to be directly compatible to LOCNESS (Granger 1998: 13).<sup>9</sup> The sub-corpus consists of 274 essays and 224,356 words. The essays were written as a part of the English language courses by the learners with a higher intermediate to advanced knowledge of English. An example of the essay topics is presented below:

---

<sup>7</sup> More information on LOCNESS is available at the official webpage of the *Université Catholique de Louvain* <https://www.uclouvain.be/en-cecl-locness.html>

<sup>8</sup> A more detailed description of the topics is available at the webpage <http://www.corpus4u.org/forum/upload/forum/2005061503142048.pdf>

- The prison system is outdated. No civilized society should punish its criminals: it should rehabilitate them.
- Most university degrees are theoretical and do not prepare students for the real world. They are therefore of very little value.
- Crime does not pay.

In this study, the only criterion for defining the Russian learner group was the mother tongue of the students. I decided not to include the criterion of country, as many people with Russian L1 background live in the former Soviet Union Republics, outside the borders of Russia and use Russian in their everyday life. It turned out, however, that all but two essays were written within the framework of the studies at the Moscow Lomonosov State University.

To extract the material, I used the ICLE 2 program, which is very flexible and allows researchers to make a selection of texts based on a variety of specified features. One can choose the type of writing, gender, country, native language, language at home, years of English at school/university, time spent in English-speaking countries and other features. The main problem in ICLE-RU is that it is often difficult to estimate its homogeneity due to the high percentage of essays with unspecified criteria. For example, we can consider the criterion of timing. Approximately 82% of the essays were untimed, and only 4% were timed. This criterion is not specified for the remaining 14%, which makes it difficult to evaluate the influence of this factor on the results. If timed essays constitute only 4% of all writing, then the overall level of involvement is not significantly influenced. However, 18% of timed essays could potentially change the results towards a higher degree of involvement. Ädel's (2008) findings discussed in Section 2.3.2 reveal the influence of timing on the level of involvement. However, ICLE-RU is a quite small corpus, and it is not preferable to reduce its size. Therefore, I decided not to exclude any of the untimed essays. Besides, LOCNESS is not homogeneous in this respect either. Unfortunately, I did not find the exact proportion of the timed essays in the NS corpus, but they seem to be a minority.

Students in ICLE-RU have a varied experience in learning English at school and at the university. In total, students spent between 3 and 12 years learning the language. Most students (73%) have never been to the English-speaking countries, 23% spent less than 6 months in the native speaker environment, and only 4% lived in English-speaking countries for over half a year.

However, the main advantage of ICLE-RU is that it can be considered representative of the Russian learner group. All learners are students with at least one year of studying English at

the university level. The corpus is rather homogeneous with regards to age: less than 9% of students are over 25 years old. All essays in ICLE-RU are argumentative and range in length from 200 up to 1800 words, and over half of the essays (53%) are within the range of 500-1000 words. According to Ädel (2008), access to secondary sources is another important criterion. The majority of students in ICLE-RU (84%) used some reference tools, whereas 12% did not have access to secondary sources, and for 4% of students this criterion is not specified.

In addition to ICLE-RU, this study refers to three more ICLE components: Bulgarian, Czech and Polish. A short description of these sub-corpora is given in Section 3.2.7 called ‘Other corpora’.

### 3.2.3 NS corpus of English academic writing

The *British Academic Written English* (BAWE) is a corpus of discipline-specific student academic writing.<sup>10</sup> The corpus consists of 3000 good-standard student assignments, which total 4,534,873 words.<sup>11</sup> The assignments represent 35 disciplines, 13 broad genre groups (essay, exercise, critique and others) and four levels of study (each of the three undergraduate years and master students’ assignments).<sup>12</sup> The corpus was collected between 2004 and 2007 at the universities of Warwick, Reading and Oxford Brookes. The present study only makes use of the linguistics part of the corpus (BAWE-LING) in order to make it most comparable to VESPA-RU. BAWE-LING is rather small and consists of 75 essays or 197,291 words<sup>13</sup>. Some assignments in BAWE-LING belong to students with Greek, German, Japanese and other L1 backgrounds. Including these assignments in the material would make my findings on English L1 writing less reliable. Therefore, using the BAWE.xls file, I identified and excluded all linguistic assignments written by students with non-English L1 background.

BAWE-LING is not totally homogeneous, as can be observed when analysing age and gender groups or the length of the assignments. Linguistic assignments in BAWE range in length from 500 up to 9,000 words, and an average length is 2,300 words. The majority of the texts (58%) are between 1,500 and 2,500 words. As for gender, female contributors are prevailing and constitute 86%, whereas male authors account for 14%. Linguistics, as well as humanities in general, is a field where gender imbalance is common. I hope this will not influence my study,

---

<sup>10</sup> More information on BAWE is available at the official webpage of the *Coventry University* <http://www.coventry.ac.uk/research/research-directory/art-design/british-academic-written-english-corpus-bawe/>

<sup>11</sup> This figure only includes the English L1 writing. BAWE incorporates the writing of students with non-English L1 backgrounds and totals 6,506,995 words

<sup>12</sup> More detailed information on genres and disciplines is available at the official webpage of the *University of Reading* <http://www.reading.ac.uk/internal/appling/bawe/BAWE.documentation.pdf>

<sup>13</sup> My study does not include the file 6189a – a report in the field of Applied Linguistics

because I expect to find similar gender distribution in VESPA-RU. The analysis of the age distribution shows that BAWE-LING contributors are rather young: 59% are between 18 and 21 year old; 15% are aged from 22 up to 27 and 24% are between 37 and 51 years old. The gap of ten years between the second and the third group is not accidental, but due to the absence of contributors of that age.

However, BAWE-LING fulfils the main requirement – it is directly comparable to VESPA-RU. Both corpora consist of discipline-specific linguistic texts written by students with various levels of academic experience. In addition, BAWE-LING is a trustworthy source of NS academic writing: this concerns the size and the quality of the material. The corpus is large enough to produce reliable results. The proficient level of writing lays a sound foundation for identifying the standard of NS student academic writing.

### 3.2.4 NNS corpus of English academic writing

Russian learner writing is studied based on VESPA-RU, the Russian sub-corpus of the *Varieties of English for Specific Purposes dAtabase* (VESPA)<sup>14</sup>. The VESPA corpus includes the writing of BA and MA students with 32 different L1 backgrounds. The texts were collected in Norway in the three higher educational establishments: *The University of Oslo* (humanities), *Høgskolen i Østfold* (linguistics and language studies), and *Handelshøyskolen BI* (business and management). As the texts were collected in Norway, a great many of the contributors has Norwegian L1 background. Fortunately for me, the number of Russian learners contributing to the corpus was high enough for me to be able to conduct my study. The sub-corpus includes 21 assignments and totals 28,651 words. Two papers by Russian students were excluded from the sub-corpus, because the students identified their mother tongue as sign language. Since I am not sure if this factor can influence the results of my investigation, and I do not have a chance to contact the contributors for more information, I decided not to include these texts into the analysis. The assignments include papers and reports in the field of linguistics and English languages studies. The majority of the writing was collected from master students, and only three assignments were written by bachelor students.

The main disadvantage of this sub-corpus is undoubtedly its small size. However, it remains the best available collection of English L2 academic writing by Russian learners. Most importantly, the corpus is large enough to obtain statistically significant results. In my study, I use relative frequencies per 10,000 words and test every overuse/underuse for statistical significance in the log-likelihood test. Another important factor is difficult to call a disadvantage, as it is rather a characteristic feature of this corpus. VESPA-RU contributors do not live in the country of their L1.

---

<sup>14</sup> More information on VESPA is available at the official webpage of the *Université Catholique de Louvain* <https://www.uclouvain.be/en-cecl-vespa.html>

As mentioned above, the texts were collected in Norway, and over half of the contributors named either English or Norwegian as the second language they use at home. Russian was of course put as their first choice. It remains unknown how many months the Russian learners have spent abroad at the time of data collection. Potentially, they may be permanent residents of Norway.

To evaluate how comparable VESPA-RU is to its NS counterpart, I investigated a number of factors as age, gender and the length of the assignments. The gender distribution is similar to the other corpora: male contributors constitute the minority. The overall majority of the Russian contributors are female, which is an advantage due to the similar gender distribution in BAWE-LING. Considering the age groups, VESPA-RU and BAWE-LING are slightly less comparable. The total of 72% of Russian learners are aged between 22 and 29 years old, and the remaining 28% were from 36 up to 43 years old. I think that the reason for such age distribution is that many Russian students take extra courses to recognize their education from Russian universities or to acquire further qualification from a Norwegian university. My assumption is supported by the amount of years the contributors have studied at university level: over half of the learners have over five years of university education. The length of the assignments is not specified for VESPA-RU. Essay is the only genre presented in the corpus. In BAWE-LING, essay is the dominant genre, but not the only one.

However, VESPA-RU is the first step on the way to acquire comprehensive data on English L2 academic writing by Russian and other Slavic learners. If Russian learner writing could previously only be investigated based on the ICLE-RU corpus, now, it is possible to study a different genre – academic writing. In addition, the study based on the Russian sub-corpus of VESPA supplements the data collected from other components of this corpus. In this way, it will help drawing conclusion on the overall English writing across learner varieties.

In Chapter 7 of this study, I also refer to two more components of VESPA: Czech and Polish. A short description of these sub-corpora is presented in Section 3.2.7.

### **3.2.5 NS corpus of Russian argumentative writing**

Russian L1 writing is important for this study in order to estimate the effect of L1 transfer. Unfortunately, none of the existing corpora of Russian can be considered comparable to ICLE and LOCNESS. One of the corpora I considered using was the *Russian National Corpus* (RNC). It is a 300-million-word collection of texts, which consists of journalistic, religious, business and scientific texts, fiction, letters and diaries. Notwithstanding its large size, the corpus does not

contain the material required for my study: student argumentative writing. Russian linguists often face similar challenges, and to respond to them, researchers have to address the Internet.

Therefore, I searched the Internet for a reliable source of student argumentative writing in Russian. The texts of the appropriate genre were collected from the webpage <http://ucheba.pro/>.<sup>15</sup> This is a resource where high-school students share the argumentative essays they wrote in preparation for the final exams in the Russian language. Students publish their essays in order for experts, sensors or other students to give their opinion on the essay's style, argumentation techniques or other possible mistakes. In the present study, this collection of texts is referred to as the *Collection of RUssian Argumentative Writing* (CRUAW).

My collection of texts consists of 71 argumentative essays and 103,911 words. The topics of the essays are similar in its nature to those in ICLE-RU and LOCNESS:

- What is more important: the beauty of the soul or the body?<sup>16</sup>
- Tolerance towards other people.
- The strength of a kind word.
- What is nobility?

My aim was to compile a balanced and homogenous collection of texts. However, it is not always easy. The original plan was to choose equally many essays from each gender, but it turned out to be impossible due to the dominance of female authors. In the end, the distribution was similar to that in ICLE-RU: 82% female and 18% male contributors.<sup>17</sup> The next important factor is age distribution. Unfortunately, this criterion is not specified for each of the writers. Nevertheless, it is likely that practically all authors were around 16-17 years old when submitting their essays, as these texts were written in preparation for the final school graduation exam in Russian. This exam is obligatory for all students planning to study further at the university. The writers in ICLE-RU are already university students, and, therefore, are some years older and have slightly more academic experience than their counterparts. I believe that the two text collections can still be considered comparable, since the age difference is not considerable. The essays are rather homogeneous with regards to length: they all vary between 200 and 500 words. The main reason for this is the requirement of the exam. In order to receive the top grade, the essay should be from 150 up to 300 words. Some essays in ICLE-RU have the same length, though the majority is somewhat longer. Another important factor is timing and access to

---

<sup>15</sup> Accessed on the 5<sup>th</sup> of December 2013.

<sup>16</sup> My translation from Russian.

<sup>17</sup> The gender distribution in ICLE-RU is as following: 84% female and 16% male contributors.

reference tools. It is hardly possible to estimate the influence of these factors, as I have no information from the writers. However, from the comments accompanying the text, I assume that the writers did not use reference tools, and the time was not rigidly limited. Most writers tried to recreate the conditions of a real exam and followed a time-schedule, but a few mentioned that they went slightly over the allowed limits.

The essays in CRUAW are written for training purposes and not for evaluation. It is possible that authors did not put all their efforts and thinking into these pieces of writing. However, students published their works online in order to receive comments and feedback. Therefore, I think that the essays are well composed and are representative of Russian L1 student argumentative writing.

Though CRUAW is not free from disadvantages, it still satisfies the main criteria of a corpus, which is comparable to ICLE-RU and LOCNESS. It is a collection of argumentative writing from non-professional writers. It has a significant size and is authentic. The writing serves educational purposes and represents real language.

### **3.2.6 NS corpus of Russian academic writing**

As discussed in the previous section, RNC is a large corpus that does not contain student writing. Academic writing is not presented there either. Fossan (2011) collected texts online through the BIBSYS search interface. Unfortunately, in Russia, no database of bachelor and master thesis is available through the Internet. The L1 academic assignments used for this study were kindly granted to me by friends, colleagues and acquaintances from Saint-Petersburg, Russia. Therefore, I have all background information about the people, who contributed to my moderate collection; and I was able to compile it to the best of my ability. In the present study, the texts are referred to as the *Collection of RUssian Master and Bachelor Assignments* (CRUMBA).

The academic works were written between 2008 and 2012, which means they fit rather perfectly in timing to the texts from VESPA-RU, as they also were composed around 2008. The collection includes 11 assignments and 162,905 words. Five works (61,408 words) are by bachelor students, and six works (101,497 words) are by master students. CRUMBA includes writing in such disciplines as linguistics, economics, business, physics and psychology. The linguistic component of CRUMBA (CRUMBA-LING) consists of 21,582 words and is used in Section 5.4 to analyse the functions of the pronoun *мы* (=we) in Russian L1 academic writing. CRUMBA-LING is most comparable to BAWE-LING and VESPA-RU.



To see if my collection of Russian L1 writing can be considered comparable to the English academic corpora, I observe the correlation of various factors between these texts. The students who contributed to my collection were from 19 up to 22 years old when they wrote the assignments. This makes my collection most comparable to BAWE-LING and slightly less comparable to VESPA-RU. The gender distribution in CRUMBA is similar to that of the English corpora. In Russian L1 writing, 64% of writers are female and 36% are male. Here, the proportion of male writers is higher than in BAWE-LING and VESPA-RU. This is not surprising, as in linguistic departments, as well as in other humanity departments, female students constitute a majority. An average length of a bachelor assignment is about 12,000 words, whereas the length of a master assignment is over 16,000 words. The entire texts of the assignments are included into my investigation, because Hyland's study (2002a) shows that author identity can be more or less visible in various parts of an academic study.

A great advantage of this collection, is that all texts were written as a part of an educational process. Students received marks for their works, which means they had put a lot of time and efforts into writing them. Such assignments represent real academic language and are a reliable material to base my investigation on.

### **3.2.7 Other corpora**

In addition to the main corpora described above, some sections of the present master thesis refer to the results from other corpora. This section gives a brief account of these corpora.

#### ***The British National Corpus (BNC)*<sup>18</sup>**

The BNC is a 100 million word collection of a written (90%) and spoken (10%) language from a wide variety of sources. The written part of the corpus consists of samples from newspapers, journals, academic writing, fiction and many other types of texts. This study addresses the academic part of the BNC to analyse the use of first person pronouns in the professional writing of NS writers. To make the results more reliable, authors' domicile is set to the UK and Ireland, and the USA. Since the results from the BNC are compared to BAWE-LING and VESPA-RU, discipline becomes one of the key criteria and is limited to humanities and arts. The texts that

---

<sup>18</sup> More information on the BNC is available at the official webpage of the corpus at the Oxford University <http://www.natcorp.ox.ac.uk/>

were written to be spoken are excluded from the search due to its interactive nature, which can result in a higher use of first person pronouns.

The spoken part of the BNC consists of the transcriptions of informal conversations and spoken language collected in various contexts: from official meetings to TV shows. The results from the spoken component of the BNC complement the findings from other corpora presented in Figure 6 of the present master thesis. It illustrates that speech has a highly involved style and that first person pronouns are more frequent here than in writing.

### **Other components of ICLE<sup>19</sup>**

In Chapter 7 of this master thesis, Russian learners' argumentative writing is compared to the argumentative writing of other Slavic learners. The aim is to identify if there are any similarities in the use of author identity between Slavic learner groups. Therefore, the use of first person pronouns is analysed in the Czech, Polish and Bulgarian components of ICLE. Similarly to ICLE-RU, I extract the texts from the ICLE2 program based on the criterion of the learners' native language.

**ICLE-Polish (ICLE-PL)** consists of 366 essays and 234,789 words. The overwhelming majority of learners live in Poland, and only three students reside in other countries. The essays range in length between 500 and 1,000 words. Most students had access to reference tools. Slightly over 60% of all essays were not timed, 29% were timed, and for 1% of the essays, the criterion is not specified.

**ICLE-Bulgarian (ICLE-BL)** is a collection of 300 texts and 199,249 words. All students live in the country of their mother tongue. The shortest essay is 216 words, whereas the longest reaches 2,400 words. However, over 56% of texts are between 500 and 1,000 words. None of the essays was timed, and 91% of the students had no access to reference tools in the process of writing.

**ICLE-Czech (ICLE-CZ)** includes 241 essays and 200,727 words. All learners reside on the territory of the Czech Republic. The length of the essays varies between around 350 and 1,400 words. Over 78% of the students had access to reference tools. About 98% of the essays were written with no time restrictions, and in 2% of the cases, the criterion is not specified.

---

<sup>19</sup> More information on ICLE is available at the official webpage of the *Université Catholique de Louvain* <http://www.uclouvain.be/en-cecl-icle.html>

## **Other components of VESPA<sup>20</sup>**

As discussed above, Chapter 7 presents a comparison of English L2 writing by different Slavic learners. Besides Russian, VESPA contains only two sub-corpora of learner writing with Slavic background: Polish and Czech. Both text collections are rather small.

**VESPA-Polish (VESPA-PL)** consists of four assignments and 5,033 words. All four writers are females and speak Polish at home. Students did not have time limitations and could use reference tools. All contributors are master students who differed quite a lot with respect to their age, years spent learning English and time spent in an English-speaking country. Therefore, it is possible to assume that students' command of English was rather diverse. This could, of course, influence the results of my investigation. However, the material was so limited that I did not dare to cut it even more. Further study would be required to confirm or reject the modest observation made on the basis of VESPA-PL and presented in Chapter 7 of my master thesis.

**VESPA-Czech (VESPA-CZ)** is a collection of five texts totalling 8,276 words. The texts were composed by two males and three females aged between 23 and 25 years old. All learners are master students and have studied English for over five years at the university level. As in VESPA-PL, the tasks were not timed, and students had free access to reference tools.

The main disadvantage of these corpora is their small size. Therefore, the results presented in Section 7.2 need to be treated with caution. It is probably more correct to call them observations, which require further investigation. I still want to underline the obvious advantages of the corpora: they are both comparable to VESPA-RU in its genre and academic experience.

## **Collection of English L2 Argumentative Writing (CELAW)**

CELAW consists of 23 essays and 5,934 words written by presumably advanced learners of English, with Russian as their mother tongue. The texts were collected in order to represent the L2 writing of Russian learners between 2010 and 2013. The material is used to evaluate the development of the 'we' perspective, which is discussed in Section 6.1.1. The argumentative essays were collected on the Internet resource [www.efl.ru](http://www.efl.ru), where many Russian students meet to discuss their preparation for various international English language tests, such as IELTS

---

<sup>20</sup> More information on VESPA is available at the official webpage of the *Université Catholique de Louvain* <https://www.uclouvain.be/en-cecl-vespa.html>

(*International English Language Testing System*), TOEFL (*Test of English as a Foreign Language*), CAE (*Certificate in advanced English*) and others.<sup>21</sup> In preparation for the written part of the exam, some students post their essays and ask for evaluation and remarks from other members of the forum. The main advantage of this collection is that it presents authentic writing of Russian students. The disadvantage is that there is little information about the contributors, as the material is collected on the Internet.

### **3.3 Concluding remarks on method and material**

The main aim of this section is to give a concise and structured overview of the main corpora and text collection that provide the basis for my investigation. Table 6 (see below) summarizes and complements the information on the material used in this study.

---

<sup>21</sup> Accessed on the 10<sup>th</sup> of December 2013.

**Table 6.** A concise overview of the key characteristic factors of the main corpora

Corpora	LOCNESS	ICLE-RU	CRUAW	BAWE-LING	VESPA-RU	CRUMBA
<b>Language of the texts</b>	British/American English	English	Russian	British English	English	Russian
<b>Native language</b>	British/American English	Russian	Russian	British English	Russian	Russian
<b>Number of words</b>	322,444	224,356	103,911	197,291	28,651	162,905
<b>Number of essays</b>	290	274	71	75	21	11
<b>Genre</b>	Argumentative	Argumentative	Argumentative	Academic (linguistics)	Academic (linguistics)	Academic (linguistics, economics, business, physics, psychology)
<b>Length of the texts (words)</b>	About 200 - 900	53% - 500 - 1,000 Range: 200 - 1,800	Range: 200 - 500	58% - 1,500 - 2,500 Range: 500 - 9,000	-	BA: 12,000 MA: 16,000
<b>Age (years old)</b>	Mainly: 18 - 20 Occasionally: 28 - 57	91% - under 25 9% - over 25	appx. 100%: 16 - 17	59% - 18 - 21 15% - 22 - 27 24% - 37 - 51	72% - 22-29 28% - 36 - 43	19 - 22
<b>Gender</b>	-	85% - female 15% - male	82% - female 18% - male	86% - female 14% - male	100% - female	64% - female 36% - male
<b>Reference tools</b>	Varied	84% - used reference tools 12% - no reference tools 4% - unknown	No reference tools	-	Free access to reference tools	Free access to reference tools
<b>Timing</b>	Varied	82% - untimed 4% - timed 14% - unknown	Not rigidly timed	-	No timing	No timing
<b>Years of learning English (total in years)</b>	-	3 - 12	-	-	6 - 17	-

#### **4. RESULTS AND ANALYSIS: QUANTITATIVE STUDY OF AUTHOR IDENTITY IN ARGUMENTATIVE AND ACADEMIC WRITING**

The present chapter introduces the quantitative results of my investigation. It incorporates two sections that present a detailed analysis of author identity features in argumentative and academic writing. The results for argumentative writing are based on the investigation of LOCNESS and ICLE-RU, which represent the writing of native speakers and learners respectively. BAWE-LING and VESPA-RU are used to conduct the investigation of academic writing. Further, the results from these sections are compared to Russian NS writing to identify the possible influence of learners' L1 on their performance in the interlanguage.

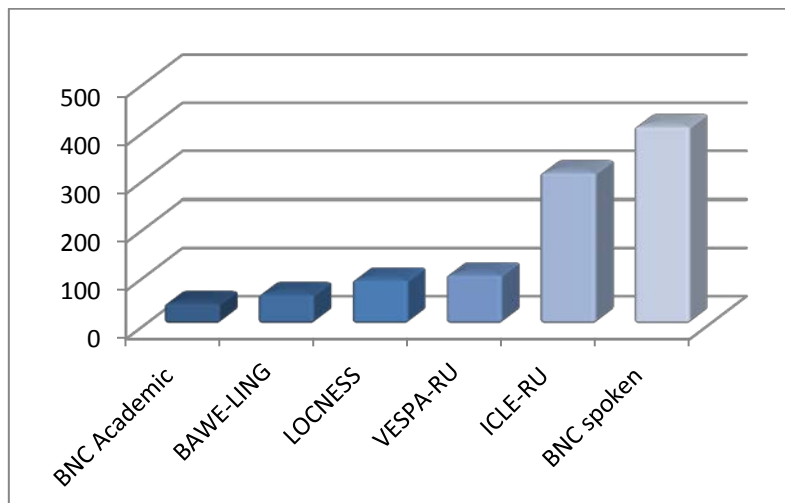
The two sections consist of sub-sections, which discuss each of first person pronouns in detail to identify how they contribute to the construction of author identity in English NS and Russian NNS writing. The pronouns are analysed separately in order to highlight the input each of them makes into constructing author identity. To account for the overuse/underuse, I analyse concordance lines for each pronoun. Concordances data give a more profound insight into the possible reasons for the deviation through analysing multi-word units, which are overused/underused at a high rate. This investigation covers the following pronouns:

**First person singular: *I, me, my, mine***

**First person plural: *we, us, our, ours***

To acquire a general understanding of the level of W/R visibility in the writing of Russian learners, I undertake a study similar to the one by Gilquin and Paquot (2008) (*cf.* Section 2.4.4). The use of first person pronouns is investigated in student and professional NS writing, learner discourse and speech. Figure 6 reveals that the overuse is more marked in learner writing than in NS professional and novice student writing, but less marked than in speech.

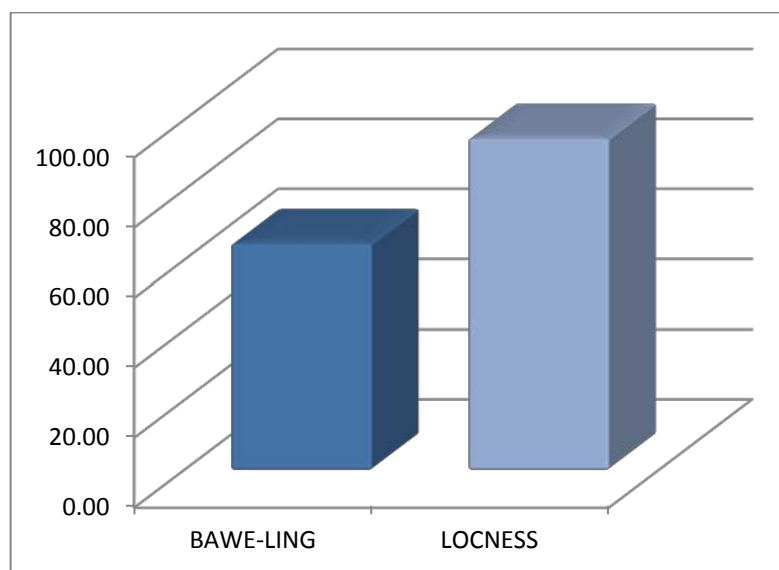
**Figure 6.** The use of first person pronouns across the corpora. Frequency per 10,000 words



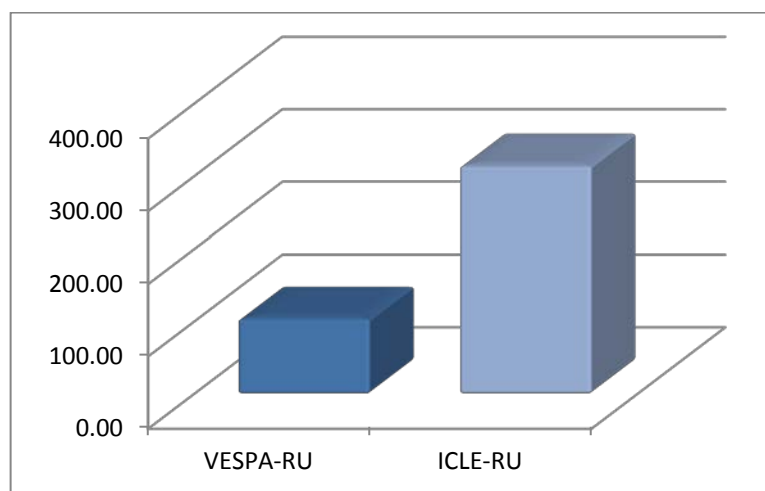
Since I study only one involvement feature, it would not be correct to discuss the overall level of involvement in the writing of Russian learners. However, the use of first person pronouns is probably the most direct and explicit expression of authorial presence in the text, which probably indicates the major tendency (Hyland 2002a).

According to Figures 7 and 8, argumentative writing makes a greater use of first person pronouns than academic writing, which is true for both NS and learner writing. The decrease in the use of the pronouns is rather sharp in learner writing, whereas in NS texts, it is more moderate.

**Figure 7.** The use of first person pronouns in NS argumentative and academic writing. Frequency per 10,000 words



**Figure 8.** The use of first person pronouns in NNS argumentative and academic writing. Frequency per 10,000 words



#### **4.1 Author identity in argumentative writing**

This section presents a comparison of first person pronouns in the argumentative writing of native speakers and learners. The results are presented in tables and figures, which contain both raw numbers and frequencies per 10,000 words. Fossan's work (2011) inspired me to include the relative difference between NS and learner results to indicate the overuse or underuse. Relative overuse/underuse is calculated by dividing the frequency per 10,000 in ICLE-RU by that of LOCNESS. A value above 1 indicates overuse. The purpose of such a comparison is not to evaluate the learners' performance, but to help Russian students realise their author identity in writing in a more effective way. At the same time, this study sheds some light on the cultural practices and academic conventions of the learner group. This provides the background for a better understanding of Russian learner writing among other learner groups and native speakers.

To make sure the discovered results are significant, I performed a log-likelihood test. It allows to check if the overuse/underuse is statistically reliable, or not reliable due to the low raw frequencies or relatively minor differences between the corpora.

At the end of the Sections 4.1 and 4.2, the results of my study are compared to the findings from other investigations. It could be interesting to compare the results for each individual form, but it is impossible, as other studies do not present the results for *I*, *me*, *my*, *mine* and *we*, *us*, *our*, *ours* separately.



#### 4.1.1 The first person singular pronoun *I*

The pronoun *I* is the most explicit way of making the author visible in a text. Table 7 shows that Russian learners overuse the pronoun *I* at a rate of 2.11. The log-likelihood test confirms the statistical significance of the overuse.<sup>22</sup> As discussed in Section 2.3.2, the overt presence of the author in the text is a characteristic feature of learner varieties.

**Table 7.** The pronoun *I* in LOCNESS and ICLE-RU. Frequency per 10,000 words

Feature	LOCNESS		ICLE-RU		ICLE-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
The pronoun <i>I</i>	982	30.45	1445	64.41	2.11

To obtain a more detailed picture, I analyse the clusters with the pronoun *I* in each corpus. The cluster size varies from 2- to 4-grams and the minimum frequency is 10 instances. I single out the clusters that contribute to expressing personal opinion or that take part in constructing author identity in the text. The results are sorted according to how frequent the clusters are in these corpora (*cf.* Table 8):

**Table 8.** Recurrent word-combinations in LOCNESS and ICLE-RU (search for 2- to 4-grams involving *I*)

LOCNESS	ICLE-RU
<i>I think</i>	<i>I think</i>
<i>I feel</i>	<i>I would like to/I'd like to</i>
<i>I believe</i>	<i>I believe</i>
<i>I know</i>	<i>I mean</i>
<i>I would like to/I'd like to</i>	<i>I suppose</i>
<i>I agree</i>	<i>I want</i>
<i>I am sure/I'm sure</i>	<i>I am sure/I'm sure</i>
	<i>I know</i>
	<i>I do not think/ I don't think</i>
	<i>I wonder</i>
	<i>I don't want</i>
	<i>I hope</i>
	<i>As for me, I</i>

Appendix 1 presents the raw and relative frequencies and the ICLE-RU overuse rate for each of the multi-word units in Table 8.

<sup>22</sup> Statistical significance – 99.99%.

Table 8 above indicates that NS writers follow the idiom principle and use the pronoun *I* in a limited number of prefabs (Erman and Warren 2000). Learners, on the other hand, adhere to the open choice principle with a tendency to strongly overuse one or two multi-word units – the so-called lexical teddy bears (Hasselgren 1994). What strikes the eye is the strong overuse of the units *I would like to* and *I think* in Russian learner writing (cf. Table 9). These two phrases amount to over half of all the clusters with a minimum frequency of 10 instances in ICLE-RU (cf. Table 47 in Appendix 1).

**Table 9.** The distribution of *I think*, *I feel*, *I believe* and *I would like to/I'd like to* in LOCNESS and ICLE-RU. Frequency per 10,000 words

Feature	LOCNESS	ICLE-RU	ICLE-RU overuse
<i>I think</i>	2.70	12.08	4.47
<i>I feel</i>	2.67	0.27	0.10
<i>I believe</i>	1.67	2.32	1.39
<i>I would like to/I'd like to</i>	0.37/-	3.17/1.52	8.57/-

The unit *I think* in ICLE-RU is about three times as frequent as it is in LOCNESS.<sup>23</sup> NS writers do not rely solely on the unit *I think*, but make use of three alternatives: *I think* (87 instances), *I feel* (86 instances) and *I believe* (54 instances). In learners' writing, *I believe*<sup>24</sup> is far less frequent than *I think*, and *I feel*<sup>25</sup> is nearly absent.

Russian learners often express their author identity without making a direct statement, but through using a negation, as for example in the patterns of subjective stance with *I do not/ I don't*:

- (6) *I don't quite agree* with the sentence. (ICLE-RU)
- (7) *I don't know* who sees any good in dominant, aggressive women and weak, tamed and vulnerable men. (ICLE-RU)
- (8) But *I don't mean* that it is very bad in fact. (ICLE-RU)
- (9) Of course *I do not mean* to say that every advertisement is useless or even harmful. (ICLE-RU)

<sup>23</sup> The frequency of 12.08 (ICLE-RU) against 2.70 (LOCNESS) per 10,000 words.

<sup>24</sup> The frequency of 2.32 per 10,000 words.

<sup>25</sup> The frequency of 0.27 per 10,000 words.

Such patterns are significantly less frequent than in LOCNESS. The log-likelihood test confirms the statistical significance of the overuse in the table below:<sup>26</sup>

**Table 10.** Patterns of subjective stance in ICLE-RU and LOCNESS. Frequency per 10,000 words

Feature	LOCNESS		ICLE-RU		ICLE-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
Patterns of subjective stance with <i>I do not/ I don't</i>	27	0.84	99	4.41	5.27

This can be caused by the influence of L1 or be a characteristic feature of learner writing in general. Fossan's results (2011) report the overuse of such patterns of subjective stance as *I don't think* and *I'm not saying* for Norwegian learners.

Russian learner writing is characterized by a high frequency of contractions, which contributes to the overall spoken-like nature of the writing. In comparison, NS authors follow the conventions of the written language and avoid using contractions.

#### 4.1.2 The first person singular pronoun *ME*

The study indicates that the pronoun *me* is not highly frequent in either corpus. The overuse is slightly higher than what is registered for the pronoun *I*. The log-likelihood test reports that the overuse is statistically significant with a high degree of certainty (*cf.* Table 11).<sup>27</sup>

**Table 11.** The pronoun *I* in LOCNESS and ICLE-RU. Frequency per 10,000 words

Feature	LOCNESS		ICLE-RU		ICLE-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
The pronoun <i>me</i>	94	2.92	175	5.43	2.68

The analysis of the top fifteen 3- and 4-grams with the pronoun *me* in LOCNESS reveals the most frequent concordance lines for the pronoun (*cf.* Table 12):

<sup>26</sup> Statistical significance – 99.99%.

<sup>27</sup> Statistical significance – 99.99%.

**Table 12.** The top fifteen 3- and 4-gram clusters with the pronoun *me* in LOCNESS. Frequency per 10,000 words

Feature	LOCNESS	
	Raw frequencies	Per 10,000 words
<i>it seems to me</i>	6	0.08
<i>this/that leads me to</i>	3	0.09
<i>told me that</i>	3	0.09
<i>important to me</i>	2	0.06
<i>is beyond me</i>	2	0.06
<i>obvious to me</i>	2	0.06

The frequency of the units is low due to the overall infrequency of the pronoun *me* in LOCNESS. Despite this, Table 12 indicates six native-like usage patterns of the pronoun. In ICLE-RU, the analysis of the top fifteen 3- and 4-gram clusters reveals that only one unit is native-like – *it seems to me*. With a frequency of 1.25 per 10,000 words, the overuse rate of the unit reaches as high as 15.09. The log-likelihood test confirms the statistical significance of the result.<sup>28</sup>

**Table 13.** The top fifteen 3- and 4-gram clusters with the pronoun *me* in ICLE-RU. Frequency per 10,000 words

Feature	ICLE-RU		ICLE-RU overuse
	Raw frequencies	Per 10,000 words	
<i>it seems to me</i>	28	1.25	15.09
<i>as for me</i>	22	0.98	-

*As for me* is another frequent cluster in L2 argumentative writing of Russian learners. The unit *as for me* is absent in LOCNESS, and has an extremely low frequency of 0.0048 instances per 10,000 words in the BNC. The only domains the unit appears in are fiction and verses, and newspapers. The use of *as for me* in argumentative writing is not native-like and may be prompted by the effect of L1 transfer. I investigated other learner varieties to see if any of them make use of the multi-word unit (*cf.* Figure 9):

<sup>28</sup> Statistical significance – 99.99%.

**Figure 9.** The sequence *as for me* across corpora per 10,000 words

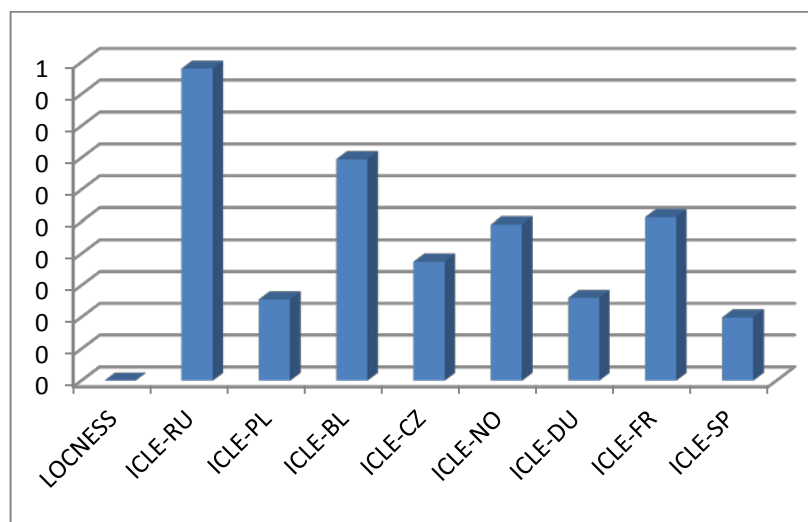


Figure 9 demonstrates that *as for me* is rather rare in other interlanguages and is specific to Russian learners, where the sequence *as for me* is repeatedly followed by the pronoun *I*. The frame is not incorrect, but definitely highly infrequent in NS writing and speech, as well as in other learner varieties. Besides, learners with other L1 backgrounds do not tend to use the pronoun *I* directly after the sequence *as for me*:

- (10) *As for me, I* am sure of it. (ICLE-RU)
- (11) *As for me, I* think that in general people are now becoming eco-friendly. (ICLE-RU)
- (12) *As for me, I* sympathise with most slogans of feminism. (ICLE-RU)

Only Bulgarian students seem to use the multi-word unit in the same way as Russian learners:

- (13) They would tell us about how to cope with everyday problems, *as for me*, in the classroom. (ICLE-NO)
- (14) *As for me* this situation was intolerable, I decided to leave home and to live on my own. (ICLE-GE)
- (15) These houses are very ugly and *as for me* very hopeless. (ICLE-CH)
- (16) *As for me, I* am quite willing to believe that everyone lives ‘in their heads’. (ICLE-BU)

The overuse is likely to result from L1 transfer as the multi-word unit has a congruent counterpart in Russian, *i.e.* *что касается меня, то я* (= *as for me, I*), which is common in argumentative writing. The examples below are extracted from the *Russian National Corpus* (RNC), a 300-million-word corpus of various texts by Russians:

- (17) **Что касается меня, то я** выступал на сцене дважды. (RNC)  
 [=As for me, I have appeared before the footlights twice in my life.]<sup>29</sup>
- (18) **Что касается меня, то я** сплю долго и крепко. (RNC)  
 [=As for me, I sleep long and deep.]
- (19) **Что касается меня, то я** лучше выпью двойную порцию чая с малиной. (RNC)  
 [=As for me, I would prefer two cups of tea with raspberry jam.]

The transfer of form may go together with the transfer of function, frequency and register. *As for me, I* repeats the stylistic profile of the Russian *что касается меня, то я* (=as for me, I) and repeatedly appears in Russian EFL learner argumentative writing. The L1 sequence is typically used in spoken or informal written language. However, it is not appropriate in formal writing. A search in VESPA-RU does not reveal a single instance of *as for me, I*. This could result from the transfer of register, but could also show a more advanced command of English among learners writing in the academic domain.

#### 4.1.3 The first person singular pronoun *MY*<sup>30</sup>

The determiner *my* is frequent in both corpora, but is over-represented in the learner writing (cf. Table 14). According to the log-likelihood test, the registered overuse is statistically significant.<sup>31</sup>

**Table 14.** The pronoun *my* in LOCNESS and ICLE-RU. Frequency per 10.000 words

Feature	LOCNESS		ICLE-RU		ICLE-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
The pronoun <i>my</i>	260	8.06	280	12.48	1.55

Similarly for both corpora, the determiner *my* collocates with nouns characteristic of argumentative writing, such as *opinion, mind, view, point of view, essay, argument, belief* and others. Together with these nouns, the pronoun forms a number of sequences, some of which are native-like, whereas others are not. *To my mind* is the most frequently used sequence with the pronoun *my* in ICLE-RU.<sup>32</sup> The unit is native-like and has a frequency of 2.38 per million words in the BNC. However, a more common unit to use would be *in my mind* or, even better, *in my*

<sup>29</sup> These glosses, as well as all further glosses from Russian into English, are provided by me.

<sup>30</sup> This study considers possessive determiners as a part of first person pronouns.

<sup>31</sup> Statistical significance – 99.99%.

<sup>32</sup> Frequency of 1.34 per 10,000 words

*opinion*. According to the BNC, the former has a frequency of 4.85 per million words, and the latter – 5.51 per million words. Both of these sequences are present in LOCNESS, in contrast to the unit *to my mind*. It is hard to explain the overuse of *to my mind* as it does not seem to be agitated by L1 patterns.

In ICLE-RU, I register the use of *my* in such non-nativelike multi-word units as *as far as my opinion is concerned, I* and *as far as my point of view is concerned*. Totally absent in either NS corpora or other learner writing corpora, these units seem to be an attempt to transfer the same L1 sequence as discussed in the Section 4.1.2 (*что касается меня, то я*) into the target language.

#### 4.1.4 The first person singular pronoun *MINE*<sup>33</sup>

The pronoun *mine* is the least frequent of first person singular pronouns, as it is used in a limited number of constructions. *Mine* shows least involvement of the author in the text and rather signals the attributive character of a preceding noun. The overuse is not found to be statistically significant according to the log-likelihood test.<sup>34</sup>

**Table 15.** The pronoun *mine* in LOCNESS and ICLE-RU. Frequency per 10,000 words

Feature	LOCNESS		ICLE-RU		ICLE-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
The pronoun <i>mine</i>	5	0.16	6	0.27	1.72

<sup>33</sup> This study considers possessive pronouns as a part of first person pronouns.

<sup>34</sup> Statistical significance less than 95%.

#### 4.1.5 The first person plural pronoun *WE*

The results show that, in argumentative writing, the pronoun *we* is even more overused by Russian learners than the pronoun *I*. The log-likelihood test shows that the overuse is statistically significant.<sup>35</sup>

**Table 16.** The pronoun *we* in LOCNESS and ICLE-RU. Frequency per 10,000 words

Feature	LOCNESS		ICLE-RU		ICLE-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
The pronoun <i>we</i>	935	29.00	1758	78.36	2.70

However, the pronoun *we* does not always refer to the writer. In its generic function, *we* can be used to refer to people in general. In Section 5.2, the functions of the pronoun in ICLE-RU are analysed and discussed in detail, which makes it clear if the overuse indicates the overt presence of the author in the text.

The concordances of *we* in ICLE-RU reveal the feature referred to by Petch-Tyson (1998) as ‘the chaining of sentence-initial features of W/R visibility’ (*ibid.*: 114):

- (20) *It is obvious, I believe*, that *we* have to distinguish between imagination and dreaming. (ICLE-RU)
- (21) *In my opinion we can't say with certainty, categorically* that a person's financial reward should be commensurate with his contribution to the society.

Here, the features of W/R visibility receive extra focus through topicalisation structures and repetition of the elements. NS writing typically has only one pre-‘topic’ element:

- (22) *I belive we* have being doing this for years by selective training of our soldiers & muscle tonning. (LOCNESS)
- (23) *I don't believe we* cannot sympathise with him when such characters as Cherea, Scipion, Caesonia and Helicon never hate him and even sympathise with him themselves. (LOCNESS)

However, this feature is typical of learner writing in general, and is not specific for Russian learners. An interesting finding in ICLE-RU is that learners choose to combine the pronoun *we* with the verb *say* to introduce conclusive statements, and with the verb *see* to develop an

<sup>35</sup> Statistical significance – 99.99%.



argument. Native speakers, on the other hand, sum up with the help of the verb *see*. The verb *say* does not collocate with *we* in LOCNESS; however, the clusters with *say* illustrate its frequent use to elaborate a topic or to introduce examples. We can compare the examples below:

Concordance of NS and NNS *we + see*:

- (24) Now that *we have seen* both sides to this argument solutions should be discussed.  
(LOCNESS)
- (25) If we take music, for example, *we will see* that all these modern synthesising machines do not leave any place for imagination, because everything is already insinuated. (ICLE-RU)

Concordance of NS and NNS *we + say*:

- (26) For example, *they say* that Britain is a small island which has stood alone for centuries.  
(LOCNESS)
- (27) Therefore *we can say* that cinema spreads two great illusions: illusion of living a more interesting and attractive life and illusion of collective or mass consciousness. (ICLE-RU)

This is an observed tendency, rather than an absolute rule. It demonstrates that in NS writing, *we* typically *see/can see/ have seen*, and *they/many/people* typically *say*. In Russian learner writing, *we* collocates both with *say/can say/can't say* and *see/will see/'ll see*, but for different purposes. The latter is used to argue, the former– to draw conclusions. The use of tenses supports my observation (*cf.* Table 17 and 18):

**Table 17.** The use of tenses in the clusters with the pronoun *we + verb see*

LOCNESS	ICLE-RU
<i>we see</i>	<i>we'll see</i>
<i>we can see</i>	<i>we will see</i>
<i>we have seen</i>	<i>we will be able to see</i>

**Table 18.** The use of tenses in the clusters with the pronoun *we + verb say*

LOCNESS	ICLE-RU
<i>they/people/many say that</i>	<i>we can say</i>
<i>to say that</i>	<i>we can't say</i>

The reason for such uses may lie in the transfer of L1 function: the verbs *видеть* (=to see) and *говорить* (=to say) collocate with the pronoun *мы* (=we) in Russian L1 writing, but tend to function differently from the similar units in English. An in-depth study of the sequences is required to reach a more detailed conclusion.

#### 4.1.6 The first person plural pronoun *US*

The overuse of the pronoun *us* is greater than the overuse of the pronoun *we*: 2.70 against 3.17 (cf. Table 16 and 19). The log-likelihood test shows that the overuse is statistically significant.<sup>36</sup>

**Table 19.** The pronoun *us* in LOCNESS and ICLE-RU. Frequency per 10,000 words

Feature	LOCNESS		ICLE-RU		ICLE-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
The pronoun <i>us</i>	212	6.57	467	20.82	3.17

The overuse of the pronoun can be explained through the high overuse of some especially frequent multi-word units.

**Table 20.** 2- and 3-gram clusters with the pronoun *us* with a minimum frequency of 5 instances in LOCNESS and ICLE-RU. Frequency per 10,000 words. Sorted according to the overuse registered in ICLE-RU

Feature	LOCNESS		ICLE-RU		ICLE-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
<i>gives us/give us/gave us</i>	9	0.28	48	2.14	7.64
<i>let us/let's</i>	15	0.62	93	4.15	6.68
<i>help up/helps us</i>	10	0.31	27	1.20	3.88
<i>makes us/make us</i>	9	0.28	23	1.03	3.67
<i>tells us/tell us</i>	17	0.53	17	0.76	1.44
<i>show us</i>	5	0.16	5	0.22	1.44

Table 20 illustrates high overuse of a limited number of multi-word units among Russian learners. The log-likelihood test proves that the overuse of each unit is statistically significant.<sup>37</sup>

The two most frequent units with the pronoun *us* in ICLE-RU are *give(s)/gave us* (in the meaning of ‘presenting someone with something’, not ‘transferring something physically’) and *let us/let's*. The overuse of *let us/let's* is also registered in French learner writing by Paquot (2008). It results from the transfer of French formal writing conventions on to the interlanguage, where first person plural imperative is often used to organize discourse: *citons* (=let us mention), *envisageons* (=let us consider), *examinons* (=let us analyse), *notons* (=let us note), *pensons* (=let us imagine) and other. In this respect, Russian language is very similar to French. First person

<sup>36</sup> Statistical significance – 99.99%.

<sup>37</sup> Statistical significance – 99.99%.

plural imperatives are common in Russian argumentative and formal writing: *Рассмотрим* (=let us consider), *проанализируем* (=let us analyse), *отметим* (=let us mention):

(28) **Вспомним** жаркое лето 2010 года, когда леса пылали пожарами. (CRUAW)

[=Let us remember hot warm summer of 2010, when woods were in fire.]

(29) **Вспомним** хотя бы человека, чьё упорство, чья безграничная сила духа, чьи знания позволили нам узнать так много об истории нашей могучей страны. (CRUAW)

[=Let us remember the person, whose persistency, tremendous spirit and knowledge allowed us to find so much about the history of our mighty country.]

Table 21 presents the frequencies of first person plural imperatives in ICLE-RU and CRUAW (for more details on CRUAW see Section 3.2.5).

**Table 21.** First person plural imperatives in ICLE-RU and CRUAW. Frequency per 10,000 words

Corpora	Raw frequencies	Relative frequency per 10,000 words
ICLE-RU	93	4.15
CRUAW	8	4.00

The results suggest that the frequency in learner's L1 may be the cause of the overuse in the interlanguage. The hypothesis is supported by the similar conclusions drawn by Paquot (*ibid.*) on the transfer of L1 conventions by French and Spanish learners. Compared to the results by Paquot (*ibid.*), Russian learners use *let us/let's* less frequently than French learners, but more frequently than Spanish, Dutch, German and Polish learners.

A similar hypothesis can be put forward for the multi-word unit *gives us/give us/gave us*. The verb *give* + first person plural pronoun *us* is commonly used in Russian formal writing:

(30) Литература **передает нам** жизненный опыт множества людей - писателей, их героев. Безусловно, книги являются важным способом развития личности. (CRUAW)

[=Literature **gives us** the life experiences of other people – writers, their characters. It is no doubt that reading books is an important ways of developing your personality.]

(31) Наука, которая **доставляет нам** внутреннее удовлетворение и служит главной причинной нашего духовного развития [...]. (CRUAW)

[=Science, which **gives us** inner satisfaction and serves as the main reason for our spiritual development [...]]

Table 22 indicates that the sequence *give/gives/gave us* has about the same frequency in ICLE-RU and CRUAW, whereas its frequency in the NS corpus is significantly lower and constitutes 0.28 instances per 10,000 words. Therefore, my hypothesis is that the frequency in the interlanguage reflects the frequency in L1.

**Table 22.** The verb *give* + the pronoun *us* in ICLE-RU and CRUAW. Frequency per 10,000 words

Corpora	Raw frequencies	Relative frequency per 10,000 words
ICLE-RU	48	2.14
CRUAW	4	2.00

The sequence *give/gives/gave us* is more frequent in ICLE-RU than in CRUAW, which may result from generalisation. In L1 writing, students use a variety of verbs with a meaning of ‘give’ combined with the pronoun *us*. Learners may not be able to find the English counterpart in L2 and use the verb with a broad semantic meaning – *to give*.

**Figure 10.** The transfer of L1 verbs with a semantic meaning ‘give’ into L2 by means of generalisation



#### 4.1.7 The first person plural pronoun *OUR*<sup>38</sup>

According to Table 23, Russian learners overuse the pronoun *our* at a high rate of 2.83. The log-likelihood test shows that the overuse is statistically significant.<sup>39</sup>

**Table 23.** The pronoun *our* in LOCNESS and ICLE-RU. Frequency per 10,000 words

Feature	LOCNESS		ICLE-RU		ICLE-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
The pronoun <i>our</i>	584	18.11	1148	51.17	2.83

<sup>38</sup> This study considers possessive determiners as a part of first person pronouns.

<sup>39</sup> Statistical significance – 99.99%.

The analysis of the collocations shows that the majority of them are nouns that are pre-defined by the topics of the essays, which especially concerns the topic ‘In *our* modern world’.<sup>40</sup> ICLE-RU and LOCNESS share a great amount of the most frequent collocations: for example, *life, society, world, country* etc. However, the overuse is significant, and in order to explain it, I analyse the 2- and 3-gram clusters in LOCNESS, ICLE-RU and CRUAW. The cluster *in our* is highly frequent among Russian learners and is overused by them with the high rate of 4.63. The log-likelihood test shows that the overuse is statistically significant.<sup>41</sup> A seemingly uninteresting multi-word unit *in our* leads to interesting observations. A search for *в наши\** (=in our) in CRUAW, shows that the unit is very frequent in L1 writing:<sup>42</sup>

**Table 24.** The 2-gram cluster *in our* in LOCNESS and ICLE-RU, and the 2-gram cluster *в наши\** in CRUAW. Frequency per 10,000 words

Corpora	Raw frequencies	Relative frequency per 10,000 words
LOCNESS	67	2.08
ICLE-RU	216	9.63
CRUAW	30	14.97

This high relative frequency of *в наши\** (=in our) in CRUAW is due to it being a part of a limited number of highly frequency prefabs: *в наши дни* (=in our days),<sup>43</sup> *в наше время* (=in our times), *в нашей жизни* (=in our life) and *в наших сердцах/душах* (=in our hearts/souls). Learners are likely to introduce some of these prefabs from their L1 to the interlanguage. In case of positive transfer, the overuse rate may be especially high.

#### 4.1.8 The first person plural pronoun *OURS*<sup>44</sup>

The overuse of the pronoun *ours* is the highest among first person pronouns in argumentative writing. However, it should be treated with caution due to the low raw frequencies of *ours* in both corpora. The overuse is statistically significant according to log-likelihood test.<sup>45</sup> The results presented in Table 25 exclude obvious misspellings from both corpora, as for example, when *ours* is supposed to be *hours*.

<sup>40</sup> The entire title is ‘Some people say that in our modern world, dominated by science technology and industrialisation, there is no longer a place for dreaming and imagination. What is your opinion?’. More information on the topics of essays is available the official webpage of the *Université Catholique de Louvain* <https://www.uclouvain.be/en-cecl-locness.html> <http://www.uclouvain.be/en-317607.html>

<sup>41</sup> Statistical significance – 99.99%.

<sup>42</sup> The Russian language has a very developed system of noun morphology, which varies depending on gender, number and case. To include all forms of the Russian pronoun *our*, I had to take the form with zero ending.

<sup>43</sup> Translations given in brackets are direct word-for-word translations of Russian units.

<sup>44</sup> This study considers possessive pronouns as a part of first person pronouns.

<sup>45</sup> Statistical significance – 99.9%.

**Table 25.** The pronoun *ours* in LOCNESS and ICLE-RU. Frequency per 10,000 words

Feature	LOCNESS		ICLE-RU		ICLE-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
The pronoun <i>ours</i>	2	0.06	13	0.58	9.34

The overuse results from grammatically incorrect uses of the pronoun *ours*. Russian language does not differentiate between the form of *our* and *ours*. This leads to incorrect uses of the pronoun *ours*:

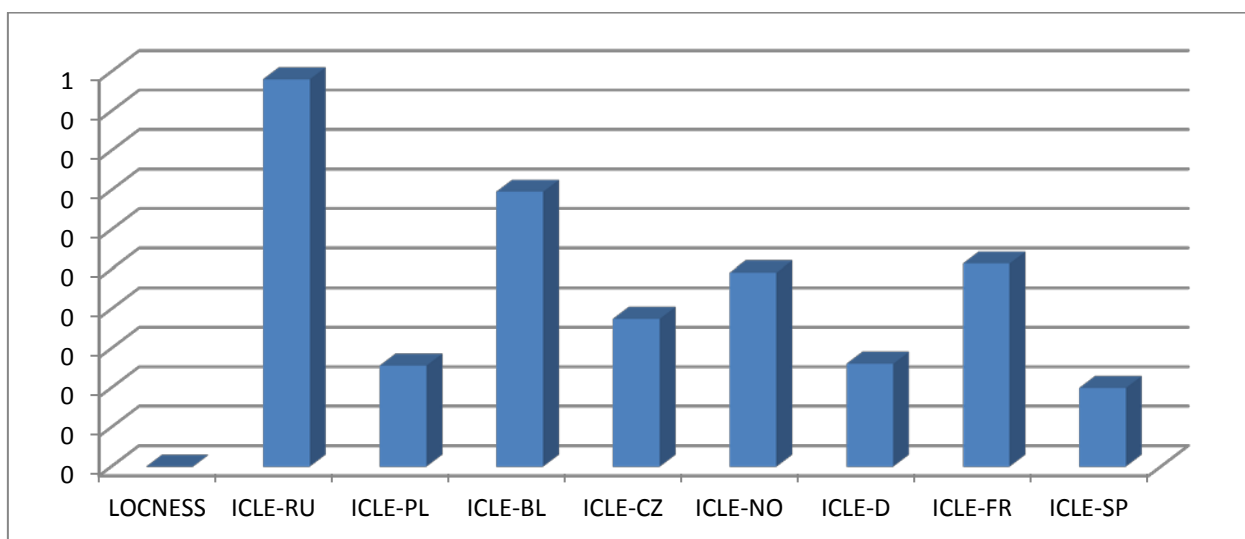
- (32) [...] as the huge amount of sleeps slips away in the ocean of inconceivable things in the very close to ***ours world*** of surreality, understandable for the most risky bodies of consciousness. (ICLE-RU)
- (33) And again very much depends on what we actually want, on ***ours intention***, and, basically, on what we want to reach in this life. (ICLE-RU)

Another frequent construction in Russian learner writing is ADJ + N + OF OURS:

- (34) Unfortunately, nothing else seems to be capable of saving him from feeling frustrated and lost in ***this horizontal world of ours***, which we inherited from the romantic XIX-th century, with no heavens above and no hell below. (ICLE-RU)
- (35) With the industrial and scientific revolution it has come about that to live and work well, to be at home in ***this new world of ours***, we need to have an education which is limited neither to the know-how of our jobs nor to our duties of citizenship. (ICLE-RU)
- (36) Let me suggest an idea: ***this noisy, bustling and sometimes ugly world of ours***, filled with machinery, can, in some cases, even help imagination to develop, because people often feel a desire to create a little world of their own, to hide from the domination of mechanisms. (ICLE-RU)

This construction increases the level of W/R visibility and highlights the descriptive adjectives, which are used to qualify nouns. The sequence is absent in LOCNESS and is very infrequent in other NS corpora. Learners tend to overuse the construction (*cf.* Figure 11):

**Figure 11.** The construction ADJ + N + *of/like/as OURS* across the corpora. Frequency per 10,000 words



#### 4.1.9 Concluding remarks on author identity in argumentative writing

The results of the investigation indicate that Russian learners overuse all first person pronouns, but at a different rate: first person singular pronouns are less overused than first person plural. The most overused pronoun is *ours*, which is partly due to incorrect grammatical uses, whereas the least overused pronouns are *my* and *mine* (cf. Table 26):

**Table 26.** The first person pronouns in LOCNESS and ICLE-RU. Frequency per 10,000 words

Feature		LOCNESS		ICLE-RU		ICLE-RU overuse
		Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
First person singular pronouns	<i>I</i>	982	30.45	1445	64.41	2.11
	<i>me</i>	94	2.92	175	7.80	2.64
	<i>my</i>	260	8.06	280	12.48	1.55
	<i>mine</i>	5	0.16	6	0.27	1.72
<b>TOTAL 1st pers. singular</b>		<b>1341</b>	<b>41.59</b>	<b>1906</b>	<b>84.95</b>	<b>2.02</b>
First person plural pronouns	<i>we</i>	935	29.00	1758	78.36	2.70
	<i>us</i>	212	6.57	467	20.82	3.17
	<i>our</i>	584	18.11	1148	51.17	2.83
	<i>ours</i>	2	0.06	13	0.58	9.34
<b>TOTAL 1st pers. plural</b>		<b>1733</b>	<b>53.75</b>	<b>3387</b>	<b>150.92</b>	<b>2.81</b>
<b>TOTAL</b>		<b>3074</b>	<b>95.33</b>	<b>5293</b>	<b>235.88</b>	<b>1.71</b>

Compared to the results from Petch-Tyson (1998), Russian learners take an intermediate position in the use of first person singular.<sup>46</sup> They use it more often than Dutch and French (the frequency of 70.69 and 62.69 per 10,000 words respectively), but less than Finnish and Swedish (the frequency of 98.23 and 88.06 per 10,000 words respectively). As for first person plural, Russian learners are last but one, with only Swedish learners overusing the pronouns at a higher rate of 266.94 per 10,000 words. If we consider the results from Ädel's study (2008) of timed and untimed argumentative essays written by Swedish learners, then the use of first person singular pronouns in ICLE-RU is very close to the untimed essays. Untimed essays are proved to have lower frequency of the pronouns in comparison to the timed ones.

The present study shows that frequencies from the students' L1 are at times reflected in the interlanguage. Together with frequencies, we can observe a transfer of function, register and form:

**Figure 12.** L1 transfer of frequency, function, register and form into the interlanguage.

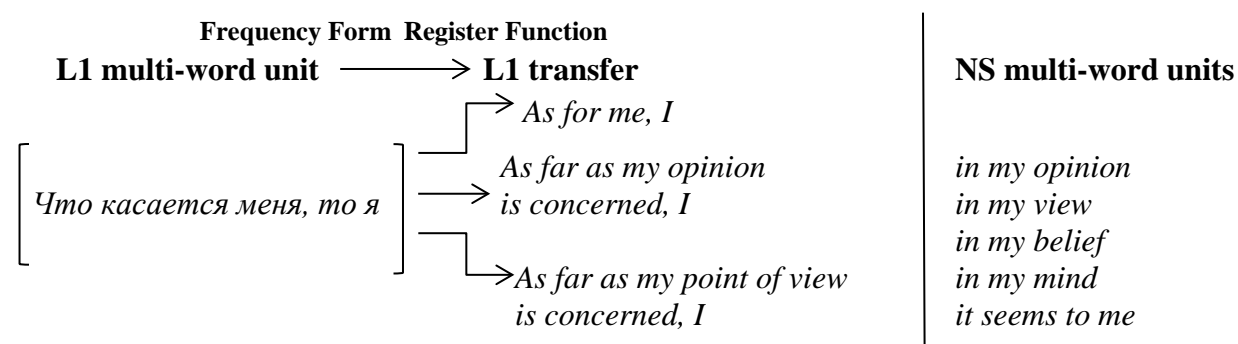


Figure 12 illustrates that when Russian learners construct their author identity in argumentative writing, they rely heavily on the devices they commonly use to argument in their L1. Therefore, instead of adopting NS techniques for expressing author identity, learners tend to either, in case of positive transfer, strongly overuse a unit, or introduce a non-nativelike construction from their L1.

<sup>46</sup> For the simplicity of comparison, I took myself a liberty to calculate the frequency per 10,000 words, based on the raw frequencies presented in Table 8.1 by Petch-Tyson (1998).



## 4.2 Author identity in academic writing

This section presents a comparison of first person pronouns in Russian learner and English NS academic writing. The overall overuse rate in academic writing is more moderate than in argumentative writing, which is due to the conventions of the domain. Just as in the previous section, the results of the study are presented in figures and tables with raw and relative frequencies per 10,000 words. Relative frequencies are important in order to compare the results of NNS writing to NS writing. As discussed in Section 2.1, NS writing is a control tool for identifying deviation in Russian learner writing. However, it is important to remember that even among NS essays the variation is often significant. For example, some texts in BAWE-LING contain no instances of the pronoun *I*,<sup>47</sup> whereas in others texts I register the frequency of 151.85 per 10,000 words or even the frequency of 199.08 per 10,000 words against the average of 28.31 per 10,000 words.<sup>48</sup>

The section on academic writing studies multi-word units in a less considerable scale than the section on argumentative writing. This is due to the size of VESPA-RU, which is 28,651 words and is significantly smaller than BAWE-LING that consists of 197,291 words. Therefore, not all features stand out clearly enough, and the results should be treated with caution. In this connection, I want to underline the importance of the work done by scholars within corpus linguistics. I was personally able to estimate the freedom a large corpus gives when I worked with ICLE-RU and LOCNESS. At a later stage, I worked with VESPA-RU and realised the constraints of a smaller corpus. The next eight sub-sections present a detailed overview of NS and Russian learner use of first person pronouns in English academic writing.

### 4.2.1 The first person singular pronoun *I*

Table 27 indicates the overuse rate of 1.61 instances per 10,000 words for the pronoun *I* in academic writing, which is lower than the corresponding overuse rate in argumentative writing – 2.11 instances per 10,000 words. The registered overuse is statistically significant according to the results of the log-likelihood test.<sup>49</sup>

---

<sup>47</sup> The file 6206a contains 3,428 words and no instances of the pronoun *I*.

<sup>48</sup> The file 6062c contains 2,239 words and 34 instances of the pronoun *I*. The frequency is 151.85 per 10,000 words. The file 6048b contains 2,311 words and 46 instances of the pronoun *I*. The frequency is 199.05 per 10,000 words.

<sup>49</sup> Statistical significance – 99.99%.

**Table 27.** The pronoun *I* in BAWE-LING and VESPA-RU. Frequency per 10,000 words

Feature	BAWE-LING		VESPA-RU		VESPA-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
The pronoun <i>I</i>	555	28.13	130	45.37	1.61

The analysis of the top ten 2- to 4-grams in each of the corpora reveals interesting results, which indicate possible functional differences between the NS and learner uses of the pronoun *I*.

**Table 28.** The top ten recurrent word-combinations in LOCNESS and ICLE-RU (search for 2- to 4-grams involving *I*)

BAWE-LING	VESPA-RU
<u><i>I will</i></u>	<u><i>I have</i></u>
<u><i>I have</i></u>	<u><i>I will</i></u>
<i>I think</i>	<u><i>I was</i></u>
<i>I feel</i>	<i>research I</i>
<i>that I</i>	<i>I know</i>
<i>I would</i>	<i>that I</i>
<i>I believe</i>	<u><i>I decided</i></u>
<u><i>I decided</i></u>	<i>I got</i>
<i>I am</i>	<i>I know you</i>
<u><i>I was</i></u>	<i>I know you didn't</i>

Appendix 2 presents the raw and relative frequencies and the VESPA-RU overuse rate for each of the multi-word units in Table 28. The clusters in the table are given in descending order of frequency. The sequences *I will*, *I have*, *I was* and *I decided* are present among the top ten 2- to 4-grams both in NS and in learner writing. The analysis of concordance lines reveals that these sequences form bundles, which according to Hyland's classification (2002a) perform the functions of *explaining a procedure* and *stating a purpose*. As discussed in Sections 2.4.1 and 2.4.2, the functions imply low risk for the author, which makes learners feel comfortable about using *I* here. Typical examples of *I have*, *I decided*, *I will* and *I am going to* in the functions of explaining a procedure and stating a purpose are presented below:

Explaining a procedure:

- (37) The text *I have chosen* to discuss is an article from a tabloid newspaper 'The Daily Mirror' dated November 17 2005, entitled 'CSA RIP'. (BAWE-LING)
- (38) After some thinking and reading *I decided* to limit my research to preverbal and postverbal positions [...]. (VESPA-RU)

Stating a purpose:

- (39) *I am going to* look at whether research into aphasia proves or disproves these claims. (BAWE-LING)
- (40) Thus *I will look at* how Theme is defined by SFL and FSP [...]. (VESPA-RU)

According to Table 28, BAWE-LING also contains units that, according to Hyland (2002a), perform high-risk native-like functions of *elaborating an argument* and *stating results/claims*: *I think, I believe* and *I feel*. These sequences are not present among the top ten 2- to 4-grams in VESPA-RU. This may indicate that learners are reluctant to express their author identity through the use of high-risk functions. The examples (41) – (44) from BAWE-LING illustrate the use of *I think, I believe* and *I feel* to elaborate an argument and state results/claims:

Elaborating an argument:

- (41) *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)
- (42) However, *I believe* the columns 'Probable' and 'Possible' in Table II (page 329) should be reversed to allow the 'likelihood' continuum to be represented in the conventional order. (BAWE-LING)

Stating results/claims:

- (43) *I feel*, though, that our students do have many different needs from the average EFL student, and more research should be done into the best way to help students equip themselves for a long period of academic study in English. (BAWE-LING)
- (44) *I think* what can be seen through the research that has been done is that there is a clear link between class (incorporating status and prestige) with gender. (BAWE-LING)

Section 5.3 presents an in-depth qualitative study of the pronoun *I* in NS and learner academic writing. However, the observations in this section allow for some preliminary conclusions on the functional differences in the use of the pronoun *I*. The analysis supports Hyland's conclusion (*ibid.*) that learners may be reluctant to overtly display their author identity in the text. Anglo-American academic conventions, however, require authors to show their personal commitment to their arguments and to stand for their claims and results.

#### 4.2.2 The first person singular pronoun *ME*

The pronoun *me* is the only pronoun that is underused in Russian learner writing. However, the underuse is not statistically significant, as the log-likelihood equals 0.19 and is well below the critical value of 3.84.

**Table 29.** The pronoun *me* in BAWE-LING and VESPA-RU. Frequency per 10,000 words

Feature	BAWE-LING		VESPA-RU		VESPA-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
The pronoun <i>me</i>	42	2.13	5	1.75	0.82

In VESPA-RU, learners avoid the two multi-word units that lead to the overuse in argumentative writing: *it seems to me* and *as for me* (cf. Section 4.1.2). It confirms the hypothesis that together with form, frequency and function, Russian learners transfer the register of the L1 unit *что касается меня* (=as for me). According to Russian academic conventions, such units are not appropriate in academic writing and should only be used in not very formal writing or speech.

Due to the low raw frequency, the pronoun *me* does not have a wide range of collocations in VESPA-RU. In three out of five instances, the pronoun collocates with the verb *help*, the two other collocates are the verbs *provide* and *enable*. BAWE-LING, on the other hand, contains a wide range of collocates, the most frequent of them being the verbs *allow*, *give*, *bring* and *lead*. These collocates increase the accuracy in the use of author identity, show a high degree of writers' proficiency and illustrate their rich vocabulary. More attention needs to be paid to such collocates in the course of learning academic writing.

#### 4.2.3 The first person singular pronoun *MY*

The pronoun *my* is overused by Russian learners in academic writing at a rate of 1.84. The log-likelihood test shows that the overuse is statistically significant.<sup>50</sup> The overuse rate is higher in academic writing than in argumentative writing: 1.84 against 1.55 respectively. However, the frequency of the pronoun *my* is lower in VESPA-RU than in ICLE-RU.

---

<sup>50</sup> Statistical significance – 99%.

**Table 30.** The pronoun *my* in BAWE-LING and VESPA-RU. Frequency per 10,000 words

Feature	BAWE-LING		VESPA-RU		VESPA-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
The pronoun <i>my</i>	116	5.88	31	10.82	1.84

In both corpora, the pronoun *my* mainly collocates with the nouns characteristic of the academic domain: *results, findings, experience, data, analysis* etc. Table 30 indicates that NS writers follow an idiom principle and collocate *my* with a limited amount of nouns. However, the prefabs they form have a rather high frequency of up to 0.35 instances per 10,000 words. Russian learners follow the open choice principle and tend to collocate the pronoun *my* with a wider range of nouns than NS authors. All, but one prefab, occur only once, but receive a high relative frequency due to the low amount of words in VESPA-RU (28,651 words) – 0.35 instances per 10,000 words. The results in Table 30 are sorted according to how frequent the collocations are in the corpora:

**Table 31.** Collocates of the pronoun *my* in BAWE-LING and VESPA-RU with a frequency above 0.15 instances per 10,000 words (collocation window span: 0 right – 1 left).

BAWE-LING	VESPA-RU
<i>results</i>	<i>research</i>
<i>findings</i>	<i>hypothesis</i>
<i>experience</i>	<i>conclusions</i>
<i>data</i>	<i>testimony</i>
<i>view</i>	<i>term</i>
<i>teaching</i>	<i>requirements</i>
<i>learning experience</i>	<i>purpose</i>
<i>analysis</i>	<i>parameters</i>
	<i>opinion</i>
	<i>observations</i>
	<i>mind</i>
	<i>material</i>
	<i>intention</i>
	<i>expectations</i>
	<i>data</i>
	<i>conclusions</i>
	<i>approach</i>
	<i>analysis</i>

Appendix 3 presents the raw and relative frequencies for each of the multi-word units in Table 30. I do not calculate VESPA-RU overuse, because the raw frequencies of the units in the corpora are very low, and the overuse rate may not reflect real tendency.

#### 4.2.4 The first person singular pronoun *MINE*

The pronoun *mine* is hardly present in either of the corpora: no instances in VESPA-RU and one instance in BAWE-LING, which is found in exemplification. The pronoun does not seem to be characteristic of the academic domain and will thus not be discussed any further here.

#### 4.2.5 The first person plural pronoun *WE*

The overuse of the pronoun *we* is the greatest in academic writing, and is statistically significant according to the log-likelihood test.<sup>51</sup> Compared to Fossan's results (2011), Russian learners make more use of the pronoun *we* than Norwegian learners. Besides, Norwegian learners overuse the pronoun *I* at a higher rate than the pronoun *we* (1.97 against 1.88 respectively), and first person singular at a higher rate than first person plural (1.87 against 1.58 respectively). Russian learners, on the contrary, make less use of the pronoun *I* than *we* (1.61 against 2.43 respectively), and first person singular than first person plural (1.60 against 2.04). Therefore, the 'we' perspective dominates in Russian academic writing.

**Table 32.** The pronoun *we* in BAWE-LING and VESPA-RU. Frequency per 10,000 words

Feature	BAWE-LING		VESPA-RU		VESPA-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
The pronoun <i>we</i>	306	15.51	108	37.70	2.43

Following the study of *we* in argumentative writing, I decided to examine how native speakers and learners collocate the pronoun *we* with the verbs *see* and *say* in academic writing. In BAWE-LING, in the same way as in LOCNESS, the pronoun *we* collocates with the verb *see* in order to introduce conclusions and sum up the findings. As discussed in Section 4.1.5, in ICLE-RU, learners tend to use *we + see* to develop an argument, whereas in VESPA-RU, I observe a switch towards a more native-like use of the prefabs for concluding purposes:

- (45) As *we can see* the results are more or less consistent if we look at the averages. (VESPA-RU)
- (46) Going back to John Algeo's claim with our results *we can see* that the progressive aspect is found to be more common in BrE [...]. (VESPA-RU)

---

<sup>51</sup> Statistical significance – 99.99%.

In ICLE-RU, the pronoun *we* together with the verb *say* are commonly used to develop an argument. In VESPA-RU, I find no such cluster. Instead, Russian learners introduce the verb *find*, which they collocate with the pronoun *we* for elaborating an argument:

(47) Throughout the text *we can find* many cases of repetition (the country, this country, we, our, us, etc). (VESPA-RU)

(48) The anaphoric reference is one of the cohesive devices that is used very often in the texts. In the text 2 *we can find* anaphoric reference almost in every paragraph. (VESPA-RU)

Table 32 indicates that the collocation *we + find* is strongly overused in VESPA-RU. The overuse is statistically significant according to the log-likelihood test:<sup>52</sup>

**Table 33.** The pronoun *we* + the verb *find* in BAWE-LING and VESPA-RU. Frequency per 10,000 words

Feature	BAWE-LING		VESPA-RU		VESPA-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
<i>we can find/we find</i>	3	0.15	9	3.14	4.43

BAWE-LING has only three instances of *we can find/we find*, which means that this verb is not commonly used for elaborating an argument in NS writing. According to Hyland (2002b), elaborating an argument is a key function in academic writing, which requires authorial presence. The fact that Russian learners use the pronoun *we* instead of *I* implies that they prefer to share the responsibility for the arguments with their readers. Further functional analysis of the pronoun *we* in Russian learner writing is presented and discussed in detail in Section 5.4.

#### 4.2.6 The first person plural pronoun *US*

The pronoun *us* is overused by Russian learners; however, the log-likelihood test reveals that the overuse is not statistically significant. Compared to argumentative writing, the overuse is more moderate, which is in line with the general tendency of lower W/R visibility in academic writing.

**Table 34.** The pronoun *us* in BAWE-LING and VESPA-RU. Frequency per 10,000 words

Feature	BAWE-LING		VESPA-RU		VESPA-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
The pronoun <i>us</i>	77	3.90	16	5.58	1.43

<sup>52</sup> Statistical significance – 99.99%.

Analysing the 2- to 4-grams, I register a general reduction of the overuse of several frequent multi-word units. However, the unit *let us/let's* retains its high frequency of 1.75 instances per 10,000 words in academic writing. It is lower than the frequency of 4.15 instances per 10,000 words in ICLE-RU, but higher than 0.10 instances per 10,000 words registered in BAWE-LING. As discussed in Section 4.1.6, first person plural imperatives are common in Russian academic writing. Therefore, the overuse is likely to result from L1 transfer. Russian students need to receive extra instruction on the use of *let us/let's* to prevent them from overusing the units. They should also be presented with alternative means of expressing their author identity in academic writing that can be used instead of *let us/let's*.

#### 4.2.7 The first person plural pronoun *OUR*

In VESPA-RU, learners overuse the pronoun *our* at a low rate of 1.18. Besides, the overuse is not statistically significant according to the log-likelihood test. Compared to argumentative writing, the overuse in academic writing is significantly lower: 2.83 against 1.18. If we compare the relative frequency per 10,000 words, the difference is even more striking with a frequency of 51.17 instances against 4.89 instances per 10,000 words.

**Table 35.** The pronoun *our* in BAWE-LING and VESPA-RU. Frequency per 10,000 words

Feature	BAWE-LING		VESPA-RU		VESPA-RU overuse
	Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
The pronoun <i>our</i>	82	4.16	14	4.89	1.18

In this section, I do not analyse any multi-word units that are overused by Russian learners in academic writing, as no patterns can be identified. The reason for this is a relatively low number of hits, which is due to the rather small size of VESPA-RU.

#### 4.2.8 The first person plural pronoun *OURS*

No instance of the pronoun *ours* is registered in either BAWE-LING or VESPA-RU. In comparison, Russian learners were found to overuse *ours* in argumentative writing (*cf.* Section 4.1.8). The overuse results from incorrect uses of the pronoun. In academic writing, I do not register such mistakes. This may indicate that learners in VESPA-RU have a better command and more experience in English than learners in ICLE-RU. Therefore, developmental factors may also play a significant role in the decrease of W/R visibility in learner academic writing.



#### 4.2.9 Concluding remarks on author identity in academic writing

In academic writing, Russian learners overuse all but two pronouns in comparison to native speakers. The pronoun *me* is the only pronoun that is underused in VESPA-RU. However, the underuse is found to be statistically not significant. The overuse of the pronouns *mine*, *us* and *our* is also not statistically significant according to the log-likelihood test.

**Table 36.** The first person pronouns in BAWE-LING and VESPA-RU. Frequency per 10,000 words

Feature		BAWE-LING		VESPA-RU		VESPA-RU overuse
		Raw frequencies	Per 10,000 words	Raw frequencies	Per 10,000 words	
First person singular pronouns	<i>I</i>	555	28.13	130	45.37	1.61
	<i>me</i>	42	2.13	5	1.75	0.82
	<i>my</i>	116	5.88	31	10.82	1.84
	<i>mine</i>	1	0.05	0	0.00	-
<b>TOTAL 1st pers. singular</b>		<b>714</b>	<b>36.19</b>	<b>166</b>	<b>57.94</b>	<b>1.60</b>
First person plural pronouns	<i>we</i>	306	15.51	108	37.70	2.43
	<i>us</i>	77	3.90	16	5.58	1.43
	<i>our</i>	82	4.16	14	4.89	1.18
	<i>ours</i>	0	0.00	0	0.00	-
<b>TOTAL 1st pers. plural</b>		<b>465</b>	<b>23.57</b>	<b>138</b>	<b>48.17</b>	<b>2.04</b>
<b>TOTAL</b>		<b>1179</b>	<b>59.76</b>	<b>304</b>	<b>106.10</b>	<b>1.78</b>

Academic writing has a lower level of involvement than argumentative writing. The results in this section indicate that Russian learners apply first person pronouns to a lesser extent than Norwegian and French learners (Paquot, Hasselgård and Ebeling, 2013). Both in argumentative and in academic Russian learner corpora, first person plural pronouns (*we*, *us*, *our*, *ours*) are overused at a higher rate than the first person single (*I*, *me*, *my*, *mine*). This is not the case for Norwegian learners, who, according to the results from Fossan's study (2011), overuse first person pronouns at a higher rate.

The results of the present study support the findings by Vassileva (1998), who investigates the first person singular 'I' perspective against the first person plural 'we' perspective in L1 German, French, Russian, Bulgarian and English writing. The results of the study show that English and Russian students construct their identity through the use of different pronouns. In Anglo-American rhetorics, the 'I' perspective is a key part of successful construction of author's academic identity, whereas in Slavic rhetoric, the 'we' perspective is dominant, and the 'I' perspective is nearly absent. The study of ICLE-RU and VESPA-RU demonstrates that Russian learners to a high degree manage to adopt the Anglo-American rhetoric, as they introduce the 'I' perspective into their writing. However, we still observe the

traces of L1 in the learner writing, as the overuse of the 'we' perspective is higher than in other learner varieties. Therefore, despite advanced command of the foreign English language, L1 transfer remains a barrier that is difficult to overcome in order to gain a native-like performance.

## 5. RESULTS AND ANALYSIS: QUALITATIVE STUDY OF AUTHOR IDENTITY IN ARGUMENTATIVE AND ACADEMIC WRITING

The previous chapter focused on the quantitative analysis of NS and learner writing and revealed that Russian students overuse the ‘we’ perspective (*we, us, our* and *ours*) at a higher rate than the ‘I’ perspective (*I, me, my* and *mine*). This chapter presents a qualitative analysis of the first person pronouns *I* and *we* in argumentative and academic writing. To carry out the in-depth study, I analysed the concordance lines with these first person pronouns and identified their functions. The functional classifications used in this chapter are discussed in Section 2.7 of the Theoretical part of this master thesis. This study focuses on the pronouns *I* and *we*, since these pronouns are most representative of author identity, and their functional classifications have already been worked out and can be used in my investigation. According to my knowledge, no classifications have been worked out for the analysis of *me, my, mine* and *us, our, ours*. This is probably due to the fact that the possessive pronouns *mine* and *ours* are infrequent, and the pronouns *me* and *us* can be prompted by the use of the verb that precedes them. The verbs *tell, inform, seem, lead* and other are typically followed by the pronouns *me* or *us*.

The pronouns *I* and *we* are investigated in four English-language corpora: LOCNESS/ICLE-RU for argumentative writing and BAWE-LING/VESPA-RU for academic writing. The Russian pronouns *я* (=I) and *мы* (=we) are analysed in CRUAW and CRUMBA. Because of the time limitations of the investigation, I could not study all occurrences of the pronouns, and analysed samples of 300 for each pronoun and each corpus. The quantity of the pronouns *I* and *we* differs from corpus to corpus. Therefore, I extracted every fifth occurrence in ICLE-RU, every third in LOCNESS and every second in BAWE-LING (*cf.* Table 26). If the example was irrelevant – typically in case of citations or examples, I took the next relevant one. In this way, the sample is considered to be random, and the results of the investigation are not influenced by the style of the writer. Unfortunately, VESPA-RU and the Russian L1 corpora contain less than 300 examples. In such cases, I used all the available examples. I believe that the results of my study are reliable and represent the major tendencies among native speakers and the learners.

In the following four sections, I discuss the results of my in-depth functional investigation conducted separately for each of the pronouns in two registers. At the beginning of each section, the identified functions are discussed and illustrated with examples. The results are

presented in tables, and the overuse/underuse is calculated for each of the functions. The log-likelihood test ensures the statistical significance of the overuse/underuse.

## 5.1 The pronoun *I* in argumentative writing

### 5.1.1 Results

Table 37 presents the results of my functional analysis of the pronouns *I* and *я* (=I) in L1 English, L2 English and L1 Russian argumentative writing. The results are displayed in raw frequencies and in percentages per 300 samples. Unfortunately, CRUAW is a very small collection of Russian L1 argumentative essays, and it contains only 98 occurrences of the pronoun *я* (=I).

**Table 37.** The functions of the pronouns *I* and *я* (=I) in the sample of up to 300 instances from L1 English, L2 English and L1 Russian argumentative writing

Function	LOCNESS		ICLE-RU		CRUAW	
	Raw frequencies	% of sample	Raw frequencies	% of sample	Raw frequencies	% of sample
Expressing personal opinion:	180	60	194	65	97	99
- reinforce a statement	3	1	17	6	-	-
- weaken a claim	-	-	6	2	-	-
Organizing the text	14	4.7	43	14	0	0
Recounting personal experience	102	34	40	13	1	1
Generic <i>I</i>	1	0.3	-	0	0	0
<b>TOTAL</b>	<b>300</b>	<b>100</b>	<b>300</b>	<b>100</b>	<b>98</b>	<b>100</b>

As correctly pointed by Fossan (2011), the results presented in the form of percentages tell us about the proportion of the function among the samples. However, we cannot judge about the overuse/underuse based on the data from Table 37. In order to make the results directly comparable, I calculate the estimated raw frequencies and the estimated frequencies per 10,000 words. The percentages from Table 37 are applied to the overall frequency of the pronoun *I* in either LOCNESS or ICLE-RU to indicate an estimated raw frequency of the function in these corpora. For example, the function of recounting personal experience accounts for 34% of all occurrences of the pronoun *I* in LOCNESS, which is 982 instances (Section 4.1.1). By taking 34% of 982, we arrive at the estimated raw frequency of the function in NS writing, namely 333.88 instances.

Based on the estimated raw frequency and the total amount of words in the corpus, one can calculate the estimated frequency per 10,000 words, namely multiply 333.88 by 10,000 and

divide by 322,444, which is 10.35 instances per 10,000 words. The overuse/underuse in ICLE-RU is calculated by dividing the estimated frequency per 10,000 in ICLE-RU by the same value for LOCNESS.<sup>53</sup>

**Table 38.** The functions of the pronouns *I* and я (=I) in L1 English, L2 English and L1 Russian argumentative writing (estimated frequencies).

Function	LOCNESS		ICLE-RU		CRUAW		OVERUSE ICLE-RU / LOCNESS
	Estimated raw frequencies	Estimated frequency per 10,000 words	Estimated raw frequencies	Estimated frequency per 10,000 words	Raw frequencies	Frequency per 10,000 words	
Expressing personal opinion:	589.20	18.27	939.25	41.86	97	9.34	2.28
- reinforce a statement	9.82	0.30	86.7	3.86	-	-	11.98
- weaken a claim	-	-	28.90	1.29	-	-	-
Organizing the text	46.15	1.43	202.3	9.02	-	-	6.50
Recounting personal experience	333.88	10.35	187.85	8.37	1	0.1	0.83
Generic <i>I</i>	2.95	0.09	-	-	-	-	-
<b>TOTAL</b>	<b>982</b>	<b>30.44</b>	<b>1445</b>	<b>64.40</b>	<b>98</b>	<b>9.44</b>	<b>2.11</b>

Presenting the author's viewpoint is a key feature of argumentative writing. Therefore, it is logical that the function of expressing personal opinion is the most frequent in NS and learner writing. Here, the pronoun *I* is typically followed by such verbs as *feel*, *believe* and *think*:

(49) *I think* that there should be an age limit on fertility treatment. (LOCNESS)

(50) *I believe* it is impossible to imagine a state without prisons or any other organs of punishment. (ICLE-RU)

In learner writing, this function is often accompanied by the sub-function of overstatement (Aijmer 2001). Learners attempt to reinforce a statement and use strings of units with personal pronouns for this purpose:

(51) *Finally I must confess I can see no reason* to oppose that most university degrees are theoretical but I would find it difficult to accept that they don't prepare students for real world. (ICLE-RU)

(52) *It is obvious, I believe, that we have to* distinguish between imagination and dreaming. (ICLE-RU)

(53) *Me personally, I prefer* to watch TV rather than to go to a cinema or a theatre. (ICLE-RU)

<sup>53</sup> No statistical testing has been performed on the estimated frequencies.

However, even in NS writing, one can find the overstatement phrase *personally I*, though it is very infrequent: 0.09 instances per 10,000 words in LOCNESS against 0.67 instances in ICLE-RU. The other non-nativelike sub-function is to signal uncertainty or to weaken a claim. In this function, the pronoun *I* usually comes in the non-initial position:

- (54) Though nothing depends on my opinion, which is far from being authoritative, still the question is very acute and important, ***I would say***. (ICLE-RU)

*I* as organizer is the second most overused function. The author uses the pronoun *I* to signal the beginning or the end of an essay, or to cross-reference within the text. The pronoun *I* in his function does not express author identity, but is used to guide the reader through the text. The function is present both in NS and in learner writing. *I* as organizer is rather simple to identify due to the typical phrases that precede it: *to begin with*, *to sum up*, *to conclude* and others.

- (55) ***Therefore, I conclude*** that although the invention and development of the human computer has kept the brain on, full-time, its use has offloaded it, to a certain extent, into redundancy. (LOCNESS)

- (56) ***To begin with I'd like to point out*** the fact that religion tends to be a very difficult subject for a proper description. (ICLE-RU)

In Petch-Tyson's study (1998), the function of recounting personal experience is recognized as a typical NS use of the pronoun *I*. Here, authors appeal to personal experience to explain or support their point of view. Learners underuse the function, and the underuse is statistically significant. In example (57), the author refers to personal interests in order to underline the significance of the presented argument. In examples (58) and (59), the argument is supported by the information received from authors' friends and acquaintances.

- (57) ***I*** am not a boxing fanatic. ***I*** wouldn't even describe myself as an enthusiast. ***I*** have never done any boxing in my entire life but ***I*** would still be prepared to defend the sport of boxing against anyone wishing to ban it. (LOCNESS)

- (58) ***I***'ve got some friends of mine who served in this system. (ICLE-RU)

- (59) ***I*** know some people who were in the jail. (ICLE-RU)

Generic *I* is the least frequently used function of the pronoun *I*. In this meaning, the pronoun *I* personifies a larger group of people: starting from a class of students and up to the whole humankind. This function is the least frequent one with only one example registered in LOCNESS and none in ICLE-RU. The generic *I* can be substituted by such alternatives as *one*,

*we* or *you*. When the author chooses to use generic *I*, then s/he expresses an opinion and attribute it to some other people as well. In example (60), the pronoun *I* is used to refer to the author together with all other people who love to eat meat, and opposed to *you*, as the group of all those, who believe it is wrong.

- (60) Whatever ***you*** believe, ***I*** shall continue to enjoy my roast beef and Yorkshire pudding, thankyou. (LOCNESS)

### 5.1.2 Analysis

As shown in Table 38, Russian learners overuse the pronoun *I* to express their personal opinion, especially in the sub-function of reinforcing a statement, and to organize the text. At the same time, they underuse *I* in the native-like function of recounting personal experience. Russian learners overuse the pronoun *I* to express personal opinion at the rate of 2.28, which may result from L1 transfer, since this is the main function of the pronoun *я* (=I) in Russian L1 writing. However, the two other overused functions of organizing the text and reinforcing a statement are unlikely to be influenced by L1 conventions. Russian students do not make use of the pronoun *I* either to reinforce a statement or to organize the text. Teaching-induced factors can be a possible explanation of the overuse. Students are often encouraged by teachers and textbooks to make a better use of introductory phrases as they are supposed to make their writing more native-like. Little instruction is given as to how these introductory multi-word units should be embedded into the body of the text in a native-like manner. If we compare example (61) from LOCNESS to examples (62) and (63) from ICLE-RU, we can observe the difference in the use of the same multi-word unit by native speakers and by learners:

- (61) ***As far as I am concerned*** fox hunting has become a pointless and outdated sport, yet it still has many followers and in parts of the country hunts do continue. (LOCNESS)
- (62) So, ***as far as my opinion is concerned, I suppose*** that the army consisting of professional soldiers would be not only convenient for the country (both for the citizens and for se state ) but also more reliable and safe in the case of danger. (ICLE-RU)
- (63) ***As far as I am concerned I cannot help admiring*** such people. (ICLE-RU)

In NS writing, the sequence *as far as I am concerned* is directly followed by author's claim. In Russian NNS writing, it is often followed by other introductory units in the function of expressing personal opinion. It means that learners aim at using as many introductory phrases as possible, which results in strings of functionally identical units. Besides, learners often do not know how to integrate these units into the text. Thus, they start a sentence with an introductory

phrase, which is followed by an independent sentence. Such introductory units can be excluded from the sentence without any change in meaning. For instance, in example (62), the unit *as far as my opinion is concerned* can be excluded with no change in meaning, since *I suppose* expresses the same idea. Similar observations can be made with regards to the overuse of pronoun *I* in the function of organizing the text:

(64) So, <sup>1</sup>*to sum up*, <sup>2</sup>*I would like to say that it is evident for everybody* <sup>3</sup> that television has become now an integral part of our life, and we can hardly survive without it. (ICLE-RU)

Though the function of the pronoun *I* is not obvious here, I think it belongs together with the unit *to sum up* and is an organizer. Russian learners want to use the unit, but do not seem to know how it can be done in a native-like manner. Therefore, they connect it to the conclusion by means of two other units. Whereas in the NNS example, the unit is accompanied by two extra units: *I would like to say* and *it is evident for everybody*, in the NS example the unit *to sum up* is directly followed by the conclusion:

(65) <sup>1</sup>*To sum up*, fox hunting is a cruel, barbaric sport that provides the upper class with a bit of fun for a few hours, but at what cost. (LOCNESS)

The function of recounting personal experience is underused. According to the findings in CRUAW, Russian L1 argumentative writing does not refer to personal experience when reasoning for or against a statement. Only one example with this function is registered in CRUAW. All in all, the underuse in L2 writing is not high, and Russian learners seem to be successfully adopting the new reasoning method, *i.e.* appealing to personal experience in argumentative writing.

The findings in ICLE-RU correlate with the findings by Petch-Tyson (1998) in ICLE-FR, ICLE-SW and ICLE-DU. Learners with these backgrounds also overuse the functions of expressing personal opinion and organizing the text, and underuse the pronoun *I* when recounting personal experience.



## 5.2 The pronoun *WE* in argumentative writing

### 5.2.1 Results

Following the classification by Herriman (2009) discussed in Section 2.7.2, I distinguish three functions of the pronoun *we* in argumentative writing: generic *we*, inclusive *we* and exclusive *we*. Since the frequencies of the inclusive *we* and the exclusive *we* are not very high, I decide not to sub-divide the sample into specific and authorial sub-functions. However, I focus on identifying the functions similar to those of the pronoun *I*: recounting personal experience, organizing the text and expressing personal opinion. My analysis reveals certain correlations between the functions of the pronoun *I* and the pronoun *we* in argumentative writing. The results of the analysis are discussed below.

**Table 39.** The functions of the pronouns *we* and *мы* (=we) in the sample of up to 300 instances from L1 English, L2 English and L1 Russian argumentative writing

Function	LOCNESS		ICLE-RU		CRUAW	
	Raw frequencies	% of sample	Raw frequencies	% of sample	Raw frequencies	% of sample
Generic <i>we</i>	243	81	247	82	171	99
Inclusive <i>we</i>	18	6	30	10	2	1
Exclusive <i>we</i>	39	13	23	8	-	-
<b>TOTAL</b>	<b>300</b>	<b>100</b>	<b>300</b>	<b>100</b>	<b>173</b>	<b>100</b>

**Table 40.** The functions of the pronouns *we* and *мы* (=we) in L1 English, L2 English and L1 Russian argumentative writing (estimated frequencies).

Function	LOCNESS		ICLE-RU		CRUAW		OVERUSE ICLE-RU / LOCNESS
	Estimated raw frequencies	Estimated frequency per 10,000 words	Estimated raw frequencies	Estimated frequency per 10,000 words	Raw frequencies	Frequency per 10,000 words	
Generic <i>we</i>	757.35	23.49	1447.42	64.51	171	16.46	2.75
Inclusive <i>we</i>	56.10	1.74	175.80	7.84	2	0.19	4.50
Exclusive <i>we</i>	121.55	3.77	134.78	6.01	-	-	1.59
<b>TOTAL</b>	<b>935</b>	<b>29</b>	<b>1758</b>	<b>78.36</b>	<b>173</b>	<b>16.65</b>	<b>2.70</b>

It was not always straightforward to identify the function of the pronouns *we* and *мы* (=we) in a sentence. Some examples had mixed functions, whereas in other examples, an exact function could be identified only by consulting the author of the essay. Below, I describe the approach adopted for classifying the examples.

The generic function of the pronoun *we* is by far the most frequent in all the three corpora. It implies minimum presence of author in the text, and is used when writer wants to present an argument on behalf of all people. Sometimes, *we* can refer to a smaller group of

people united by some feature, as for example, people driving cars or living in the same country etc. In example (66), the pronoun *we* unites all Brits as a nation where fox-hunting is a tradition and opposes them to other nations where this ritual is prohibited by law. The pronoun *we* in example (67) denotes the people on this planet, or more precisely people using aerosols.

(66) The idea raises disgust from other countries, and in some places is laughable, so why do *we* British hang on to this cruel and heartless exercise. (LOCNESS)

(67) Moreover *we* are destroying the ozone layer by using aerosols and it is changing climate and causing cancer diseases. (ICLE-RU)

The generic *we* partly correlates with the most frequent function of the pronoun *I*. The generic *we* helps authors to reason for the arguments they support, and so does the pronoun *I* in the function of expressing personal opinion. In this function, the pronoun *I* is often a part of a multi-word unit (for example, *as far as I am concerned*), which is not typical of the generic *we*. My hypothesis can be indirectly confirmed by the similar frequencies per 10,000 words for the pronoun *I* expressing personal opinion and for the generic *we*: in LOCNESS, 18.27 and 23.49, and in ICLE-RU, 41.86 and 64.51 respectively. Comparing examples (61) and (66) shows that both examples express the same idea, *i.e.* that foxhunting is cruel and pointless, but authors apply different methods of supporting their argument. In example (61), the reasoning relies on the authority of the author, whereas in example (66), the writer generalizes and uses the practises of other countries as an argument.

The inclusive *we* unites the author and the reader, but excludes other people. This function correlates with the organizing function of the pronoun *I*. The overuse in ICLE-RU is similar and is high both in case of the inclusive *we* (4.50) and in case of *I* as organizer (6.50). The focus here is on guiding the reader through the text. Thus, in example (68) and (70), the inclusive *we* is used to summarize the discussion, and in example (69) – to show the development of the argument.

(68) Thus *we* see that the effects on the agricultural industry would be widespread. (LOCNESS)

(69) Before *we* discuss the case of ‘Baby Richard’, *we* might want to go back a couple of years to the case of ‘Baby Jessica’. (LOCNESS)

(70) On the other side, after deeper observation *we* can spot some similar circumstances that show how it can happen that both religion and TV have such a great influence on our everyday life. (ICLE-RU)

The exclusive *we* refers to a group of people that includes the writer. None of the texts in either LOCNESS or ICLE-RU was written by more than one author. Therefore, the exclusive *we* does not refer to the authors of the text, but is used to recount personal experience. For instance, it is used to refer to the author and her husband in example (71), or to the author and his classmates in example (72).

(71) *We* did [*author's note*: come downtown to identify some of the property] and they had recovered some of our items from the trunk of an old car. (LOCNESS)

(72) Not knowing where *we* would serve and what *we* would do *we* had to study various subjects. (ICLE-RU)

### 5.2.2 Analysis

All the three functions of the pronoun *we* are overused by Russian learners in argumentative writing. The main function of *we* in argumentative writing is undoubtedly the generic function. The overuse of the generic *we* can result from L1 transfer, as in Russian L1 argumentative writing, the pronoun *we* occurs almost solely in its generic function.

The inclusive *we* is the most overused function in ICLE-RU, and the overuse is not due to L1 transfer. I did register two samples of this function in CRUAW, but they account for less than 1% of all uses, and is, therefore, nearly non-existent.

The most unexpected results are registered for the exclusive *we*, which is proved to be overused by Russian learners at the rate of 1.59. No exclusive *мы* (=we) is registered in L1 Russian argumentative writing. This indicates that recounting personal experience is not common in Russian NS student writing. The overuse is not high, but it indicates that Russian learners appeal to their personal experience to support an argument in L2 writing.

Comparing my results to the findings by Fossan (2011) reveals a number of similarities. Norwegian learners overuse the generic *we* at the rate of 2.99, which is very close to the rate of 2.75 registered in Russian learner writing. Both the inclusive *we* and the exclusive *we* are overused by these learner groups. However, Norwegian learners strongly overuse the exclusive *we* and slightly overuse the inclusive *we*. These results are opposite to the findings for Russian learners, where the inclusive *we* is highly overused, whereas the exclusive *we* is hardly overused at all. I can hypothesize that using personal experience to support an argument is acceptable and widely used in Norwegian argumentative writing, which could prompt the overuse of this function in L2 writing.

### 5.3 The pronoun *I* in academic writing

#### 5.3.1 Results

For my investigation of the pronoun *I* in academic writing, I adopt Hyland's functional classification (2002a) with the addition of one function – organizing the text. In Russian learner writing, the pronoun *I* is most often used in its low-risk function of explaining a procedure. *I* is least frequent in the most high-risk function of stating results/claims. Below, the functions are presented and discussed according to their frequency in VESPA-RU.

Tables 41 and 42 do not include the results for academic L1 Russian writing due to the very low frequency of the pronoun *я* (=I) in the corpus. I think that presenting the results based on three hits would be misleading.

**Table 41.** The functions of the pronoun *I* in the sample of up to 300 instances from L1 and L2 English academic writing.

Function	BAWE-LING		VESPA-RU	
	Raw frequencies	% of sample	Raw frequencies	% of sample
Explaining a procedure	123	41	92	71
Elaborating an argument	84	28	13	10
Stating a purpose	42	14	12	9
Organizing text	6	2	10	8
Stating results/claims	45	15	3	2
<b>TOTAL</b>	<b>300</b>	<b>100</b>	<b>130</b>	<b>100</b>

**Table 42.** The functions of the pronoun *I* in L1 and L2 English academic writing (estimated frequencies).

Function	BAWE-LING		VESPA-RU		OVERUSE ICLE-RU / LOCNESS
	Estimated raw frequencies	Estimated frequency per 10,000 words	Raw frequencies	Frequency per 10,000 words	
Explaining a procedure	227.55	11.53	92	32.11	2.79
Elaborating an argument	155.40	7.88	13	4.54	0.58
Stating a purpose	77.70	3.94	12	4.19	1.06
Organizing text	11.10	0.56	10	3.49	6.63
Stating results/claims	83.25	4.22	3	1.05	0.25
<b>TOTAL</b>	<b>555</b>	<b>28.13</b>	<b>130</b>	<b>45.37</b>	<b>1.61</b>

Explaining a procedure is a low-risk function according to Hyland (*ibid.*). Compared to high-risk functions, this use of *I* implies less personal exposure. However, this does not make this function less significant: a profound description of research procedures is one of the key

features of a successful academic paper. This is the most frequent function both in native speaker and in learner writing. Both student groups have a good command of this function, and use *I* willingly here. Russian learners overuse *I* in this function at the rate of 2.79. Students use *I* to recount the choices they make (see example 73) and the material they use (see example 74):

- (73) In each word class ***I chose*** a selection of frequencies, i.e. high and lower frequencies. (BAWE-LING)
- (74) For the collection of data ***I have used*** COCA, Corpus of Contemporary American English, and the BNC, the British National Corpus. (VESPA-RU)

Elaborating an argument is a high-risk function, which involves the responsibility of the authors for their arguments. In both corpora, the function is the second most frequent, though Russian learners are proved to underuse it. In Anglo-American rhetoric, authors should show responsibility for their arguments and express agreement or disagreement with a position. Here, the pronoun *I* is typically followed by cognitive verbs such as *think*, *feel*, *believe* and others. In example (75) and (76), students present their observations, and use the pronoun *I* to show their commitment to the ideas.

- (75) ***I think*** what can be seen through the research that has been done is that there is a clear link between class (incorporating status and prestige) with gender. (BAWE-LING)
- (76) At the first sight ***I thought*** that those questions expressed polarity, but the use of ‘want’ in each of those questions is a bit confusing and ***on a second thought I'd rather say*** that they express an offer. (VESPA-RU)

Stating a purpose is a low-risk function that is slightly overused by Russian learners. When using *I* in this function, authors do not expose themselves to possible criticism from the reader's side. It is often used at the beginning of a paper or a section to introduce the topic (see example 77) or to state the intentions of the text (see example 78):

- (77) In this essay ***I am going to*** concentrate on the role of prestige and how this affects the way we speak. (BAWE-LING)
- (78) Thus ***I will look*** at how Theme is defined by SFL and FSP, whether there are any counterparts of Marked Theme in FSP [...]. (VESPA-RU)

When classifying the samples from BAWE-LING and VESPA-RU, I felt that the five of Hyland's (*ibid.*) functions were not enough, and some examples did not fit any of these functions. The pronoun *I* is often used by students for cross-referencing within the text.

Therefore, I introduce this extra function into my classification. *I* as organizer is overused at the high rate of 6.63.

(79) *As I have discussed*, findings have shown that there can be 'exceptions' as to where the lesion is located and the type of aphasia found. (BAWE-LING)

(80) During the research process *I* came across some problems, which *I described earlier*. (VESPA-RU)

The other high-risk function is stating results and claims. The presence of the author is critically important here, as the main purpose of every investigation is to yield results. Student writers may be tempted to avoid the responsibility for their findings by adopting various grammatical structures without personal pronouns. This function is especially problematic for Russian learners, who are less aware of the conventions of English academic writing. They underuse the function at the rate of 0.25. The function of stating results can be observed in the first two uses of the pronoun *I* in example (81) and the second use in example (82). These are not very explicit examples of this function, which is probably due to the students' lack of academic experience. Hyland's study (*ibid.*) gives examples from professional academic writing, and they are more sophisticated and persuasive.

(81) In this project, *I have shown* that although online news sites share some features with broadsheet newspapers *I also identified* some shared features that the evidence *I gathered* from the BNC corpus suggests are part of language use in general. (BAWE-LING)

(82) Now that *I have studied* and *compared* the two varieties of the English language, *I can say that* the PSF classification pattern remains dominant overall and it is even slightly more frequent in the BrE. (VESPA-RU)

Expressing self-benefits is the last function in Hyland's classification (*ibid.*), who discusses it as a typical learner use of the pronoun *I*. It is used to recount students' positive experiences when working on the project. This function is not present in NS writing, and only one example is present in Russian learner writing. Since this function is very infrequent, I do not include it in Tables 41 and 42. I think that frequencies that are based on one instance cannot be considered reliable and can lead to confusion. The one example I extracted from VESPA-RU is presented below:

(83) All in all, *I* have had a very interesting and at times exciting experience while working with the corpora. (VESPA-RU)

Expressing self-benefits may be a culture-determined function of the pronoun *I*. If Hong Kong academic conventions imply that students explain what they personally have gained from the study, it may result in the transfer of the L1 function into the interlanguage. However, this is not the case for Russian academic conventions.

### 5.3.2 Analysis

My analysis reveals that Russian learners overuse three and underuse two of the functions in L2 academic writing. The underuse is registered in the two high-risk functions of elaborating an argument and stating results and claims. The overuse is registered in the three low-risk functions: explaining a procedure, stating a purpose and organizing the text.

Unfortunately, it is not possible to compare my results to Hyland's results (*ibid.*). Hyland presents his results in percentages as a total for all self-mentioning pronouns (*I, me, my, mine* and *we, us, our, ours*), and I investigate the pronouns *I* and *we* separately. Besides, the use of percentages is rather controversial, as it is more important to look at relative frequencies in the text and identify overuse/underuse, which can be used for analyses.

The overuse of *I* to explain a procedure has a value of 2.79, which is rather high, especially taking into account the fact that the function constitutes 71% of all uses of *I* in Russian learner writing. The use of *I* to organize a text is overused at a rate of 6.63. However, we can see a decrease in the use of this function compared to the results from argumentative writing. The decrease of 2.5 times is registered both in NS and in learner writing. In LOCNESS, *I* is used to organize a text with an estimated frequency of 1.43 instances per 10,000 words, and it is 2.55 times less often in BAWE-LING – 0.56 instances per 10,000 words. In ICLE-RU, estimated frequency of *I* as organizer is 9.02 instances per 10,000 words, and it drops down to 3.49 instances per 10,000 words in VESPA-RU.

## 5.4 The pronoun *WE* in academic writing

### 5.4.1 Results

*We* in academic writing is probably the most challenging, but at the same time interesting pronoun to investigate in this chapter. The reason is that I could not decide on the classification to apply here, as Hyland (*ibid.*) uses the same functions to analyse *I* and *we* in academic writing, and Fossan (2011) adopts Herriman's classification (2009) of *we* from argumentative writing to academic writing. Both of these classifications provide interesting perspectives on the pronoun

*we*. To decide on the classification, I identify the function of each example first according to Hyland's classification (2002a) and then according to Herriman's classification (2009). It turns out that these classifications complement each other and give a deeper insight if discussed together. Therefore, in this section, I take Hyland's classification (2002a) as the basis and comment on which type of *we* (generic, inclusive or exclusive) is common to express the function. At the end of this section, the values for Herriman's classification (2009) of *we* are presented in corresponding tables.

The tables below present the distribution of various functions of *we* in academic writing. In contrast to я (=I) in the previous section, the Russian L1 writing contained enough examples of мы (=we) to provide a comparison with BAWE-LING and VESPA-RU.

**Table 43.** The functions of the pronouns *we* and мы (=we) in the sample of up to 300 instances from L1 English, L2 English and L1 Russian academic writing

Function	BAWE-LING		VESPA-RU		CRUMBA-LING	
	Raw frequencies	% of sample	Raw frequencies	% of sample	Raw frequencies	% of sample
Elaborating an argument	204	68	48	44	21	36
Organizing text	27	9	19	18	3	5
Stating results/claims	27	9	16	15	13	22
Stating a purpose	15	5	15	14	19	32
Explaining a procedure	27	9	10	9	3	5
<b>TOTAL</b>	<b>300</b>	<b>100</b>	<b>108</b>	<b>100</b>	<b>59</b>	<b>100</b>

Table 44 presents the values, which are more meaningful for this study, *i.e.* the rate of the overuse and the estimated relative frequencies.

**Table 44.** The functions of the pronouns *we* and мы (=we) in estimated frequencies in L1 English, L2 English and L1 Russian academic writing

Function	BAWE-LING		VESPA-RU		CRUMBA-LING		OVERUSE VESPA-RU / BAWE- LING
	Estimated raw frequencies	Estimated frequency per 10,000 words	Raw frequencies	Frequency per 10,000 words	Raw frequencies	Frequency per 10,000 words	
Elaborating an argument	208.08	10.54	48	16.75	21	9.73	1.59
Organizing text	27.54	1.40	19	6.63	3	1.39	4.74
Stating results/claims	27.54	1.40	16	5.58	13	6.02	3.99
Stating a purpose	15.30	0.77	15	5.23	19	8.80	6.79
Explaining a procedure	27.54	1.40	10	3.49	3	1.39	2.49
<b>TOTAL</b>	<b>306</b>	<b>15.51</b>	<b>108</b>	<b>37.68</b>	<b>59</b>	<b>27.33</b>	<b>2.43</b>



When analysing samples from VESPA-RU, I find all five functions of *we*. All the functions are overused by Russian learners, but at different rates. The functions are discussed in a descending order according to their frequency in VESPA-RU.

The function of elaborating an argument is overused at a rate of 1.59. The pronoun *we* is more frequent in this function than the pronoun *I*. The major difference between the NS and Russian learner writing is that native speakers typically use the generic *we* to elaborate an argument, whereas Russian learners alter for the inclusive *we*:

- (84) However plasticity lessens as *we* grow into adulthood with the brain becoming more stable and the ability to learn language being lost. (BAWE-LING)
- (85) In our text one *we* have several examples of ellipsis. (VESPA-RU)

As in all the three previous sections, *we* as organizer is overused at a high rate. Both in NS and learner writing, *we* in this function is mainly inclusive. It is used to refer to something that has previously been mentioned or will be mentioned later in the text. *We* as organizer is often followed by verbs as *mention*, *see*, *discuss* and others.

- (86) Linguistically this is not the case *as we have seen* through the discussion of accent and dialect. (BAWE-LING)
- (87) Also, *as we have seen earlier*, the distribution of the PP is significantly higher in the BrE (8% vs. 3%). (VESPA-RU)

The overuse of 3.99 is registered in the function of stating results and claims. As in the previous function, *we* is typically used in its inclusive meaning. The author includes the reader in drawing the conclusions, and thus shares the responsibility for them.

- (88) A general *statement we can make* from this data, is that it seems Spanish is the language preferred in more formal domains, whilst Guarani is the language preferred in more informal domains. (BAWE-LING)
- (89) *Overall we see that* the Past Progressive has scored 13% in BrE, which is a significant number if compared to the results from COCA, representing AmE (4%). (VESPA-RU)

The function of stating a purpose is overused. In this function, *we* indicates the direction of research and the development of an argument. Here, the majority of *we* is used with inclusive meaning and is followed by verbs referring to the future:

- (90) *We will now elaborate on* gender and its effect on speech. (BAWE-LING)

(91) To prove the hypothesis *we will try to* find the examples of cohesion, 'the linguistic devices by which the speaker can signal the experiential and interpersonal coherence of the text' (*ibid.*) and express coherent meanings. (VESPA-RU)

Explaining a procedure is the least frequent function of the pronoun *we* in VESPA-RU. In case of the pronoun *I*, this function, on the contrary, has the highest frequency in Russian L2 academic writing. The function is overused at the rate of 2.49. In the examples below, *we* refers to the author and the reader, but not to collective authorship of two or more writers.

(92) In order to discover whether they are correct, *we must examine* the criteria of human language and use them to assess the chimpanzees' abilities. (BAWE-LING)

(93) In order to see certain transitivity patterns *we need to analyse* verbs / verbal groups appearing in the text. (VESPA-RU)

#### 5.4.2 Analysis

The main finding in the present section is that the two high-risk functions are among the top three most frequently used functions of the pronoun *we*. This contrasts with my findings for the pronoun *I* in academic writing, which has highest frequencies in low-risk functions. This means that Russian learners assign the leading academic role to the pronoun *we*. Another observation concerns the L1 transfer of function. High overuse is registered in the two functions of stating results/claims and stating a purpose. The overuse seems to be promoted by the functions *мы* (=we) fulfils in the Russian language. High overuse is also registered in *we* as organizer, the function that has a lower frequency in Russian L1 writing than in L2 writing. Here, the overuse is not prompted by L1 conventions, but is rather a general learner issue. Similar overuse is registered by Fossan (2011) in Norwegian learner writing.

My hypothesis about the functional L1 transfer can be observed more clearly in the other classification of the pronoun *we*. The inclusive *we* is overused at the rate of 4.62. Besides, the function accounts for 76% of all the uses in VESPA-RU. The transfer of L1 function is the most likely reason for the overuse here, since it is definitely the dominant function in the Russian L1 writing, and estimated frequency per 10,000 words is similar in the two Russian corpora. The other functions are underused.

**Table 45.** Generic/inclusive/exclusive *we* and *мы* (=we) in the sample of up to 300 instances from L1 English, L2 English and L1 Russian academic writing

Function	BAWE-LING		VESPA-RU		CRUMBA-LING	
	Raw frequencies	% of sample	Raw frequencies	% of sample	Raw frequencies	% of sample
Generic <i>we</i>	171	57	25	23	2	3
Inclusive <i>we</i>	120	40	82	76	56	95
Exclusive <i>we</i>	9	3	1	1	1	2
<b>TOTAL</b>	<b>300</b>	<b>100</b>	<b>108</b>	<b>100</b>	<b>59</b>	<b>100</b>

When analysing examples according to the two classifications, I noted that Russian learners use inclusive *we* to elaborate an argument, whereas native speakers majorly use generic *we* in this function. Therefore, the main pedagogical implication is that Russian learners need to be made aware of how an argument can be elaborated with the help of generic *we*.

**Table 46.** Generic/inclusive/exclusive *we* and *мы* (=we) in L1 English, L2 English and L1 Russian academic writing (estimated frequencies).

Function	BAWE-LING		VESPA-RU		CRUMBA-LING		OVERUSE VESPA-RU / BAWE- LING
	Estimated raw frequencies	Estimated frequency per 10,000 words	Raw frequencies	Frequency per 10,000 words	Raw frequencies	Frequency per 10,000 words	
Generic <i>we</i>	174.42	8.84	25	8.73	2	0.93	0.99
Inclusive <i>we</i>	122.40	6.20	82	28.62	56	25.95	4.62
Exclusive <i>we</i>	9.18	0.47	1	0.35	1	0.46	0.75
<b>TOTAL</b>	<b>306</b>	<b>15.51</b>	<b>108</b>	<b>37.7</b>	<b>59</b>	<b>27.34</b>	<b>2.43</b>

### 5.5 Concluding remarks on the qualitative analysis of author identity

Both in argumentative and in academic writing, Russian learners overuse the first person pronouns *I* and *we* to organize the text. The comparison of learner writing and Russian L1 writing shows that the overuse of the function is not induced by L1 transfer. Besides, the overuse of this function is also registered in Norwegian learner writing (Fossan 2011). Therefore, I assume that the use of author pronouns as text organizers is a problem common for different learner varieties. The reasons for such tendency among learners require further investigation. However, I can hypothesize that the main reason lies in teaching-induced factors, *i.e.* textbooks initiating learners to use introductory phrases. Little instruction is given on how to use such phrases, which results in chaining of several units at the beginning of the sentence.

The qualitative analysis supports my hypothesis that the ‘we’ perspective is leading in Russian learner academic writing. Russian learners overuse the pronoun *I* in low-risk functions, whereas the pronoun *we* is overused in high-risk functions. Another finding is that Russian learners use the inclusive *we* to elaborate an argument, whereas native speakers use the generic *we* for this purpose.

The present chapter provides evidence regarding the influence of L1 functions on L2 writing. The effect of L1 transfer may cause the overuse of *I* to express personal opinion and the underuse of *I* to recount personal experience in argumentative writing. The functional L1 transfer leads to the overuse of the generic *we* in argumentative writing and the inclusive *we* in academic writing. To achieve nativeness, learners should use the pronouns *I* and *we* in native-like functions.

## 6. RESULTS AND ANALYSIS: REASONS FOR OVERUSE AND UNDERUSE OF AUTHORIAL PRONOUNS IN ENGLISH L2 WRITING

The previous chapters showed that Russian learners construct their L2 author identity in a different manner than native-speakers. In this chapter, I put forward possible explanations to account for these differences. Following the study by Gilquin and Paquot (2008), I distinguish four possible explanation: the effect of L1 transfer, teaching-induced factors, the effect of developmental factors and the spoken medium influence. In what follows, I section by section consider in detail each of these effects in order to identify the major ones.

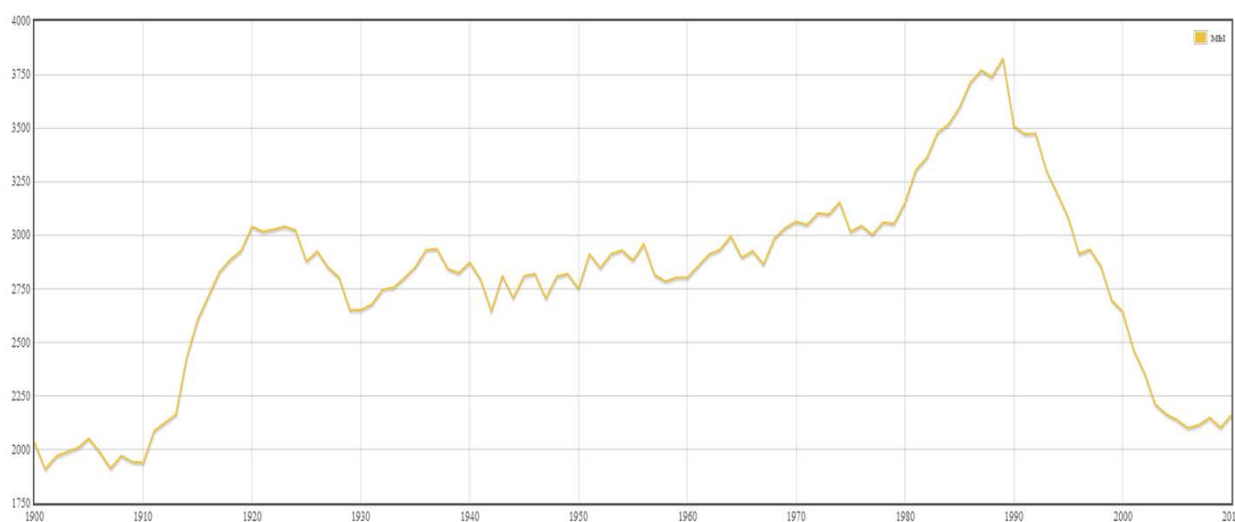
### 6.1 The effect of L1 transfer

L1 transfer seems to have a significant effect on the way learners construct author identity in interlanguage. Throughout Chapter 4, I have paid attention to the units that are highly overused by Russian learners. Sometimes these units lead to the overall overuse of a personal pronoun. Another observation in Chapter 4 is that Russian learners overuse first person plural pronouns (the ‘we’ perspective) at a higher rate than first person single (the ‘I’ perspective). Furthermore, the qualitative analysis reveals that the pronoun *we* fulfills high-risk functions, whereas the pronoun *I* performs low-risk functions in academic writing. I assume that the deviation results from the effect of L1 transfer. This section aims at comparing the results from L2 writing to Russian L1 writing, in order to test my assumption.

An interesting observation on author identity in Russian L1 writing can be made based on Figure 13 presented below. It illustrates that the use of the first person pronoun *мы* (=we) increases drastically at around 1915 and grows up until mid-1980s and then drops continuously. These tendencies correspond to political situation in Russia. In 1918, the Communist Party took over the country and pronounced the common ownership of production means. It is of major interest for this study that the ‘we’ perspective became dominant over the ‘I’ perspective, which continued during the whole rule of the Communist Party until 1989. One of the well-known sayings of the time was that ‘Не я-кай, “я” – последняя буква алфавита’ (translation: Do not say ‘I’, as ‘I’ (= Russian ‘я’) is the last letter of the alphabet. After the fall of the Communist regime, I register a decrease in the use of the pronoun *we* (=мы). These observations are made on the basis of the RNC corpus, which is a mixed corpus of speech, literature, journalism and

other types of writing. The results from RNC cannot be directly compared to the results from this study, as writing conventions take longer time to change than, for example, speech, especially when it concerns academic writing. I still think that this data reflects the general tendency and can be useful for explaining some of the findings presented in this chapter.

**Figure 13.** Historical perspective on the use of the pronoun *мы* (=we) in Russian L1 writing in RNC.



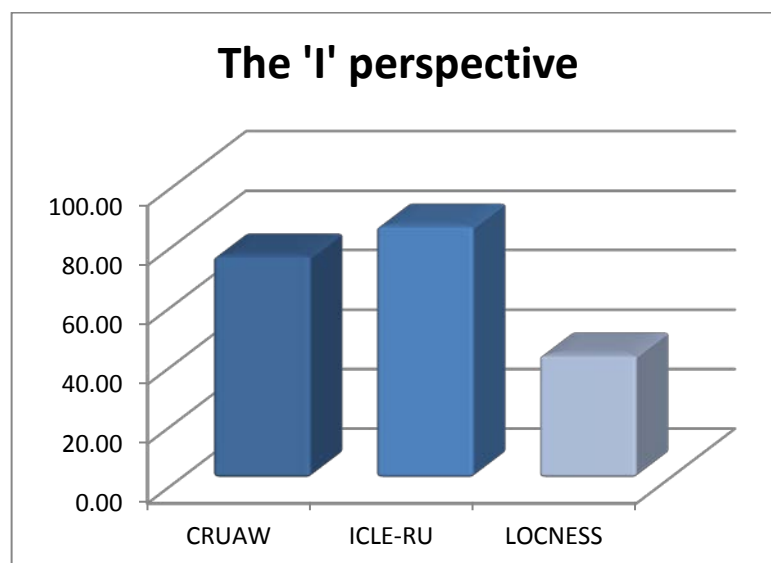
In the following two sections, I investigate the influence of L1 transfer in argumentative and academic writing. The data from LOCNESS and ICLE-RU is compared to Russian L1 writing presented by my own collection of argumentative essays, and the data from BAWE-LING and VESPA-RU is compared to my own collection of Russian L1 student academic writing.

In the sections on L1 transfer, I discuss author identity in terms of the ‘I’ perspective and the ‘we’ perspective. The terminology is adopted from the study by Vassileva (1998), who refers to first person single pronouns as the ‘I’ perspective and first person plural pronouns as the ‘we’ perspective. I believe that these terms illustrate more clearly the main focus of these sections, as I mainly discuss the group of the pronouns *I*, *me*, *my* and *mine* as opposed to the other group of the pronouns *we*, *us*, *our* and *ours*. In other words, I contrast the collective mind and the individual mind settings.

### 6.1.1 The effect of L1 transfer in argumentative writing

The material I use to study the effect of L1 transfer is discussed in detail in Section 3.2.5. The analysis of the material with the AntConc corpus tool reveals the following results (*cf.* Figure 14):

**Figure 14.** The 'I' perspective in CRUAW, ICLE-RU and LOCNESS. Frequency per 10,000 words



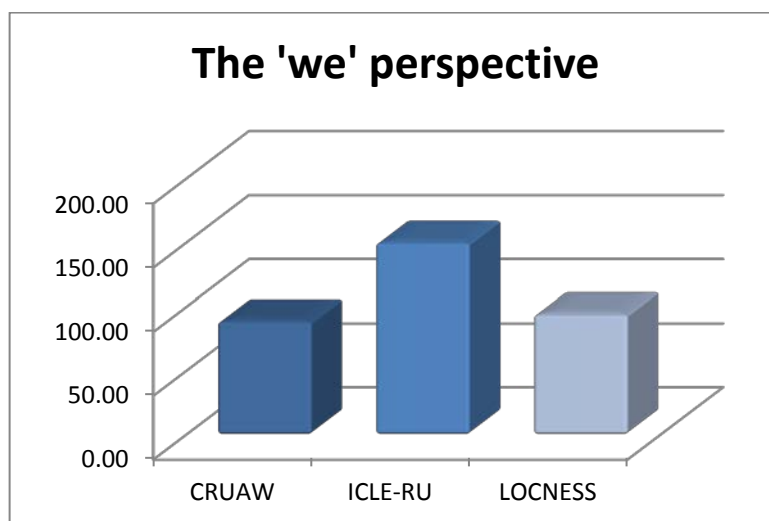
As shown in Figure 14, the use of the 'I' perspective is only slightly higher in ICLE-RU than in CRUAW, and about double of what is registered in the English NS corpus. The difference between the Russian L1 and English L1 writing is statistically significant.<sup>54</sup> Thus, I can assume that the L1 transfer takes place in Russian L2 argumentative writing. Though the overuse in ICLE-RU compared to CRUAW is very slight (at the rate of 1.14), it is still statistically significant.<sup>55</sup> This means that L1 transfer is probably not the only reason for the learners' overuse of first person single pronouns.

Figure 15 presents the findings for the 'we' perspective in English L1, English L2 and Russian L1 argumentative writing. As discussed in Section 5.2, not all instances of *we* refer to the writer, but could have been used in its generic function. The findings for the 'we' perspective are not as clear-cut and raise a number of questions. First person plural pronouns are less frequently in Russian L1 writing than in English L1 and L2 writing. This could indicate that the overuse of the 'we' perspective in L2 writing does not result from the effect of L1 transfer. However, we should not forget about the data in the Figure 13, which shows the drastic decrease in the use of the 'we' perspective in the Russian language after the fall of Communism. The essays in ICLE-RU date back to the 1990s, whereas the essays in CRUAW were written between 2010 and 2012. This could potentially explain such a distinct difference in the use of first person plural pronouns in ICLE-RU and CRUAW.

<sup>54</sup> Statistical significance – 99.99%.

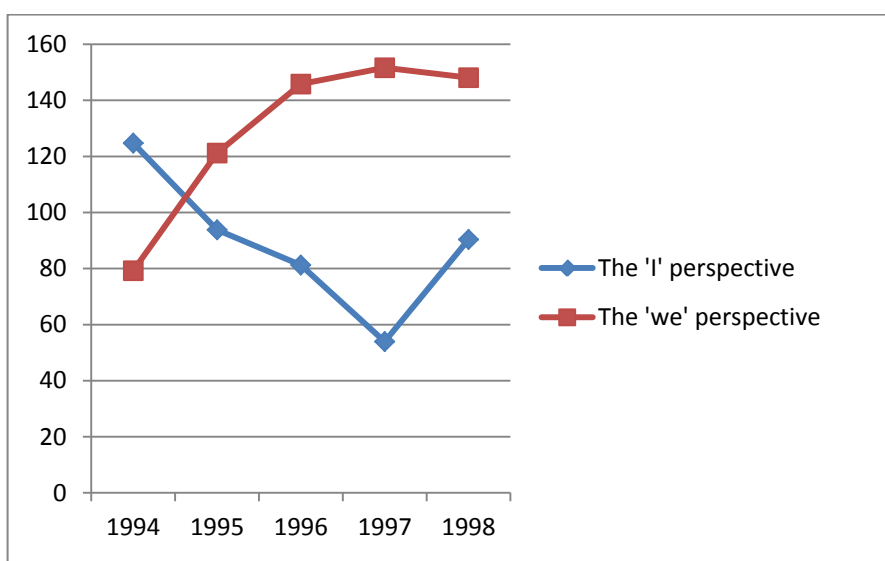
<sup>55</sup> Statistical significance – 99.99%.

**Figure 15.** The 'we' perspective in CRUAW, ICLE-RU and LOCNESS. Frequency per 10,000 words



To the best of my knowledge, ICLE-RU is the latest corpus of Russian L2 writing, which makes it complicated for me to test my hypothesis properly. However, to make some estimate about the use of the 'I' and the 'we' perspective in the 1990s, I decided to investigate the use of first person pronouns in the ICLE-RU texts dating between 1994 and 1998 year by year. According to ICLE 2, the amount of texts varies from 21 texts in 1997 and 23 texts in 1994 up to 111 texts in 1998. I think that this amount of texts is enough to achieve reliable results. Besides, the log-likelihood test proves the results to be statistically significant.<sup>56</sup> The year-by-year analysis of the texts reveals the following results:

**Figure 16.** The 'I' and the 'we' perspective from 1994 until 1998 in ICLE-RU. Frequency per 10,000 words



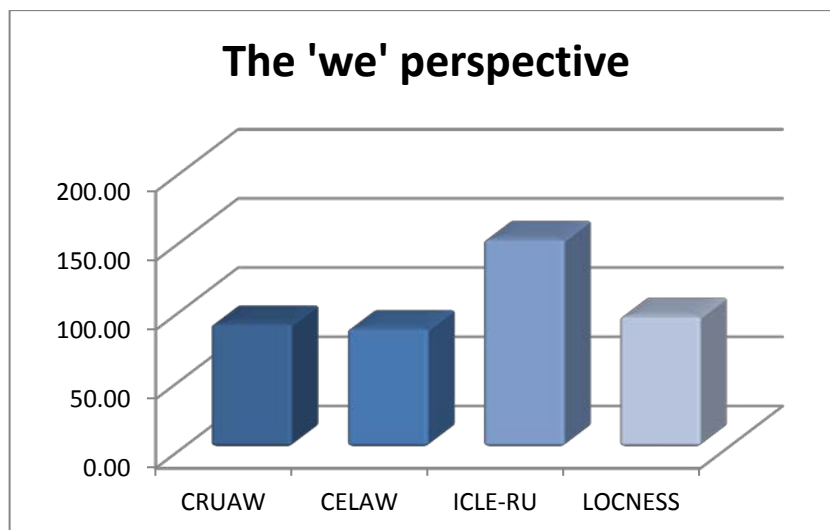
<sup>56</sup> Statistical significance varies from 95% to 99.99% depending on the years in question



According to Figure 16, the use of the ‘we’ perspective was increasing and reached its peak in 1997. In 1998, I register a slight decrease. The use of the ‘I’ perspective, on the contrary, was decreasing steadily and was at its lowest in 1997. The year 1998 showed a strong change in that tendency, and the ‘I’ perspective reached the rate of 1995. The results may reflect the political and economic situation in the country. In 1998, a severe economic crisis struck Russia, whose government announced the default on its debts and devaluated the rouble. Many Russians lost their money and found themselves alone in the fight for their well-being. People could not anymore rely on the social system or the state. I assume that after the events in 1998, Russians were more willing to make statements using the ‘I’ perspective.

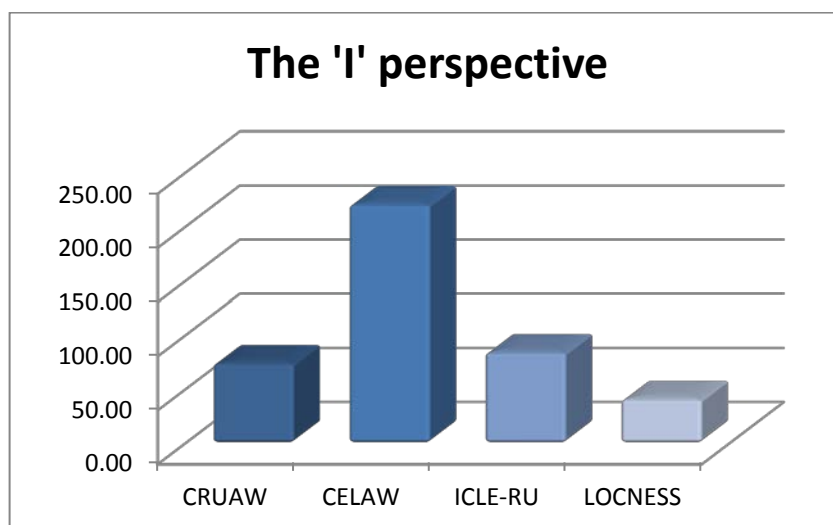
Unfortunately, I have no data from 1998 up until 2010 and cannot analyse the further development during this period. However, I was able to collect a sample of Russian L2 writing between 2010 and 2013 (*cf.* Section 3.2.7). The observations made on the basis of CELAW offer only a brief insight as to how the use of the ‘we’ perspective might have changed. However, these results should be treated with great caution due to the size and the origin of CELAW. The analysis of the ‘we’ perspective in this collection of essays reveals the following results:

**Figure 17.** The ‘we’ perspective in L1 Russian, L1 and L2 English argumentative writing. Frequency per 10,000 words



My moderate collection of argumentative essays suggests that since the turning point in 1998, the ‘we’ perspective experienced a decline. Figure 17 indicates that L2 learner writing reflects the tendency of L1 writing, which emphasises the role that L1 transfer plays in the L2 overuse/underuse. The same investigation was carried out for the ‘I’ perspective (*cf.* Figure 18):

**Figure 18.** The 'I' perspective in L1 Russian, L1 and L2 English argumentative writing. Frequency per 10,000 words

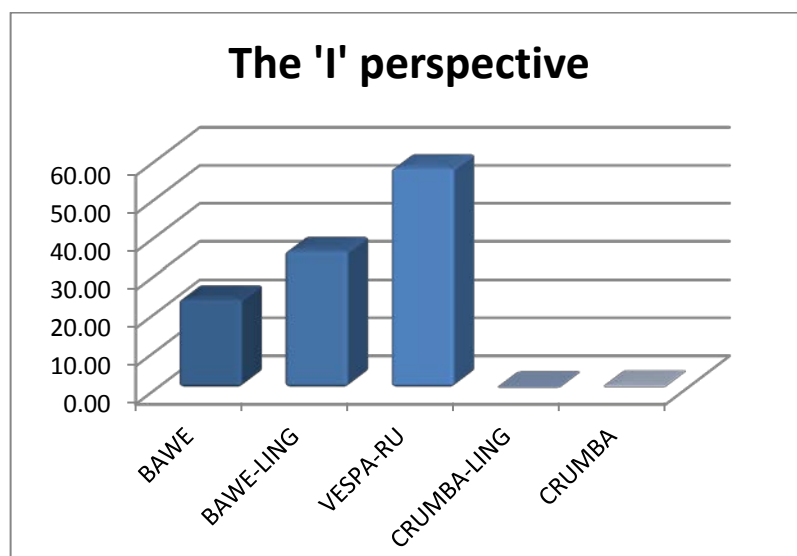


The results in Figure 18 indicate that the 'I' perspective is strongly overused in CELAW compared to the Russian L1 writing (CRUAW). The overuse cannot be explained from the point of view of L1 transfer and is probably the result of other effects. I hypothesize that the main influence comes from the teaching-induced factors, as in the 1990s, textbooks available to Russian learners had focus on grammar, and few of them prompted the use of introductory phrases. During the last decade, the focus has shifted, as many learners have to write argumentative essays within the framework of international English language exams. The new teaching material contains lists of introductory units, which result in the chaining of sentence-initial features of W/R visibility (Petch-Tyson 1998: 114) and may lead to the overuse of the pronoun *I*. However, the limitations of this study do not allow me to investigate this hypothesis further.

### **6.1.2 The effect of L1 transfer in academic writing**

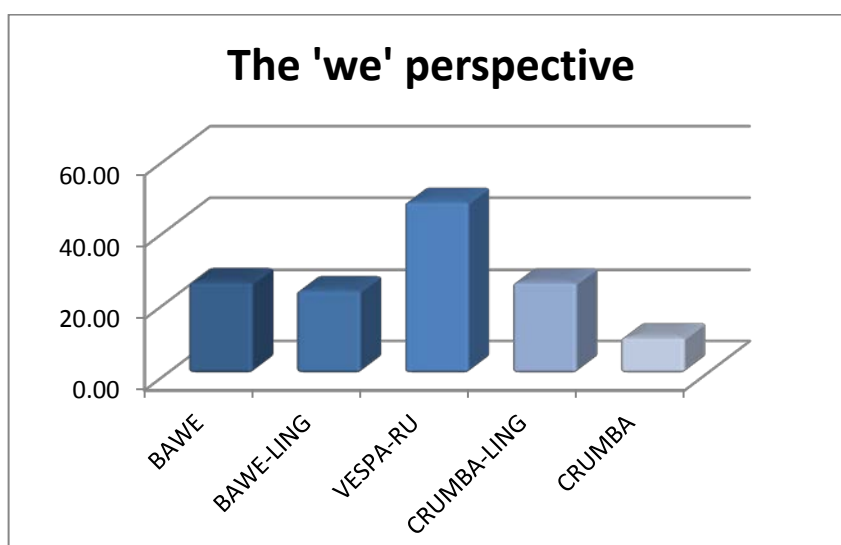
The material used to analyse Russian L1 academic writing is described in detail in Section 3.2.6. The results of the study are presented in Figure 19, which indicates that the 'I' perspective is practically absent in Russian L1 academic writing. These results support the findings by Vassileva (1998), who reports the minor role of the 'I' perspective in Slavic academic writing. At the same time, Figure 19 indicates the overuse of first person single pronouns in the English L2 writing. Thus, the overuse of 'I' perspective in academic writing is not caused by L1 transfer and could be a result of other influences.

**Figure 19.** The 'I' perspective in L1 Russian, L1 and L2 English academic writing. Frequency per 10,000 words



According to Figure 20, the 'we' perspective is significantly more frequent in L1 Russian academic writing. The rate of its use is very close to what is registered in the NS writing. However, the 'we' perspective in L2 English writing is about twice as frequent as it is in L1 Russian and L1 English academic writing. At first sight, the results in Figure 20 seemed confusing to me, since my original hypothesis had been that the dominant role of the 'we' perspective would result from L1 transfer of academic conventions.<sup>57</sup> However, the next section on developmental factors reveals significant variation in the use of the 'we' perspective in bachelor and master learner writing, which partly explains the results in Figure 20.

**Figure 20.** The 'we' perspective in L1 Russian, L1 and L2 English academic writing. Frequency per 10,000 words



<sup>57</sup> The 'we' perspective is proved to be dominant according to the quantitative and qualitative studies, presented in Section 4.2.9 and Section 5.4 respectively.

## 6.2 Developmental factors

The second possible explanation for the overuse/underuse is the influence of developmental factors. The term developmental factors is understood by Gilquin and Paquot (2008) as a feature that does not necessarily attribute the overuse/underuse to non-nativeness of the author, but to the lack of academic experience. Learners may not have a good command of academic written language or may not be too familiar with academic conventions.

To study the influence of developmental factors, I separately analyse the bachelor and master assignments in the VESPA-RU corpus. To distinguish between the former and the latter, I refer to them as VESPA-RU-BA and VESPA-RU-MA. The bachelor part of the corpus contains three works, whereas the master part has seventeen works. Unfortunately, the material is not of the same size, but the results are presented in relative frequencies per 10,000 words, which makes them comparable. The frequencies from the VESPA corpus are analysed in the context of L1 Russian and L1 English academic writing.<sup>58</sup> To make the corpora more comparable, I divided the L1 Russian writing into bachelor and master assignments. The results are presented for the linguistic component of CRUMBA, as it is more comparable to VESPA-RU and BAWE-LING.

The effect of L1 transfer was not apparent in the previous section, but the perspective of developmental factors sheds some light on the data. Figures 21 and 22 indicate that the results from the bachelor part of VESPA-RU reflect the tendencies of L1 Russian academic writing: the 'I' perspective is hardly present, whereas the 'we' perspective is dominant. Figure 21 reveals that bachelor students make the first steps towards adopting the NS academic conventions, as the 'I' perspective emerges in their L2 writing. The analysis of the master assignments, on the other hand, does not reveal the apparent effect of the L1 transfer. Learners at this stage seem to have gained a good command of the written language and are knowledgeable about the English academic conventions. The results from Figures 21 and 22 show significant change in the way master students express their author identity. They make more use of the 'I' perspective and even overuses it. At the same time, the use of the 'we' perspective decreases and is about to reach the NS writing rate.

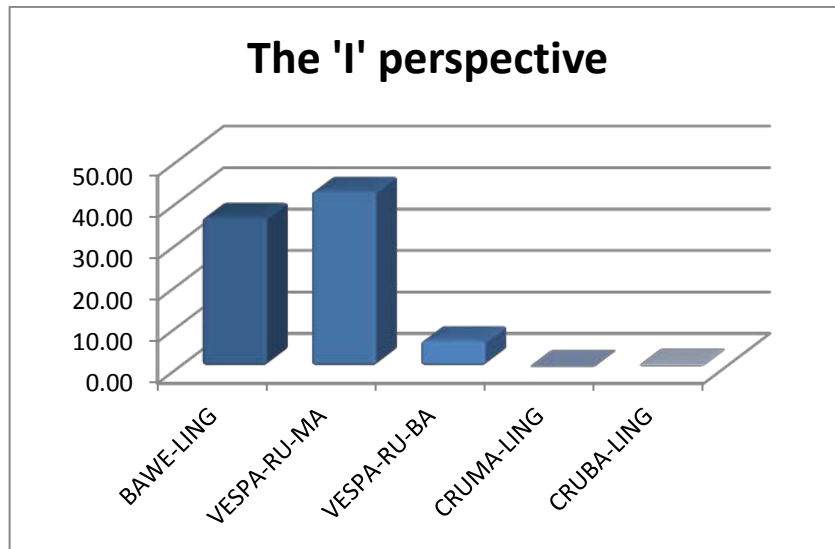
Thus, developmental factors have influence on L2 English writing. Bachelor student writing experiences the strong impact of their L1 academic writing conventions. Here, writers strongly underuse the individual voice of the author and overuse the collective voice. However, the development of academic skills changes the way authors express their identity. In master

---

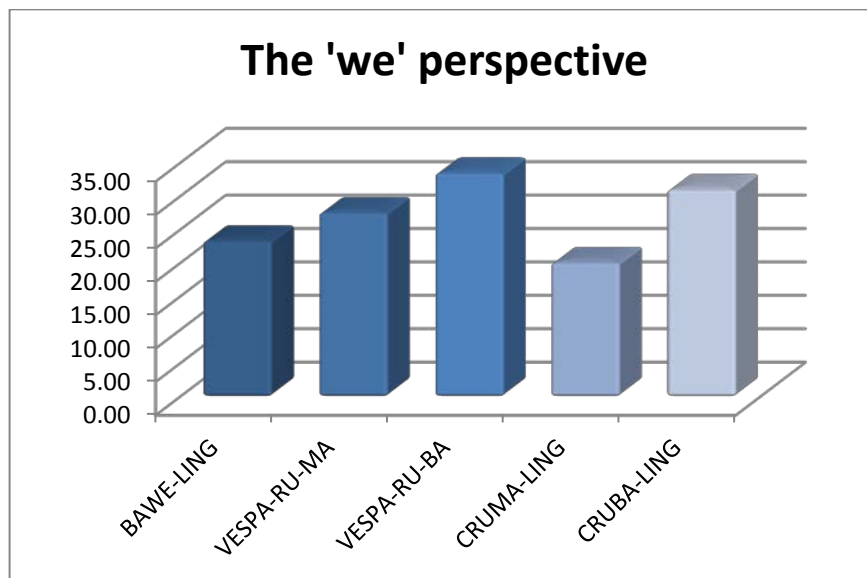
<sup>58</sup> The frequencies should be read with some caution due to the small size of the VESPA-RU-BA corpus.

assignments, students are more willing to put forwards their individual identity. Here, there is less influence of L1 transfer. Master students continue to overuse the ‘we’ perspective, though at a lower rate than bachelor students.

**Figure 21.** The ‘I’ perspective in the master and bachelor L1 Russian, L2 and L1 English academic writing. Frequency per 10,000 words



**Figure 22.** The ‘we’ perspective in the master and bachelor L1 Russian, L2 and L1 English academic writing. Frequency per 10,000 words



All in all, developmental factors, or more precisely academic experience, have its influence on L2 student writing. Through the years of education, students read professional academic writing, exercise in writing themselves, receive feedback from their supervisors and

visit conferences. I believe that such experience helps students to abstract from L1 conventions and adapt to new English writing practices.

### 6.3 Teaching-induced factors

In this section, I analyse three textbooks that focus on developing students writing skills.<sup>59</sup> Two of the textbooks focus specifically on academic writing, whereas one textbook provides training in argumentative writing for the CAE (*Certificate of Advanced English*) Paper 2 – ‘Writing’. In these textbooks, I was looking for explanations on how authors should portray their identity in the text, express their personal opinions or use first person pronouns. The findings for argumentative writing are presented first and are followed by the findings for academic writing.

The textbook *Advanced writing with English in use. CAE.* helps students in improving their skills in writing argumentative essays. One of the units in the book is dedicated to the strategies of reasoning and expressing personal opinion. The analysis of the unit reveals that the majority of the advice and exercises is not tailored for L2 learners of English. The book takes up the general issues of writing an argumentative essay. For example, it advises to think through the subject before writing, to group ideas together into paragraphs and arrange them in a coherent order:

*‘In this type of writing, the way ideas are connected is very important; this is achieved by the use of connecting words, by the way the ideas are grouped together into paragraphs, and by the way the paragraphs themselves are ordered’ (Hugh 1999: 69)*

*‘Your opinion will not be interesting or worth expressing until you have thought about the subject. So the first stage in writing your opinion happens in your head, while you get your ideas straight’ (ibid: 72)*

*‘[...] it is necessary to look at the product’s bad points as well as its good points.’ (ibid.)*

The advice is very similar to what NS students receive in Academic Writing Guides. One of the examples could be *The DEV Guide to Academic Writing* supplied by the University of East Anglia:

*‘Try to identify and list key themes and use this as a starting point for preparing an outline for your assignment.’ (ibid.)*

---

<sup>59</sup> Hugh, C. (1999). *Advanced writing with English in use. CAE.* Oxford: Oxford University Press.  
Jordan, R. R. (1999). *Academic writing course: Study skills in English.* Edinburgh: Longman.  
Hamp-Lyons, L. and B. Heasley. (2006). *Study writing: A course in written English for academic purposes.* Cambridge: Cambridge University Press.

*'Your argument should also be balanced – it should consider different points of view and take into account evidence for and against a particular position.'* (ibid.)

*'Writing your assignment is a multiple-step process - organising your ideas and information, preparing an outline[...] and shaping your text (paragraphing).'* (ibid.)

Students undoubtedly require some guidance in how to structure an argument. However, learners have more profound and specific needs, as their aim is not only to write a successful essay, but also to write it in the most native-like manner. Especially students with Russian cultural background, whose academic conventions are rather different from the English ones, would benefit from a more detailed explanation of how they should construct author identity in writing. Nevertheless, the textbooks give only one general comment on author identity in writing – it should be impersonal. It is often presented in the form of exercises, where learners are asked to make a personal argument into a more impersonal one:

*'The text below contains many cases of personalisation. Rewrite the text so that it is less personal, but still reads as an argument.'* (Hamp-Lyons and Heasley 2006: 111)

*'This film review is mainly the writer's opinion, but he never says "I" or "in my opinion". How does he convey such a strong opinion without using the first person?'* (Hugh 1999: 73)

*'Many students spoil their articles and reviews by writing in a chatty, informal style as if they were talking to a friend. On the contrary, essays, articles and reviews should be relatively impersonal. Your readers are not particularly interested in you: they need information, description and narrative more than they need your opinion.'* (ibid.: 74)

Learners receive a signal that personal pronouns are not appropriate in writing. However, my preliminary study shows that NS writing uses first person pronouns both in argumentative and academic writing. Hyland (2002a) illustrates that the pronoun *I* is used in NS professional writing in high-risk functions.

A different feature that unites the NS writing guide and the learners' textbook, is that both present students with a list of cohesive devices that can be used to structure an argument: *in my opinion, to my mind, I would like to* and others. Gilquin and Paquot (2008) conclude that learners are likely to deduce that connecting words are interchangeable and mix the register. Therefore, some teaching methods that are appropriate for native speakers may not be good enough for learners. The lack of instruction leads to the chaining of sentence-initial features of W/R visibility (Petch-Tyson 1998: 114), which results in the overuse of first person pronouns (*cf.* Section 4.1.5) and the use of the pronouns in a non-nativelike function of overstatement (*cf.* Section 5.1.1). My study also reports that the overuse of first person pronouns can be prompted by the frequent use of the prefabs, which result from L1

transfer. In Russian learner writing, one of such prefabs is the unit *let us* (cf. Sections 4.1.6 and 4.2.6). The overuse of *let us* is also registered in French learner writing and results from L1 transfer (Paquot 2008).

All in all, learners' textbooks present a simplified version of NS writing guides. They teach students that an argument should be balanced and well-structured. Author identity is discussed in terms of incongruity of personalization and the pronoun *I*. However, Russian learners make more use of the 'we' perspective than the 'I' perspective in academic writing. Thus, textbooks' instructions will only prevent learners from achieving nativeness, as Russian learners should, on the contrary, aim to use more of the 'I' perspective and less of the 'we' perspective. Besides, prefabs with first person pronouns should be discussed in more detail. As accurately pointed out by Granger (1998), we should provide learners with language-specific EFL materials and not try to aim at all learner groups at once. My preliminary study supports Granger's finding (*ibid.*) that L1 influences the use of prefabs in interlanguage. For example, in the L2 writing of Russian learners, the overuse of *as for me* is prompted by L1 unit *что касается меня, то* (=as for me) and the overuse of *let us* is prompted by the typical use of first person plural imperative in L1 writing conventions. Textbooks should take learners' mother tongue into account. Writing aid that addresses the learners in the same way as the native speaker cannot help L2 students achieve native-like performance.

#### **6.4 The influence of speech**

Granger (1998) points out that learners have a tendency to borrow oral strategies in their writing. Therefore, the influence of speech is one of the possible explanations of the overuse of first person pronouns in learner writing. Students first master the oral communicative strategies. At a later stage, they start adopting writing conventions, but continue to rely on what they know best, namely the oral conventions. This is especially true with regards to the use of the pronouns *I* and *we*, which are often used as agents in spoken language.

On the other hand, to estimate the influence of speech, one should evaluate the possible sources of NS speech in Russia. In everyday life, learners have little opportunity to listen to a native speaker talk or to come in personal contact with a native speaker. The television and radio stations broadcast solely in the Russian language. English movies and TV programs are always dubbed. Due to the Internet, learners have a chance to watch movies in the original or listen to English-language radio. However, this contact with L1 environment is limited. The main speech input comes in the form of teaching instructions or study dialogues between learners. The input



typically comes from a NNS source. Such little exposure to NS speech is not likely to have great influence.

When analysing learner writing textbooks, I made a surprising observation: some examples of writing presented in the textbooks include spoken-like features such as contractions:

*'I'm often told I'm lucky to be a man/woman, but in fact it's a mixed blessing.'*

(Hugh 1999: 69)

The examples are given in textbooks to create a model that learners can aspire to. The fact that such examples include contractions can be misleading for learners. They can interpret it as a signal that English writing conventions are close to speech.

To sum up, I do not believe in the strong influence of speech in Russian learner writing. The exposure to English is rather low compared to countries such as Norway, where a lot of broadcasting is in English. Learners can introduce *I* or *we* as an agent in their writing due to it being a simpler structure. Besides, some confusion can result from the examples that are designed to show learners how they should write in English, but include informal markers as discussed above.

### **6.5 Concluding remarks on the reasons for overuse and underuse of authorial pronouns in English L2 writing**

The fact that learners construct their author identity differently from native speakers cannot be explained through the influence of one single factor. All the four factors discussed in this chapter play their role in the way Russian learners present their author identity in argumentative and academic writing. Besides, Sections 6.1.1 and 6.1.2 illustrate how dramatic changes in the history of a nation can affect the way people write. During just over a decade, Russian argumentative writing conventions experienced a major change: the use of the 'we' perspective decreased and gave way to the more individually oriented 'I' perspective. The academic writing conventions are typically more persistent to changes, and the 'we' perspective continues to be dominant there.

According to my findings, L1 transfer is the main factor that influences learners' author identity in writing. By L1 transfer, I mean the transfer of frequencies and register. Due to the difference in the conventions of argumentative and academic writing in the Russian language, I am able to observe the L1 transfer of register. Another significant factor here is the

developmental factor. My findings illustrate that with the increase of academic experience, learners gain a better understanding of the L2 English writing conventions.

The effect of teaching-induced factors and the influence of speech are difficult to evaluate and require a further in-depth study. However, I believe that students are in need of language-specific textbooks that would set the focus on the problematic areas for students with the same L1 background. For Russian learners, special attention needs to be paid to the differences in the use of the 'I' and 'we' perspectives in their mother tongue and the target English language. Academic writing requires most attention, since here, the differences between these perspectives is the greatest. Argumentative writing undergoes changes and reveals more common features with English argumentative writing. Concerning the influence of speech, I want to remark that Russian L2 students experience little exposure to the spoken language in their everyday life. Therefore, I am inclined to think that this factor has least influence on the L2 writing of the Russian students.

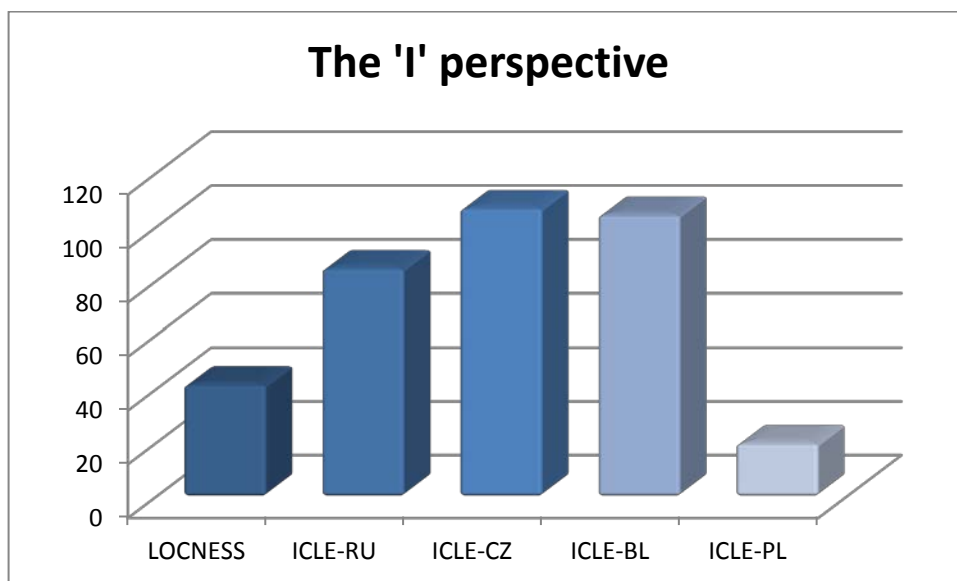
## **7. RESULTS AND ANALYSIS: AUTHOR IDENTITY IN L2 WRITING OF LEARNERS WITH SLAVIC L1 BACKGROUND**

Previous chapters studied quantitatively and qualitatively the use of author identity in Russian learner writing. The results report that the ‘we’ perspective is dominant in L2 writing of learners with Russian background. Chapter 6 indicates that the deviation may result from the transfer of L1 conventions into the interlanguage. The aim of this chapter is to put the findings for Russian L2 writing in the context of other Slavic learner varieties. The two main sections are discussing argumentative and academic writing. In argumentative writing, the results are based on the data from the ICLE corpus for three learner groups: Bulgarian, Czech and Polish. In academic writing, the observations are made based on the VESPA corpus for two varieties: Czech and Polish; since Bulgarian L2 writing is not present in VESPA. The results are presented in the form of charts, as my main focus is on illustrating the main tendencies, and exact values are of less importance. The charts show separately the results for the ‘I’ perspective and the ‘we’ perspectives. The former includes the first person singular pronouns *I*, *me*, *my* and *mine*, and the latter consists of the first person plural pronouns *we*, *us*, *our* and *ours*.

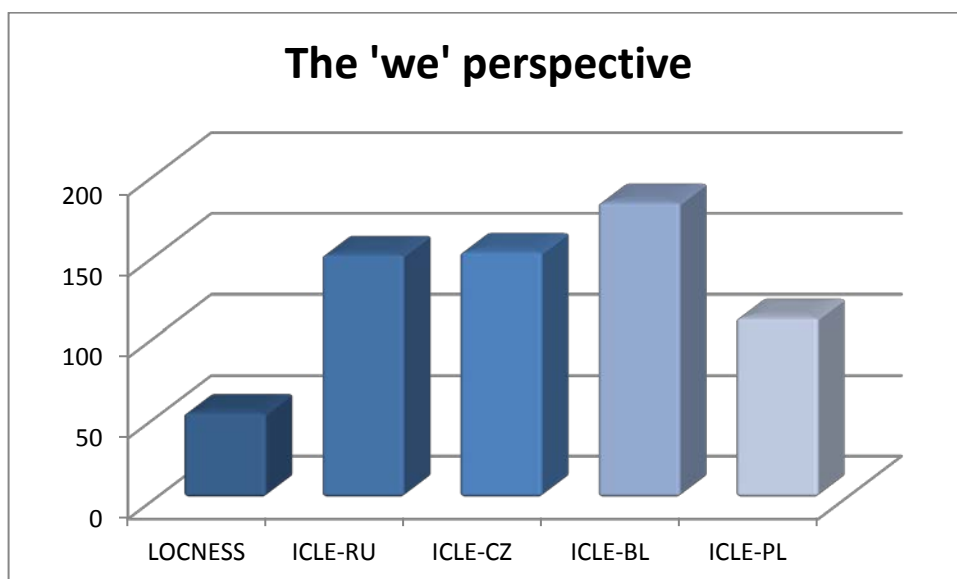
### **7.1 Argumentative writing of Slavic learners**

According to Figures 23 and 24, learners with Slavic L1 background share certain common features in English L2 argumentative writing. With the exception of Polish learners, all the Slavic varieties overuse the ‘I’ perspective at a rather high rate. The overuse of the ‘we’ perspective is even higher, and is registered in all four varieties, including Polish, Bulgarian, Czech and Russian. The overuse is found to be statistically significant according to the log-likelihood test. First person pronouns are least frequent in the Polish L2 writing, and are most frequent in Czech and Bulgarian writing. Russian L2 writing takes a middle position among other learner varieties in the use of the ‘I’ and ‘we’ perspectives.

**Figure 23.** The use of the 'I' perspective in NS and Slavic learner argumentative writing. Frequency per 10,000 words



**Figure 24.** The use of the 'we' perspective in NS and Slavic learner argumentative writing. Frequency per 10,000 words



## 7.2 Academic writing of Slavic learners

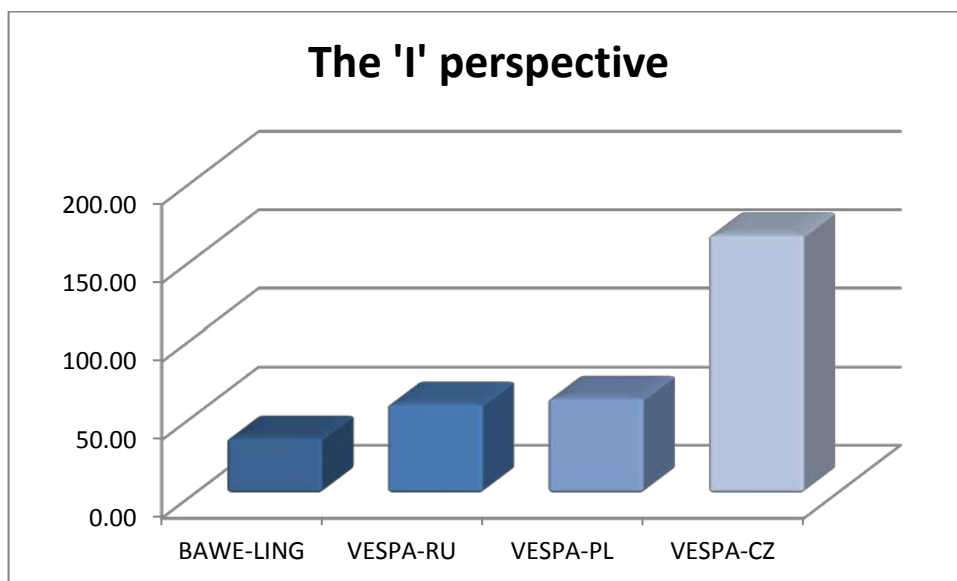
Except for the Russian learner writing, VEPSA contains texts from two more Slavic learner groups: Polish and Czech. The results in this section should be treated with caution due to the small size of the corpora (*cf.* Section 3.2.7). Russian, Polish and Czech learners overuse the ‘I’ perspective in academic writing, and the log-likelihood test shows that the overuse is statistically significant.<sup>60</sup> Similarly to argumentative writing, Czech learners overuse the ‘I’ perspective stronger than other learners varieties, but in academic domain the overuse is strikingly high. The time limitations of this study do not permit me to investigate further the reasons for such overuse.

Figure 26 illustrates the findings that are most relevant for the present study. What strikes the eye here is that Russian learners make significantly higher use of the ‘we’ perspective than the other Slavic learner groups. As discussed in Section 5.4 and 6.1.2, the overuse is most likely to be the result of L1 transfer. It is difficult to explain low frequencies of the ‘we’ perspective in Polish and Czech learner writing, especially taking into account the reader-oriented nature of Slavic academic writing (Čmejrková 1996). This result in the Polish learner writing is probably due to the overall low level of involvement, which is also registered in argumentative writing and in the ‘I’ perspective in academic writing. I think more investigation is required to explain the low frequency of first person plural pronouns in Czech learner academic writing. Comparing the values from Figure 24 and Figure 26 reveals that in argumentative writing Czech learners make significantly higher use of the ‘we’ perspective, than in academic writing. Further study of Czech learner writing is required to account for these observations.

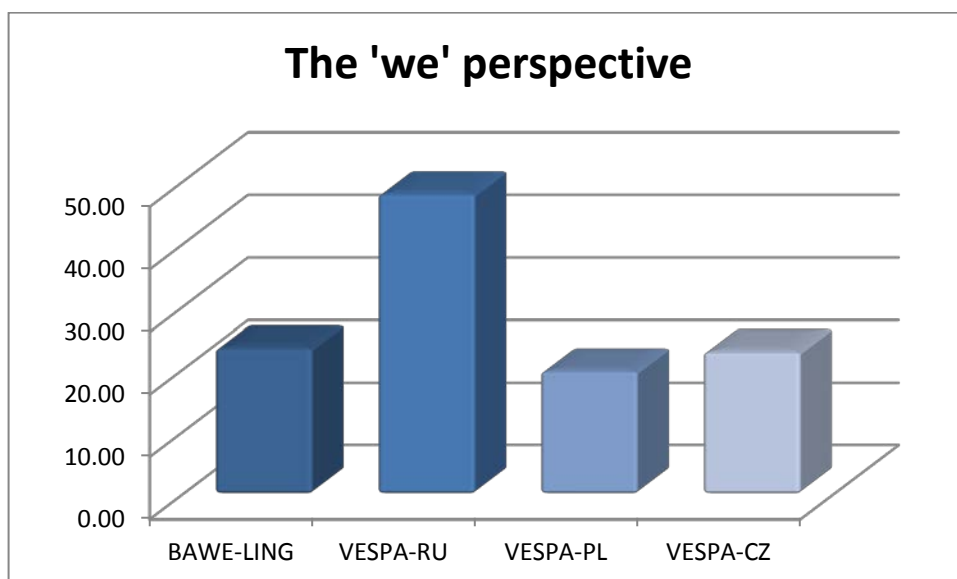
---

<sup>60</sup> Statistical significance of 99.99% in Russian and Czech writing, and 99% in Polish writing.

**Figure 25.** The use of the 'I' perspective in NS and Slavic learner academic writing. Frequency per 10,000 words



**Figure 26.** The use of the 'we' perspective in NS and Slavic learner academic writing. Frequency per 10,000 words



### **7.1 Concluding remarks on author identity in Slavic learner writing**

Slavic learner writing is characterized by a rather high level of W/R visibility in argumentative writing. The 'we' perspective is strong here, and the 'I' perspective is significantly overused as well. Polish learners demonstrate a slightly lower tendency to overuse first person pronouns. However, even they overuse the 'we' perspective in argumentative writing. A significant drop in the use of authorial pronouns is registered in academic writing. The overuse of the 'we' perspective by Russian learners stands out and does not seem to reflect the general tendency among Slavic languages. I should mention, however, that both Poland and Czech Republic take a middle position geographically, politically and socially between Russia and the Western European countries. During the Soviet times, people in these countries had more contact with the Western world than people in Russia. Therefore, academic conventions may have experienced less influence of the communistic ideology. It could be interesting to analyse the L2 writing of Ukrainian and Belarusian learners, as these countries underwent similar ideological changes to what people experienced in Russia. Unfortunately, these learner varieties are not present either in ICLE or in VESPA.

I think that though languages belong to the same Slavic group, they still vary in the way they express author identity. Many factors are responsible for the formation of the writing conventions including linguistic, historical and political factors. The L2 writing is not a direct reflection of L1 conventions, though, as shown in Sections 5.1.1 and 5.1.2, the effect of L1 transfer can be a major reason for overuse/underuse. I would assign the dominance of the 'we' perspective in academic writing to a combination of the above-mentioned factors and partly to the Slavic origin of the Russian language.

## 8. CONCLUSION

The aim of this study was to investigate how Russian learners construct their author identity in English L2 writing. A combination of quantitative and qualitative research methods was applied to analyse Russian learner writing from various perspectives. The results of this study indicate that Russian students overuse the ‘we’ perspective at a higher rate than the ‘I’ perspective, which is likely to be the result of L1 transfer. The overuse of the ‘we’ perspective is among the highest registered in learner writing. The ‘I’ perspective is overused at a more moderate rate. The overuse is likely to result from teaching-induced and developmental factors. The next section discusses the key findings of this study in more detail. Section 8.2 points out the pedagogical implications of these findings. The final Section 8.3 suggests the topics for further research.

### 8.1 Summary of the findings

The preliminary quantitative study in Chapter 4 indicates that Russian learners share a number of features with other learner varieties. In argumentative writing, such features include the overuse of all first person pronouns, subjective stance markers (*I think, to my mind, it seems to me, I would like to* etc) and the chaining of sentence-initial features of W/R visibility. This study also reveals some problems typical of Russian learners only. The major one is the dominance of the ‘we’ perspective over the ‘I’ perspective. The overuse of the ‘we’ perspective in argumentative writing is among the highest registered in L2 writing. The ‘I’ perspective is overused at a more moderate rate and takes a middle position between the overuse registered in the writing of French and Dutch learners, and Scandinavian learners. Section 4.1 demonstrates that the overuse often is an effect of L1 transfer. The frequencies of Russian L1 units are reflected in English L2 writing. Together with frequencies, we observe a transfer of function, register and form. Russian learners introduce L1 devices for constructing author identity in L2 writing. In case of positive transfer, it leads to strong overuse of units (*as for me; let us/let’s; I would like to*), and in case of negative transfer, it results in the use of a non-nativelike construction (*as far as my opinion is concerned, as for me, I*).

In the academic writing, the overuse rate is lower than in the argumentative genre, which is a general tendency for various learner groups. The pronoun *me* is underused, but the underuse is not statistically significant. The registered overuse is lower than in the writing of French and Norwegian learners (Paquot, Hasselgård and Ebeling 2013), but is far from the underuse registered in the reports of Hong Kong students (Hyland 2002a). The ‘we’ perspective is less frequent here than in argumentative writing, but the overuse rate remains rather high. Benefiting



from academic experience, Russian learners manage to adopt Anglo-American rhetoric and distribute the use of the 'I' and the 'we' perspectives in a more native-like way.

The qualitative study confirms that the 'we' perspective is dominant in L2 writing. In Hyland's (2002a) terms, Russian learners overuse the pronoun *I* in low-risk functions, whereas the pronoun *we* is overused in high-risk functions. It supports the findings of the preliminary quantitative study, which indicates a high overuse rate and frequency of first person plural pronouns. The results of the analysis highlight the influence of L1 functions on L2 writing. The effect of L1 transfer leads to overuse of the pronoun *I* to express personal opinion, underuse of the pronoun *I* to recount personal experience and overuse of the generic *we* in argumentative writing. In academic writing, it results in overuse of the inclusive *we* and underuse of the generic and exclusive *we*.

Another finding of the functional analysis is that Russian learners overuse the pronouns *I* and *we* as part of text-organizers. The overuse of this function has also been registered for Norwegian learners (Fossan 2011). It is likely that this is a general learner issue, but it is difficult to claim anything as little research has been completed in this field. The results of the study indicate that the registered overuse does not result from L1 transfer. I can hypothesize that the reason for the overuse lies in teaching-induced factors, *i.e.* textbooks encouraging learners to use more introductory phrases. The preliminary study indicates that it often leads to the chaining of sentence-initial introductory units.

Four main reasons for the overuse include the effect of L1 transfer, developmental factors, teacher-induced factors and the influence of speech. The study reports that the first three factors are likely to have most impact on the learner writing, whereas the last factor is not applicable to Russian students. The effect of L1 transfer is noticeable in argumentative writing. The overuse of the 'I' perspective in ICLE-RU reflects the tendency registered in Russian L1 writing. The results for the 'we' perspective are slightly ambiguous. The comparison of ICLE-RU and Russian L1 writing does not reveal the presence of L1 transfer. My assumption is that Russian L1 writing was in the process of changing after the collapse of the Soviet Union and the economic crisis. The L2 essays dating between 2010 and 2012 show a more moderate use of the 'we' perspective, and the effect of L1 transfer is more obvious there.

The effect of L1 transfer is only partly registered in the academic genre. Despite the absence of the 'I' perspective in L1 Russian academic writing, learners still make use of it in L2 assignments. The 'we' perspective is overused at a high rate, and the overuse is partly due to the L1 academic conventions. The study of developmental factors sheds some light on the use of the

'we' perspective in L2 academic writing. Russian bachelor students apply the 'we' perspective to a greater extent than Russian master students. This tendency is registered in Russian L1 as well as English L2 writing, which indicates the effect of L1 transfer. The 'I' perspective is underused by bachelor learners and overused by master learners. With academic experience comes a better understanding of Anglo-American rhetoric and decreases the overuse/underuse of authorial pronouns.

Teaching material is not completely tailored for learners. Textbooks for learners give the same advice as the guides to academic writing that are aimed at native speakers. The focus is on structuring the text and organizing the exposition. Such advice may be sufficient for native speakers, as they have a good command of their own language and an understanding of writing conventions. Learners with a different cultural background require more detailed instruction on how author identity functions within the framework of Anglo-American rhetoric. The 'I' and the 'we' perspective should be discussed in terms of the frequency and the function they fulfil in the argumentative and academic genre. Lists of introductory phrases are supplied in guides for native speakers and in the textbooks for learners. English L1 writers benefit from such lists, which may help them diversify their discourse and make it sound more academic. Learner writing often suffers from the incompetent use of introductory phrases. It leads to the overuse of first person pronouns in non-nativelike function and the chaining of sentence-initial features of author identity. A more in-depth analysis of teaching-induced factors is presented in the section on pedagogical implications.

The influence of speech is likely to have least impact on the learner writing of Russian students. The exposure to English is rather slight compared to countries such as Norway, where a lot of broadcasting is in English. Learners may of course borrow oral strategies, as this is what they learn first and know best. Therefore, it is especially important that textbooks do not send confusing signals in the form of contractions (*can't, I'm, he's*) and other informal markers in the examples of writing that students are exposed to and aspire to.

The study of other Slavic learner varieties is a rather small-scale investigation due to the scarcity of material. It reveals that the 'we' perspective is strong in L2 writing of Slavic learners, especially in the argumentative genre. Compared to the results from Petch-Tyson (1998), Slavic learners overuse the 'we' perspective at a higher rate than Dutch, French and Finnish learners, but at a lower rate than Swedish learners. Norwegian learner writing uses first person plural pronouns at about the same rate as the Slavic learner group (Fossan 2011). In the use of the 'I' perspective, Slavic learners take the middle position. They use it more often than Dutch and French learner groups, but less often than Swedish, Finnish and Norwegian learners. The only

exception is the Polish learner group, as they do not tend to be overtly present in English L2 argumentative writing.

In academic writing, Russian learners apply the ‘we’ perspective to a greater extent than other Slavic and Norwegian learners. The ‘I’ perspective is overused by all Slavic learner groups. The overuse in VESPA-RU and VESPA-PL is at about the same rate as in VESPA-NO. However, Czech learners overuse the ‘I’ perspective at a significantly higher rate than the three above-mentioned varieties. The deviations in Polish argumentative and Czech academic writing may present an interesting topic for further investigation.

A substantial part of the findings relies on the frequency variation between the learner and NS writing. Therefore, the reliability of results crucially depends on the material. The considerable size of the corpora of argumentative writing opened up a possibility for me to conduct an in-depth study of overused sequences and account for the overuse of first person pronouns. The observations for the academic part of my study are less detailed due to the modest size of VESPA-RU, VESPA-PL and VESPA-CZ. It was at times challenging to account for the overuse and underuse of authorial pronouns, as not all features stood out clearly enough. Besides, the findings on L1 transfer rely on my own collection of Russian L1 texts. Since most of them were collected from the Internet, the data about the contributors is not extensive. Therefore, the results should be treated with caution, and further in-depth study is required to make more detailed observations.

This study and these findings would hardly be possible without the existence of corpus linguistics. Therefore, I want to underline the importance of the work done by scholars within this field. I was personally able to appreciate the freedom a large corpus gives when I worked with ICLE-RU and LOCNESS. Corpus-based investigations result in important findings, which raise our awareness about the difference in the writing conventions of various cultures and help learners gain a better command of the rhetorical devices of the target language to communicate their message more successfully. The next section presents some such findings, which resulted from my corpus-based study of English L2 writing of Russian learners.

## **8.2 Pedagogical implications**

This section starts with outlining pedagogical implications, which can be helpful for any learner variety, and then gives a more detailed overview of L1-specific pedagogical implications for Russian learners.

The main implication is that textbooks should introduce learners to the concept of author identity within the framework of Anglo-American rhetoric. Even though learners tend to overuse first person pronouns, it would be wrong to advise them to be less visible in the text. A conscious use of first person pronouns is not a disadvantage, but a key component in communicating an authorial viewpoint. Learners should be made aware of the overuse and be instructed on what functions require the presence of authorial identity, and in what functions it can be downplayed. First person pronouns primarily imply the presence of the author to recount personal experience in argumentative writing and to elaborate an argument or state the results in academic writing. The low-risk text-organizing function is a typical learner use, which does not help the writer to construct author identity in the text. Students should probably be introduced to alternative and more native-like text-organizing devices. In this way, they can steer the reader through the text without self-referencing.

In many cases, the overuse of first person pronouns results from the chaining of sentence-initial features of W/R visibility. Textbooks and language instructors should not create an impression that the more introductory units learners use, the more credit they earn in an examination process. Incorrect use of these units leads to a less native-like performance and stronger deviation in the presentation of author identity. To decrease the overuse and the chaining of sentence-initial features of W/R visibility, textbooks should supply a detailed description of each introductory phrase with good examples to contrast the connotative meanings.

Even more progress can be achieved if learning material takes into account students' L1 background. This study focuses on the Russian learner variety, and the pedagogical implications are primarily aimed at this learner group. The major pedagogical implication concerns the use of the 'we' perspective in English L2 writing. In Section 6.1.1, we saw a decrease in the use of the 'we' perspective in argumentative writing, though it still retains its strong position. In academic writing, the 'we' perspective is overused to a greater extent, especially in bachelor assignments. This deviation may be caused by the instructions Russian students receive at school. As for example, Hyland (2002a) assumes that Hong Kong students are taught to hide their author identity, and Fossan (2011) suggests that Norwegian students are encouraged by teachers to be personally involved with the discussion, Russian students may be instructed that authors should not promote their own self in the text, but rather hide behind the impersonal *we*. Since writing conventions and academic norms are changing from culture to culture, it is crucial that learners are made aware of this variation in the view of author identity.

In the academic genre, the deviation is greater in bachelor student writing, as they have little academic experience. Bachelor students could benefit from a course dedicated to academic conventions in Anglo-American rhetoric. This study reports on the strong influence of Russian academic conventions on the L2 bachelor student writing. It results in the overuse of the ‘we’ perspective and the underuse of the ‘I’ perspective. Russian learners should receive instructions on the functional use of first person pronouns. Since they use *we* in high-risk functions and *I* in low-risk functions, they need to be instructed about the leading role of the ‘I’ perspective in Anglo-American rhetoric. Academic experience helps master students to avoid such strong deviation in the use of first person pronouns. The deviation at master level may decrease, if bachelor students learn about the differences of Russian and Anglo-American rhetoric.

The textbooks aimed specifically at Russian learners should include consciousness-raising activities to prevent the overuse of multi word-units typical of this learner group. Such units often result from positive L1 transfer and, therefore, high rates of overuse may not be visible even to Russian language instructors. This study points out some such cases. The unit *as for me* is strongly overused by Russian learners due to the positive L1 transfer of the unit *что касается меня* (= *as for me*). Other learner varieties do not tend to overuse the unit at such a high rate. Russian students could be presented with alternatives (*e.g. It seems to me that...* instead of *As for me, I...*) to express the same meaning in a different way. The imperative *let’s/let us* is overused by Russian and French learners (Paquot 2008). In both learner groups, the overuse results from the positive L1 transfer and requires learners’ special attention. Textbooks could include rewriting exercises, which instruct students to rewrite the sentence avoiding imperative *let’s/let us*.

Russian learners may overuse the pronouns to maintain direct word order and construct grammatically correct sentences. This may be especially challenging for this learner group since direct word order and the presence of a subject in a sentence is not obligatory in their L1 grammar. The use of pronouns is taught at the very beginning of language courses and is the simplest way to make sure a sentence is grammatically correct. It particularly concerns the writers of argumentative essays, who have an intermediate or upper-intermediate level of English. Little language experience often means the lack of alternatives or confidence to use them. To fill the gap, textbooks should discuss alternative ways of expressing author identity in the text. More attention can be paid at the use of passive voice, nominalization, omission of the agent of the action and participle clauses. Students should be taught how to exploit such structures consciously, as it is no use downplaying author identity in high-risk functions.

A quote from Granger's study on Prefabricated Patterns in EFL Writing (1998: 159) gives a summary of what I consider to be the key pedagogical implication of this study:

*'[...] L1 plays an important role in the acquisition and use of prefabs in the L2. For obvious commercial reasons, most EFL material is aimed at all learners, irrespective of their mother tongue. Given the essentially language-specific nature of prefabs, this is a major issue that must be addressed if we are serious about giving learners the most efficient learning aids.'*

This study illustrates the need for language-specific EFL material, which could take into account not only the general learner issues, but also students' L1 background.

### **8.3 Looking ahead**

This study has brought to light the major tendencies among Russian learners of English to construct their author identity in argumentative and academic writing. The research process revealed new areas of further investigation. It would be interesting to see if Russian learners overuse, underuse or misuse other involvement features such as disjuncts (*e.g. perhaps, maybe*), emphatic particles (*e.g. just, really*), reference to situation of writing/reading (*e.g. here, now, this essay*), sentence types (*e.g. questions, exclamations*) and other. My original plan included an investigation of the writing of professional authors and discipline-specific writing. Due to objective limitations, I do not study these factors in my master thesis, but it could be an interesting and fruitful field for further research. It would be desirable to investigate the L2 academic writing of Russian learners based on a larger corpus. The preferable size would be about 200,000 words or more, which would match BAWE-LING. A further functional study of first person pronouns in master and bachelor learner writing could tell us if together with academic experience learners become more aware of the role of personal pronouns.

As pointed out by Granger (1998), commercial reasons prevent publishing agencies from printing language-specific EFL material. However, if certain similarities could be identified for Slavic learner groups, the first step on the way to language-specific EFL material would be textbooks that tackle the problems of Slavic learners. New studies on Polish, Czech and Bulgarian learner writing could be a useful step in this direction. More work is required to collect samples of argumentative and academic writing of learners with Slavic L1 backgrounds, as the four corpora available today are rather small (it includes Russian, Polish, Czech and Bulgarian).

The overuse of the 'we' perspective identified in this study requires further investigations. The historical development of the perspective is especially interesting. The decrease in the use of first person plural pronouns seems to be reflected in the learners' L2 writing. This indicates a close connection between the changes in L1 and in L2. To be able to state this with a higher degree of certainty, more material needs to be collected representing learner writing from 1998 up until today. The results also imply the influence of drastic political and economic changes on the use of the 'I' and the 'we' perspective by Russian learners. It would be interesting to analyse how the change from communism to capitalism influenced the Russian language and the use of first person pronouns.

## BIBLIOGRAPHY

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

Aijmer, K. (2001). *I think* as a marker of discourse style in argumentative Swedish student writing. In K. Aijmer, ed. *A Wealth of English - Studies in Honour of Göran Kjellmer*. Gothenburg: Acta Universitatis Gothoburgensis, pp.247-58.

Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In Granger, Hung & Petch-Tyson, eds. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp.55-76.

Altenberg, B. (1997). Exploring the Swedish component of the International Corpus of Learner English. In Lewandowska-Tomaszczyk & P.J. Melia , eds. *PALC'97: Practical Applications in Language Corpora*. Łódź: Łódź University Press, pp.119-32.

Applebee, A.N., Langer, A. & Mullis, I.V. (1986). *The Writing Report Card: Writing Achievement in American Schools*. Princeton, NJ: Educational Testing Services.

Arnaudet, M. & Barrett, M. (1984). *Approaches to Academic Reading and Writing*. Englewood Cliffs, NJ: Prentice Hall.

Bakhtin, M. (1986). The problem of speech genres and the problem of the text in linguistics, philology and the human sciences: An experiment in philosophical analysis. In C. Emerson & M. Holquist, eds. *Speech Genres and Other Late Essays*. Austin: University of Texas Press, pp.250-317.

Becher, T. & Trowler, P. (2001). *Academic Tribes and Territories*. Buckingham: SRHE and the Open University Press.

Biber, D. (2006). *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.

Biber, D., Johansson, S, Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson.



Clyne, M. (1987). Cultural differences in the organization of academic texts: English and German. *Journal of Pragmatics*, 11, pp.211-47.

Čmejrková, S. (1996). Academic writing in Czech and English. In E. Ventola & A. Mauranen, eds. *Academic Writing: Intercultural and Textual Issues*. Amsterdam/Philadelphia: John Benjamins, pp.137-52.

Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, 34(2), pp.213-38.

Crawford, W. (2005). *Is L2 Writing Like Native-English Conversation?* Paper presented at ICAME 26 (International Computer Archive of Modern and Medieval English) – AAACL 6 (American Association of Applied Corpus Linguistics). Michigan: University of Michigan.

Crowhurst, M. (1983). *Revision Strategies of Students at Three Grade Levels*. ERIC Document Reproduction Service No. ED 238 009).

Day, R. & Gastel, B. (2006). *How to Write and Publish a Scientific Paper*. Cambridge : Cambridge University Press.

Erman, B. & Warren, B. (2000). The idiom principle and the open choice principle. *Text - The Hague then Amsterdam then Berlin*, 20(1), pp. 29-62.

Fløttum, K., Gedde-Dahl, T. & Kin, T. (2006a). "We now report on." versus "Let us now see how.": Author Roles and Interaction with Readers in Research Articles. In K. Hyland & M. Bondi, eds. *Academic Discourse across Disciplines*. Bern: Peter Lang, pp.203-24.

Fløttum, K., Gedde-Dahl, T. & Kin, T., eds. (2006b). *Academic Voices: Across Languages and Disciplines*. Amsterdam/philadelphia: John Benjamins.

Fossan, H. (2011). *The Writer and the Reader in Norwegian Advanced Learners' Written English*. Oslo: University of Oslo.

Gardner, S. & Holmes, J. (2009). Can I use headings in my essay? Section headings, macrostructures and genre families in the BAWE corpus of student writing.. In: M. Charles, D. Pecorari & S. Hunston, eds. *Academic Writing: At the Interface of Corpus and Discourse*. London/New York: Continuum, pp. 251-271.

Gilquin, G. & Paquot, M. (2007). Spoken features in learner academic writing: Identification, explanation and solution. In Davies, M., Rayson, , Hunston, & Danielsson, , eds. *Proceedings*

of the Fourth Corpus Linguistics Conference CL2007, University of Birmingham, 27-30 July 2007. Birmingham : University of Birmingham.

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

Gong, G. & Dragga, S. (1995). *A Writer's Repertoire*. New York: Longman.

Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg & M. Johansson, eds. *Languages in Contrast. Text-Based Cross-Linguistic Studies*. Lund: Lund University Press, pp.37-51.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. In P. Cowie, ed. *Phraseology: Theory, Analysis and Applications*. Oxford: Oxford University Press, pp.145-60.

Granger, S. & Paquot, M. (2009). In search of a General Academic Vocabulary: A corpus-driven study. Paper presented at the international conference 'Options and Practices of L.S.A.P Practitioners'. Heraklion, Crete: University of Crete. Available at: [http://www.uclouvain.be/cps/ucl/doc/adri/documents/In\\_search\\_of\\_a\\_general\\_academic\\_english.pdf](http://www.uclouvain.be/cps/ucl/doc/adri/documents/In_search_of_a_general_academic_english.pdf) [Accessed 20 August 2013].

Granger, S. & Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger, ed. *Learner English on Computer*. London/New York: Addison Wesley Longman, pp.119-31. Available at: [http://www.comp.lancs.ac.uk/~paul/publications/gr98\\_leoc.pdf](http://www.comp.lancs.ac.uk/~paul/publications/gr98_leoc.pdf) [Accessed 25 August 2013].

Hamp-Lyons, L. & Heasley, B. (2006). *Study Writing: A Course in Written English for Academic Purposes*. Cambridge: Cambridge University Press.

Harwood, N. (2005). 'We do not seem to have a theory ... The theory I present here attempts to fill this gap': Inclusive and exclusive pronouns in academic writing. *Applied Linguistics*, 26(3), pp.343-75.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4(2), pp.237-60.

Hasselgård, H. (2009). Thematic choice and expressions of stance in English argumentative texts by Norwegian learners. In K. Aijmer, ed. *Corpora and Language Teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp.121-39.

- Herriman, J. (2009). The Use of *we* in argumentative essays by Swedish advanced learners of English. In R. Bowen, M. Mobärg & S. Ohlander, eds. *Corpora and Discourse - and Stuff Papers in Honour of Karin Aijmer*. Gothenburg: University of Gothenburg, pp.105-20.
- Hinds, J. (1987). Reader versus writer responsibility: A new typology. In U. Connor & R.B. Kaplan, eds. *Writing Across Languages: Analysis of L2 Text*. Reading, MA: Addison Wesley, pp.141-52.
- Howarth, P.A. (1996). *Phraseology in English Academic Writing*. Tübingen: Max Niemeyer Verlag.
- Hugh, C. (1999). *Advanced Writing with English in Use. CAE*. Oxford University Press: Oxford UP.
- Hyland, K. (2002a). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics*, 34, pp.1091-112.
- Hyland, K. (2002b). Options of identity in academic writing. *ELT Journal*, Volume 56//4.
- Hyland, K. (2002c). What do they mean? Questions in academic writing. *Text*, 22(4), pp.529–57. Available at: [http://www2.caes.hku.hk/kenhyland/files/2012/08/What-do-they-mean\\_questions-in-academic-writing.pdf](http://www2.caes.hku.hk/kenhyland/files/2012/08/What-do-they-mean_questions-in-academic-writing.pdf) [Accessed 25 August 2013].
- Hyland, K. & Tse, P. (2007). Is there an "Academic vocabulary"? *TESOL Quarterly*, 41(2), pp.235-53.
- Ivanic, R. & Camps, D. (2001). I am how I sound. Voice as self representation in L2 writing. *Journal of Second Language Writing*, 10(1), pp.3-33.
- Ivanič, R. (1998). *Writing and Identity: The Discoursal Construction of Identity in Academic Writing*. Amsterdam/Philadelphia: John Benjamins Publishing.
- John, S. (2012). Identity without the 'I': A study of citation sequences and writer identity in literature review sections of dissertations. *Academic Writing in a Second or Foreign Language. Issues and Challenges Facing ESL/EFL Academic Writers in Higher Education Context*, pp.186-203.
- Jordan, R.R. (1999). *Academic Writing Course: Study Skills in English*. Edinburgh: Longman.
- Kertz, E. & Haas, F. (2009). The aim is to analyse NP: the function of prefabricated chunks in academic texts. In R. Corrigan, E. Moravcsik, H. Ouali & K. Wheatley, eds. *Formulaic*

*Language: Volume 1. Distribution and Historical Change*. Amsterdam, Philadelphia: John Benjamins, pp.97-117.

Kuo, C. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes*, 18, pp.121-38.

Mauranen, A. (1996). Discourse competence - evidence from thematic development in native and non-native Texts. In E. Ventola & A. Mauranen, eds. *Academic Writing. Intercultural and Textual Issues*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp.195-230.

McCann, T.M. (1989). Student argumentative writing knowledge and ability at three grade levels. *Research in the Teaching of English*, 23(1), pp.62-76. Available at: <http://www.jstor.org/stable/40171288> [Accessed 10 October 2013].

McEnery, T. & Hardie, A. (2012). *Corpus Linguistics. Method, Theory, Practice..* Cambridge: Cambridge University Press.

Narita, M. & Sugiura, M. (2006). The use of adverbial connectors in argumentative essays by Japanese EFL college students. *English Corpus Studies*, 13, pp.23-42.

Neff, J., Ballesteros, F., Dafouz, E., Martínez, F. & Rica, J.-P. (2007). A contrastive functional analysis of errors in Spanish EFL university writers argumentative texts: corpus-based study. In Fitzpatrick, ed. *Corpus Linguistics Beyond the Word: Corpus Research from Phrase to Discourse*. Amsterdam: Rodopi, pp.203-25.

Ohta, A.S. (1991). Evidentiality and politeness in Japanese. *Issues in Applied Linguistics*, 2(2), pp.183-210.

Paquot, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In F. Meunier & S. Granger, eds. *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins Publishing, pp.101-19.

Paquot, M. (2010). *Academic Vocabulary in Learner Writing. From Extractions to Analysis*. London: Continuum.

Paquot, M., Hasselgård, H. & Ebeling, S.O. (2013). Writer/reader visibility in learner writing across genres: A comparison of the French and Norwegian components of the ICLE and VESPA Learner Corpora. *Twenty Years of Learner Corpus Research: Looking back, Moving, Corpora and Language in Use – Proceedings 1*, pp.377-87.

Pawley, A. & Syder, F. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In R.W. Richards & S.C. Schmidt, eds. *Language and Communication*. London: Longman, pp.191-226.

Petch-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In S. Granger, ed. *Learner English on Computer*. London/New York: Longman, pp.107-18.

Recski, L.J. (2004). Expressing standpoints in EFL written discourse. *Revista Virtual de Estudos da Linguagem*, 3(2). Available at:

<http://132.248.9.34/hevila/Revistavirtualdeestudodalinguagem/2004/vol2/no3/3.pdf> [Accessed 20 October 2013].

Ringbom, H. (1998). Vocabulary frequencies in advanced learner English: A cross-linguistic approach. In S. Granger, ed. *Learner English on Computer*. London/New York: Addison Wesley Longman, pp.41-52.

Rundell, M. & Granger, S. (2007). From corpora to confidence. *English Teaching Professional*, (50), pp.15-18.

Scollon, R. (1994). As a matter of fact: The changing ideology of authorship and responsibility in discourse. *World Englishes*, 13, pp.34-46.

Seidlhofer, B. (2005). Key concepts in ELT. English as a Lingua Franca. *ELT Journal*, 59/4, pp.339-41. Available at: [http://people.ufpr.br/~clarissa/pdfs/ELF\\_Seidlhofer2005.pdf](http://people.ufpr.br/~clarissa/pdfs/ELF_Seidlhofer2005.pdf) [Accessed 6 September 2013].

Simon-Vandenberg, A.M. (2000). The function of *I think* in political discourse. *International Journal of Applied Linguistics*, 10(1), pp.41-63.

Spencer, C. & Arbon, B. (1996). *Foundations of Writing: Developing Research and Academic Writing Skills*. Lincolnwood, IL: National Textbook Co.

Stein, N.L. & Glenn, C.G. (1979). An analysis of story comprehension in elementary school children. In R.O. Freedle, ed. *New Directions in Discourse Processing*. Norwood, NJ: Ablex, Inc.

Tang, R. & John, S. (1999). The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes*, 18, pp.S23-39.

*The DEV Guide to Academic Writing*. (2009). University of East Anglia.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam/Philadelphia: John Benjamins Publishing.

Vassileva, I. (1998). Who am I/Who are we in academic writing. *International Journal of Applied Linguistics*, 8(2), pp.163-90.

Virtanen, T. (1998). Direct questions in argumentative student writing. In S. Granger, ed. *Learner English on Computer*. London: Longman, pp.94-106.

Zaytseva, E. (2011). Register, genre, rhetorical functions: Variation in English native-speaker and learner writing. *Multilingual Resources and Multilingual Applications*, pp.239-42.

Ädel, A. (2008). Involvement features in writing: do time and interaction trump register awareness? In G. Gilquin, S. Papp & M.B. Diez-Bedmar, eds. *Linking up Contrastive and Learner Corpus Research*. Amsterdam/New York: Rodopi, pp.35-53.

## **CORPORA USED**

### **ICLE: ICLE-Bulgarian, ICLE-Czech, ICLE-Polish, ICLE-Russian**

The texts were studied using Wordsmith Tool and ICLE 2.

### **LOCNESS**

The texts were studied using Wordsmith Tool.

### **VESPA: VESPA- Czech, VESPA- Polish, VESPA- Russian**

The texts were studied using Wordsmith Tools.

### **BAWE (BAWE-LING)**

The texts were studied using Wordsmith Tool.

### **BNC: written and spoken parts**

The texts were studied through the online interface at the official website of *BNCWeb*  
<http://www.tekstlab.uio.no/bnc/BNCquery.pl?theQuery=search&urlTest=yes>

## APPENDIX 1. The frequencies of the 2- to 4-gram clusters with the pronoun *I* in LOCNESS and ICLE-RU

The log-likelihood test confirms the statistical significance of the overuse in Table 47.<sup>61</sup>

**Table 47.** 2- to 4-gram clusters with the pronoun *I* with a minimum frequency of 10 instances in LOCNESS. Frequency per 10,000 words

Cluster with the pronoun <i>I</i>	LOCNESS	
	Raw frequencies	Per 10,000 words
<i>I think</i>	87	2.70
<i>I feel</i>	86	2.67
<i>I believe</i>	54	1.67
<i>I know</i>	27	0.84
<i>I would like to</i>	12	0.37
<i>I agree</i>	12	0.37
<i>I am sure</i>	11	0.34

**Table 48.** 2- to 4-gram clusters with the pronoun *I* with a minimum frequency of 10 instances in ICLE-RU. Frequency per 10,000 words

Cluster with the pronoun <i>I</i>	ICLE-RU		ICLE-RU overuse
	Raw frequencies	Per 10,000 words	
<i>I think</i>	271	12.08	4.47
<i>I would like to/I'd like to</i>	71/34	3.17/1.52	8.57/-
<i>I believe</i>	52	2.32	1.39
<i>I mean</i>	42	1.87	-
<i>I suppose</i>	34	1.52	-
<i>I want</i>	34	1.52	-
<i>I am sure/I'm sure</i>	22/11	0.98/0.49	2.88/-
<i>I know</i>	20	0.89	1.06
<i>I do not think/I don't think</i>	18/18	0.80/0.80	-
<i>I wonder</i>	16	0.71	-
<i>I don't want</i>	15	0.67	-
<i>I hope</i>	15	0.67	-
<i>As for me, I</i>	11	0.49	-

<sup>61</sup> Statistical significance for each instance of overuse – 99.99%.



**APPENDIX 2. The frequencies of the 2- to 4-gram clusters with the pronoun *I* in BAWE-LING and VESPA-RU**

**Table 49.** The top ten 2- to 4-gram clusters with the pronoun *I* in BAWE-LING. Frequency per 10,000 words.

Feature	BAWE-LING	
	Raw frequencies	Per 10,000 words
<i>I will</i>	73	3.70
<i>I have</i>	67	3.40
<i>I think</i>	24	1.22
<i>I feel</i>	23	1.17
<i>that I</i>	23	1.17
<i>I would</i>	20	1.01
<i>I believe</i>	19	0.96
<i>I decided</i>	19	0.96
<i>I am</i>	18	0.91
<i>I was</i>	18	0.91

**Table 50.** The top ten 2- to 4-gram clusters with the pronoun *I* in VESPA-RU. Frequency per 10,000 words.

Feature	VESPA-RU		VESPA-RU overuse
	Raw frequencies	Per 10,000 words	
<i>I have</i>	16	5.58	1.64
<i>I will</i>	14	4.89	1.32
<i>I was</i>	9	3.14	3.44
<i>research I</i>	7	2.44	-
<i>I know</i>	5	1.75	-
<i>that I</i>	5	1.75	1.50
<i>I decided</i>	4	1.40	1.45
<i>I got</i>	4	1.40	-
<i>I know you</i>	4	1.40	-
<i>I know you didn't</i>	4	1.40	-

### APPENDIX 3. The collocations of the pronoun *me* in BAWE-LING and VESPA-RU (collocation window span: 0 right – 1 left)

**Table 51.** The collocations of the pronoun *me* in BAWE-LING with a frequency over 0.15 instances per 10,000 words (collocation window span: 0 right – 1 left).

Feature	BAWE-LING	
	Raw frequencies	Per 10,000 words
<i>results</i>	7	0.35
<i>findings</i>	6	0.30
<i>experience</i>	6	0.30
<i>data</i>	4	0.20
<i>view</i>	3	0.15
<i>teaching</i>	3	0.15
<i>learning experience</i>	3	0.15
<i>analysis</i>	3	0.15

**Table 52.** The collocations of the pronoun *me* in VESPA-RU with a frequency over 0.15 instances per 10,000 words (collocation window span: 0 right – 1 left).

Feature	VESPA-RU	
	Raw frequencies	Per 10,000 words
<i>research</i>	2	0.70
<i>hypothesis</i>	2	0.70
<i>conclusions</i>	2	0.70
<i>testimony</i>	1	0.35
<i>term</i>	1	0.35
<i>requirements</i>	1	0.35
<i>purpose</i>	1	0.35
<i>parameters</i>	1	0.35
<i>opinion</i>	1	0.35
<i>observations</i>	1	0.35
<i>mind</i>	1	0.35
<i>material</i>	1	0.35
<i>intention</i>	1	0.35
<i>expectations</i>	1	0.35
<i>data</i>	1	0.35
<i>conclusions</i>	1	0.35
<i>approach</i>	1	0.35
<i>analysis</i>	1	0.35