University of Latvia
Faculty of Humanities

STUDENT-COMPOSED ELECTRONIC DISCOURSE AS A RESULT OF APPLIED LINGUISTIC RESEARCH

Studentu elektroniskais diskurss kā lietišķās valodniecības pētījuma rezultāts

Promocijas darbs – disertācija filoloģijas doktora grāda iegūšanai valodniecības zinātņu nozares lietišķās valodniecības apakšnozarē

Zigrīda Vinčela

Scientific adviser:
Dr. habil. philol. I.Kramiņa

2010

DECLARATION OF ACADEMIC INTEGRITY

I, Zigrīda Vinčela, hereby declare that this study is my own and does not contain any unacknowledged material from any source.

December 2, 2010 ……..
(signature)

ACKNOWLEDGEMENTS

I would first like to express my deepest gratitude to Dr. habil. philol. Ingrīda Kramiņa, who stimulated and encouraged me throughout this research. Her thoughtful scientific advice, comments, and help were invaluable.

My grateful thanks are also due to colleagues of the Faculty of Humanities, the Institute of Mathematics and Computer Science, the University of Latvia who shared their expertise and insights in a number of ways.

Finally, I want to express my debt to my family for being around to offer care, support and encouragement in difficult times throughout this research.

CONTENTS

ABBREVIATIONS

ASCII	American Standard Code for Information Interchange
AWL	Academic Words of English
BNC	British National Corpus
BSP	Bachelor Study Programme
CARS	Create a Research Space (model)
CLAWS	Constituent Likelihood Automatic Word-tagging System
CMC	Computer-mediated Communication
DAP	Data Access and Protection
EAP	English for Academic Purposes
EF	Eutropian Federation
ELT	English Language Teaching
ESP	English for Specific Purposes
EU	European Union
ICT	Information Communication Technology
IDEELS	Intercultural Dynamics in European Education through Online Simulation
IRC	Internet Relay Chat
MD	Multi-dimensional
MDA	Multi-dimensional Analysis
MOOs	Multi-user Object-oriented Dimension
MUDs	Multi-user Dungeons or Dimensions
NNSE	Non-Native Students' of English
OPUSi	Messaging System Used during the Participation in the Simulations of Project *IDEELS*
POS	Part-of-speech
S/FL	Second/Foreign Language
SFL	Systemic Functional Linguistics
S/FLW	Second/Foreign Language Writing
SGML	Standard Generalized Markup Language
SPSS	Statistical Package for the Social Science
TEI	Text Encoding Initiative
UCREL	University Centre for Computer Corpus Research

TABLES

FIGURES

INTRODUCTION

A characteristic feature of the end of the 20th and the beginning of the 21st century is the worldwide advent of information communication technology (*ICT*). In the 1990s and onwards, computer networks became popular in a range of countries including Latvia. The current technology has provided written communication with facilities that were hardly imaginable before the era of computer networks. Versatile electronic communication systems (e-mail, discussion forums, and synchronous conferences) that serve the needs of written communication enable a person to switch over from one electronic document to another flexibly and quickly. Moreover, the online embedding of documents in different formats as well as virtual drafting, editing, and submitting of them diversifies the choice of lexical and grammatical elements in contemporary electronic written communication.

Latvia is a European Union member state and the English language proficiency enables participation in international projects as well as successful competition in the labour market. The English language is widely used in international communication in various domains and at various levels, where electronic written communication is particularly essential. However, the English language use in the contemporary electronic communication can cause difficulties for non-native speakers of the English language since the appropriate choice of lexical and grammatical elements in this communication is a complicated process. It requires the correlation of the mentioned linguistic elements with the goal of each electronic communicative situation. Therefore the use of lexical and grammatical elements in electronic texts and the investigation of their communicative functions in these texts are topical for English language students, who are non-native speakers of English (*NNSE*). The electronic discourse of students within the framework of the present study comprises the electronic texts written by the students during the writing activities of e-course *English Academic Writing III* and online simulation of the project *IDEELS* (*Intercultural Dynamics in European Education through Online Simulation*).

The research of electronic texts is based on theories on linguistic characteristics of discourse investigated by Halliday (1985), Brown and Jule (1983), Cook (1989), Biber (1988; 1995) and other researchers. Biber, who has applied multi-dimensional analysis (MDA) of discourse, has revealed the patterns of interaction of lexical and grammatical elements in various interactional and transactional texts.

The theories of the linguists mentioned have served as a basis for researchers investigating language use in electronic written texts. Warschauer (1996), Herring (2001), Crystal (2001), Whittaker (2003), Kern (2006) have explored texts of electronic discourse and concluded that this discourse differs from the conventional discourse, as electronic interactional and transactional texts draw on linguistic units used in written as well as spoken language. Moreover, Carter et al (2001) and Kern (2006) emphasise that a particular messaging system can be used for writing, editing and forwarding of interactional as well as transactional texts. The earlier research of electronic texts of students' and their writing process by Biesenbach-Lucas, Weasenforth (2001), Pennington (2004), Granger (2004), Cigankova (2004), Scott, Tribble (2006), Farneste (2006), Vinčela (2008) reveal that students experience difficulties in choosing lexical and grammatical elements in their electronic texts. However, to date no detailed *MDA* analysis of lexical and grammatical elements in students' electronic texts written in the English language has been undertaken in Latvia. Therefore the purpose of this research is to undertake a detailed investigation of lexical and grammatical elements in the students' electronic texts that are written in English. The present study aims at identifying statistically computable differences in the frequency of occurrence of linguistic elements in students' electronic discourse. In order to reveal these differences, the Latvian statistical term *variācija* (Raščevska and Kristapsone 2000) was used in the thesis as the equivalent to the English term *variation* (Biber, 1988).

The research subject

Linguistic variation in electronic texts of students who are non-native speakers of English (*NNSE*).

The research object

Electronic texts compiled in a corpus. These texts were collected during students' participation in online simulations of the project *IDEELS* (*Intercultural Dynamics in European Education through Online Simulation*) and writing activities envisaged for an e-course *English Academic Writing III*.

The research goal

The goal of the present study is to investigate and compare theories on the use of lexical and grammatical features in English written and, in particular, in electronically written

texts in order to design writing process model of electronic texts in English as well as to design a structurally and part-of-speech marked corpus of students' electronic texts and explore variation of lexical and grammatical features in the texts of this corpus by undertaking the multi-dimensional analysis (*MDA*).

The research hypothesis

The variation of lexical and grammatical features and electronic medium-specific means of expression in students electronic texts are conditioned by discourse type the students, who are non-native speakers of English, choose in order to achieve communicative purpose of these electronic texts in the academic environment, thus developing their competences.

The research objectives

1 To investigate and analyse linguistic theories to provide a theoretical basis for the concepts of text and discourse, the use of linguistic elements in spoken and written communication and the linguistic knowledge required for the development of electronic discourse in a foreign language.

2 To develop a model for the writing process of electronic texts and design e-courses that would foster the study and application of linguistic theories within the framework of Bachelor study programme of English Philology.

3 To compile a corpus of students' electronic texts that comprises part-of-speech annotated interactional and transactional texts.

4 To investigate electronic texts of the students' corpus in order to identify theoretically grounded variation of linguistic units in their interactional and transactional texts.

5 To undertake quantitative and qualitative investigation involving descriptive statistics.

Theoretical research methods

The theoretical background of the present research is the comparative analysis of theories of general linguistics and applied linguistics of the 20th century and of the beginning of the 21st century. The theoretical background for the present research has been based on linguistic theories of Harris (1952), Brown and Jyle (1983), Ceplītis et.al. (1989), Chaffe (1992), Cook, (1989) Widdowson (2004), Bušs et. al. (2007) on the concepts of text and discourse, on Jakobson's (1960), Halliday's (1964), Lyons' (1977), Biber's (1988), Trosborg's (1997), Buhler's (1999), Valdmanis' (1999, 2002), Hyland's (2002), Martin's (2003), Bhatia's (2004), Eggins' (2004) theories on genre, register and linguistic

characteristics of spoken and written discourse, on the theories of Warschauer (1996), Crystal (2001), Herring (2001), Whittaker (2003), Kern (2006) on the linguistic characteristics of electronic discourse as well as on the theories of written discourse development models proposed by White and Arndt (1991), Tribble (1996, 2002), Bjork (2003), Coffin (2003), Pennington (2003).

Empirical research methods

1 The quantitative research method, namely, the multi-dimensional analysis (*MDA*) proposed by Biber (1988) has been applied in the present study to investigate the frequency of lexical and grammatical units in 1412 electronic texts (220 012 tokens) written by 102 *NNSE*. The target audience of the investigation is English Philology students of the University of Latvia, Faculty of Humanities. The structurally marked and part-of-speech annotated corpus of student-composed electronic texts was created with the help of *TEI* (Text Encoding Initiative) *CLAWS* (*Constituent Likelihood Automatic Word – tagging System*) developed at *UCREL* (*University Centre for Computer Corpus Research*)**.** The lexico-grammatical units were extracted with the help of *WordSmith Tools*.

2 In order to explore the frequency variation of lexico-grammatical units in the corpus of electronic texts, the following statistical analysis has been performed:

2.1 The identification of the dispersion of lexico-grammatical units with the help of *Microsoft Office Excel* and *SPSS-15.0*

2.2 The calculation of factors/dimension scores in students' electronic texts with the help of *Microsoft Office Excel* and *SPSS-15.0*

2.3 The identification of the significance of differences among electronically written texts with the help of Pearson's Chi-Square.

3 The qualitative analysis of the corpus in order to explore characteristic features of electronic texts going beyond the quantitative analysis.

Target audience

The target audience of the present Thesis is students, who are *NNSE* and study the English philology at the Faculty of Humanities, the University of Latvia. Further in the Doctoral Thesis the target audience is labeled students.

The research novelty

1 The methodology of the multi-dimensional analysis of discourse proposed by Biber (1988) has been first applied in the investigation of electronic texts composed by the students who are *NNSE*.

2 The author has presented the theoretical basis for the technology-enhanced writing process model that has been integrated in the e-course *English Academic Writing III.*

3 The author has designed a structurally marked and part-of-speech annotated corpus of complete electronic texts composed by the students (who are *NNSE*) and such corpus of texts has been for the first time compiled and investigated in Latvia.

4 *WordSmith Tools* have been applied for the first time in the extraction of lexical and grammatical features from the corpus of complete electronic texts composed by the afore-mentioned students.

5 The quantitative data obtained during the statistical analysis of the innovative corpus can reveal the knowledge that students have acquired about the variation of lexical and grammatical features in various electronic texts that they have composed in the English language.

Theoretical significance

1 The research presents a theoretically substantiated analysis of linguistic variation in students' electronic texts, thus providing the theoretical framework for the model of linguistic variation in electronic texts written in English by students who are non-native speakers of English. The said model can be applied as the methodological framework in research in applied linguistics.

2 The research contributes to the development of the theory of linguistic variation and provides the theoretical basis for e-course design and application in foreign language studies.

Practical significance

1 A comprehensive linguistic analysis of the corpus containing students' electronic texts (1412 texts, 220 012 tokens) written in English has been undertaken. The designed corpus can be applied in the further research.

2 The performed research provides the theoretical basis for e-course design applicable in foreign language studies.

The approbation of research

1 The results of the research have been discussed and approved at the meetings of the Department of English Studies, the Faculty of Humanities, the University of Latvia.

2 The theoretical framework developed in the present Thesis and the results of the research have been applied in the designing of language teaching materials, namely, four e-courses within the framework of the E-university Project at the University of Latvia as well as in the text-book *Written and Spoken Communication in English* and during the participation in two projects: *Socrates* project *Nordic-Baltic Network for Foreign Language Teaching* (2002-2003) and *EU* Project *Science and Technology Students' Foreign Language Skills Development* (2007-2008).

3 The investigation results have been presented in 16 papers and conference proceedings as well as reported at 17 conferences in Latvia and abroad (in Canada, France, Greece, Lithuania, Sweden and Turkey) from 2000 till 2010.

Outline of the Thesis

The thesis consists of an introduction, two chapters, conclusions, bibliography and annexes.

Part 1, *A Theoretical Basis to the Investigation of Student-composed Electronic Discourse,* includes three subchapters. The first chapter *Text and Discourse* outlines the major theoretical principles in the application of two core terms in the present study, namely, *text* and *discourse*, paying special attention to the theories on their interrelation that explain their interchangeable use in the study. The second chapter, *Linguistic Variation in Discourse*, defines the term *register* and is devoted to theories on register-based variation of linguistic features in written discourse. The given chapter provides the theoretical background on factors determining linguistic variation across texts. It focuses on language use in written discourse that leads to the theories on the multi-dimensional analysis (*MDA*) of this discourse. The final part of the given chapter is devoted to the theoretical considerations on the application of theories on linguistic variation in electronic discourse. The third chapter, *Models of Written Discourse Development and Role of Information Technology in their Application,* discusses the three main approaches to written discourse development. The subchapter emphasizes that these approaches have merged within the contemporary technology-enhanced reader-oriented approaches of discourse development. These theoretical considerations serve as the basis for a three-stage technology-enhanced model for *FLW* instruction proposed by the author of the study that

aims at students' involvement in register pertinent linguistic variation in their electronic texts.

Part 2, *Research on Linguistic Variation in Student-composed Electronic Discourse,* contains two subchapters. The first chapter, *Research Methodology and Procedure*, provides details of the research methodology applied in the present study, a description of the characteristic elements of electronic environments as well as the contextual framework of the technology-enhanced individual and collaborative writing activities of students who are *NNSE*. An explanation is provided for the procedure for implementation of these activities, the selection of the linguistic units for the investigation of the register transfer in students' electronic discourse, followed by details about the arrangement of students' who are *NNSE* corpus of electronic texts, the annotation of the corpus and its processing to identify the application of selected linguistic units by students. The chapter provides details on the descriptive statistics applied in the analysis of students' electronic texts. The second chapter, *Analysis of Student-composed Electronic Texts as Discourse Manifestations*, provides the analysis of the detected frequency counts of linguistic features. The first part of the chapter is devoted to the analysis of particular linguistic units that have displayed the most significant register-pertaining variation and their comparison with Biber's (1988) study. The second part of this chapter presents an analysis of the overall language use variation across the individually and collaboratively developed texts.

Conclusions of the study emphasize the main findings of the whole investigation.

1 A THEORETICAL BASIS TO THE INVESTIGATION OF STUDENT-COMPOSED ELECTRNIC DISCOURSE

The first two chapters of this study are devoted to an analysis of the theoretical considerations on the notions: text – discourse – communicative context – register, as well as upon linguistic variation within written and spoken communication. The third chapter of the first part of this study explores the role of technology in the various 'writing process' models that support the author's own paradigm of electronic text development and the investigation of linguistic variation within student-composed electronic discourse.

1.1 Text and Discourse

Theories on the definitions of *text* and *discourse* reveal that they range from almost interchangeable use to drawing a clear line between these notions. Such variety of approaches, as McCartney (1991: 7) maintains, is caused by the fact that 'discourse analysis has grown into a wide-ranging and heterogeneous discipline which finds its unity in the description of language above the sentence and an interest in the contexts and cultural influences which affect language in use'. Consequently, the definitions of *text* and *discourse* vary depending on the theoretical frames of reference adopted by linguists. Therefore, whenever both terms are used in a particular study, it is essential to define them in the context of that particular study. The purpose of the present overview on definitions is to consider them from the perspective of written discourse analysis in order to highlight linguistic theories underlining these terms and their application in the present study.

The emergence of the term *discourse* was, according to Cook (1989: 12), preceded by numerous investigations of language in contexts 'under various guises'. For example, grammar was separated from rhetoric by scholars of Greece and Rome 'the former being concerned with the rules of language as an isolated object, the latter with how to do things with words to achieve effects and communicate successfully with people in particular contexts' (ibid). Other, approaches that have focused on language in its full context as part of society and the world, were proposed in the first half of the twentieth century in North America. Scholars who were both anthropologists and linguists investigated the languages and societies of the Native Americans (American Indians). A similar tradition has been developed in Britain by Firth (1935), who considers language to be a part of culture, which is responsive to the environment.

In order to substantiate the application of the terms *text* and *discourse* in the present study, a more detailed analysis has been undertaken on the recent linguistic theories devoted to these notions. The overview reveals two overlapping approaches that underlie the move from the quantitative (according to the size of the linguistic unit) to the conceptual (according to the context that a linguistic unit manifests) definition of the terms *text* and *discourse*:

The definition of the terms *text* and *discourse* became topical with the emergence of the term *discourse* in the middle of the twentieth century.

This term was introduced by Zellig Harris, the distinguished Prague School linguist, and it entered the general use as the title of his paper *Discourse Analysis* published in 1952 and was elaborated by linguists (Brown and Jule, 1983, Cook, 1992, Gee, 1996, van Dijk, 1997, Chafe, 2003, Bhatia, 2004, Wioddowson, 2004).

Harris notes in the paper that there are two approaches to discourse analysis, 'continuing descriptive linguistics beyond the limits of a single sentence' and 'correlating culture and language' or non-linguistic and linguistic behaviour (Harris, 1952: 3). Harris in his paper *Discourse Analysis* focuses on the first approach. Therefore, his claim that 'Language does not occur in stray of words or sentences, but in connected discourse' (ibid) is based on the examples showing that the connectedness is established via recurrent morpheme sequences due to their structural similarity. It shows that Harris has investigated how language is organized above a sentence by dealing with the structural frameworks of 'larger linguistic units than a sentence' that underlie his definition of discourse. It shows that Harris distinguishes the terms *text* and *discourse* according to the magnitude of linguistic units to be investigated. Discourse according to Harris is larger than a sentence.

Stubbs points out that both terms *text* and *discourse* refer to 'language above the sentence, or above the clause' and admits that the terms *text* and *discourse* are 'often ambiguous and confusing'; however, refrains from disambiguation of them. He notes that the distinction can be made according to the communication mode 'one often talks of "written text" versus 'spoken discourse [...] and *discourse* implies length whereas a *text* may be very short' (Stubbs 1983: 9). The inconsistency in the use of the terms *text* and *discourse* remains topical also in his further research:

> *There is considerable variation in how terms such as text and discourse are used in linguistics. Sometimes this terminological variation signals important conceptual distinctions, but often it does not, and terminological debates are usually of little interest.*
> (Stubbs, 1996: 4)

The distinction of the notions *text* and *discourse* according to the magnitude of linguistic units is recognized by Bhatia (2004: 4) as the textualization of lexico-grammar, which reflects the tendency of narrowing the term *discourse* to *text.* Bhatia points out that these theories refer to the 1960s and the early 1970s when linguists were strongly influenced by frameworks in formal linguistics, namely, the analysis of the surface-level statistical significance of one or several lexico-grammatical features in a 'particular subset of texts associated with a particular discipline' (ibid.). For example, Barber (1962) has identified significant grammatical features in a corpus of scientific texts. Spencer (1975) has investigated noun-verb combinations in legal discourse. Bhatia and Swales (1983) have identified nominalizations in legislative discourse. These studies demonstrate a relative lack of attention to the functional variation of lexical and grammatical features in various discourse forms; however they have initiated the functional investigation of these features, although mostly within clause boundaries without reference to discourse organization.

The indiscriminate use of the terms *text* and *discourse* and the definition of *discourse* as language beyond sentence boundaries persist in more recent publications, for example, Chafe (1992: 356, 2003: 439 – 40) in the definition that is included in two impressions of *Oxford International Encyclopaedia of Linguistics* maintains:

> *The term discourse is used in somewhat different ways by different scholars, but underlying the differences is a common concern for language beyond the boundaries of isolated sentences. The term text is used in similar ways. Both terms may refer to a unit of language larger than the sentence: one may speak of a discourse or a text.*
>
> Chafe (1992: 356, 2003: 439 – 40)

Hoey (1991: 197) and Virtanen (1990: 448) have concluded that there is a considerable indifference towards conceptual distinction of the terms *text* and *discourse*. However, Hoey supposes that 'distinction continues to be made' as '[...] it is as if some basic differentiation is felt to exist that people cannot quite agree on but cannot leave alone'.

The conceptual definition of the terms *text* and *discourse* is addressed by linguists who focus on the limitations of the definition '*discourse* is language above sentence' and bring out the functional approach to *discourse* as language event in a context. The conceptualization of this term, therefore, originates from the linguists' stance towards dichotomy between language use and the abstraction of language use. Thus Brown and Jule (1983: 1) maintain that the discourse analysis involves the analysis of the language in use and that 'it [discourse analysis] cannot be restricted to the description of linguistic forms independent of purposes or functions which these forms are designed to serve in human affairs'.

The focus on functional properties of language underlies influential approaches that undertake the conceptual differentiation of the terms *text* and *discourse* irrespective of the communication mode: written (text) and spoken (discourse). Brown and Yule (1983:1) maintain that *discourse* equals language in use, while *text* is defined as the 'verbal record of a communicative act'. Widdowson (1979: 90) in his early works also distinguishes between *text* as 'sentences in combination' and *discourse* as the 'use of sentences in combination'. A similar view is supported by Edmonson (1981: 4) who defines *text* as 'a structural sequence of linguistic expressions forming a unitary whole', and *discourse* as 'a structured event manifest in linguistic (and other) behaviour'.

A similar position on the definition of text is adopted by Latvian linguists (Bušs et.al. 2007, Ceplītis et. al. 1989, Rozenbergs, 2004) who claim that a text is the spoken or written actualisation of linguistic communication. Its content, form and scope are determined by the specifics of the information, the setting of the communication and the subjective relationship of the message sender towards the content that is expressed and towards the message recipient.

Consequently, the perception of *discourse* as a process (language in use) advances the linguists' definition that widens the scope of the concept *discourse* through its contextualization and draws attention to the peculiarities of the interrelationship between *discourse* and the linguistic code. Linguists bring out an important observation: the concept *discourse* is wider than *text* in given qualitative terms. Thus *text* can equal discourse without context 'I shall use the term *text* to refer to any written record of communicative event (Nuan, 1993: 6), while *discourse* includes *text* and its situational context 'I shall reserve the term discourse to refer to the interpretation of the communicative event in context'(ibid.).

This direction towards contextualization of *discourse* leads linguists to the theories on the scope of contextualization. Bhatia (2004: 9) distinguishes three levels of contextualization: textual, genre and social context level. The first or textual level is recognized by Bhatia as the textualization of lexical and grammatical features, which reflects the tendency of narrowing the term *discourse* to *text*. The second – genre level, refers to a particular communicative situation 'the immediate context' (ibid.), namely, to written discourse organization in the terms of cognitive structures that are used to distinguish particular genres. The conceptualization of this contextualization level contributes to the elaboration of the genre theory in the written discourse analysis. The following linguists have investigated the discourse organization: Hoey (1983) investigated

problem solution structures, Swales (1990), Bhatia (1993) have contributed to the development of the genre theory. The third level of discourse contextualization distinguishes the context in a much broader sense in the form of 'what makes a particular text possible'. It involves the social context and aims at investigating of discourse as a powerful instrument of social control (Bhatia, 2004: 11). Bhatia holds the opinion that this trend in written discourse analysis is investigated in the studies that examine the disciplinary variation, for example, Bhatia (1999), Hyland (2000). Bhatia shows with the help of the discourse analysis of particular texts that a text manifests all three discourse contextualization levels. Bhatia's theory of contextualization levels is summarized in his multi-perspective model of the written discourse analysis that illustrates his conclusion: 'I am using *discourse* as a general term to refer to any instance of the use of written language to communicate meaning in a particular context. The analysis of *discourse*, however, can be viewed in different ways, as a text, as genre, as professional or social practice' (Bhatia, 2004, 18- 19). In other words, he implies that the contextual interpretation of a *text* reveals the discourse manifestation in the said text.

Contextualization levels are also distinguished by Gee (2002: 7-8). He points out that 'we, as" applied linguists", are interested in how language is used "on site" to enact activities and identities'. He calls such language in use 'discourse with little d' and emphasizes that activities and identities are rarely ever enacted through language alone. Gee is convinced that When "little d" discourse (language in use) is welded with non-language"stuff", the"big D" Discourses are involved'. He, therefore, suggests dealing with the study of discourse within Discourses, which correlates with the theory proposed by Bhatia (2004) on the overlapping contextualization levels of his multi-perspective approach to the interpretation of language in use. Gee (2002: ibid) concludes that 'Discourses' always involve 'coordinating language with ways of acting, interacting, valuing, believing, feeling, and with bodies, clothes, non-linguistic symbols, objects, tools, technologies, times, and places' (ibid.). The notion of context and the way it manifests in linguistic code is considered in the chapter *Language Variation in Discourse*.

The multi-perspective approach to *discourse* complies with the theories that deal with the interrelationship between *discourse* and *text* from the point of their magnitude. According to Cook and also Widdowson's (2004) recent works, flexibility towards *discourse* and *text* magnitude confidently proves the role of discourse contextualization:

> *Discourse can be anything from a grunt or single expletive, through short conversations and scribbled notes up to Tolstoy's novel War and Peace, or a lengthy legal case.*
> Cook (1989: 7)

Cook (ibid) believes that the distinguishing feature of discourse is that 'it [discourse] communicates and is recognised by its receivers as coherent'. The idea about the discourse potential of very small, isolated textual units, for example, such words as public notices *private, silence, open, closed, push, pull* is the cornerstone of Widdowson's (2004: 8) theory that the interpretation depends on 'relating the text [irrespective of its bulk] to something outside itself, in other words, to other dimensions, that is to say, to the context: to where it is located, on the one hand, and how, on the other hand, it keys in with my knowledge of reality as shaped and sanctioned by the society I live in'. Therefore, Widdowson believes the conceptualization of the terms *text* and *discourse* to be crucial:

> *I recognize a piece of language as a text not because of its linguistic size, but because I assume it is intended to key into this reality. Texts can come in all shapes and sizes: they can correspond in extent with any linguistic unit: letter, sound, word, sentence, combination of sentences. I identify a text not by its linguistic extent but by its social intent. We achieve meaning by (interpretation) using language to engage our extralinguistic reality. Unless it is activated by this contextual connection, the text is inert.*
> (Widdowson, 2004: 8)

The contextual interpretation of the linguistic code or according to Bhatia, interpretation across the contextualization levels or according to Gee (2002) involving of the study of the discourse with 'little d' within the Discourse with 'big D' is what Widdowson (2004) refers to as *discourse*.

The approach to the conceptual definition of the terms *text* and *discourse* emphasizes the interrelationship between them and tends to widen the notion of *text* towards the concept of *discourse.*

> *A text, which is the product of that process, not only embodies the same kind of polyphonic structuring as is found in the grammar (for example in the structure of the clause, as message, exchange and representation), but also, since it is functioning at a higher level of the code, as the realization of semiotic orders 'above' the language, may contain in itself all the inconsistencies, contradictions and conflicts that can exist within and between such higher-order semiotic systems. Because it has this potential, a text is not a mere reflection of what lies beyond; it is an active partner in the reality-making and reality-changing processes.*
> (Halliday, 1985: 318)

Since *text*, a coherent unit with a beginning and end, is viewed as part of discourse it represents, linguists emphasize that it is a pragmatic unit (Quirk et al. 1985:1423) that fulfils communication in a communicative situation. Widdowson also (2004: 14) points out that *text* is not a simple manifestation of linguistic data, but 'it exists as a symptom of pragmatic intent', and it has discourse significance. As the production of text is the process, it reflects a process of the negotiating of intended meaning, or, in other words, the pragmatic process of discourse realization. Consequently, the resources of the language code are used 'to engage with the context of beliefs, values, assumptions that constitute the user's social and individual reality'.

Since Widdowson (2004) views *text* as a part of the discourse that has *discourse* significance, it seems that the notion of *text*, as Halliday and Hasan (1976: 2) have put forward and Virtanen has noted (1990: 450), has significantly expanded towards the concept *discourse*, thus developing from a descriptive structural unit towards a more processual unit. Widdowson elaborates this idea and concludes that *text* possesses a texture as well as is a unified whole in a given context, and therefore has discourse implication.

The above discussed linguistic theories that emphasise the conceptual definition of the notions *text* and *discourse* originate from the peculiarities of their interrelationship and allow concluding that *text* is part of discourse. It is a performance, pragmatic process of discourse realization and as such it comes from a particular contextual reality or *discourse*. Therefore *text* is part or manifestation of *discourse* and contains discourse potential. This approach widens the scope of the notion *text* towards the concept *discourse* and allows an interchangeable use of these concepts in the present study. This approach reveals *text* as a part of discourse, its manifestation before its discourse potential is activated through contextual interpretation of a text. Consequently, *discourse* is the activation of the connection of the *text* with its extralinguistic, contextual reality. Thus, the contextual reality or *discourse*, according to theories that are at the basis of the definition of *text* and *discourse* has widened through expanding the contextualization levels towards the social space in the interpretation of the discourse manifestation or *text*. The interrelation of the notions of *text* and *discourse* has been recognised in the present study. It is at the basis of the relation of language functions formulated and grouped by linguists with the clustering of the lexico-grammatical features used to perform these functions in texts as discourse manifestations (see chapter *Linguistic Variation in Discourse*).

SUMMARY

The conceptual definition of the notions *text* and *discourse* that proves their interrelationship is adopted in the present study. The term *Text* is used to denote a linguistic unit with a beginning and an end as a particular discourse manifestation and part of that discourse. *Discourse* is used to denote the contextual reality that embraces overlapping levels of context that a particular *text* belongs to and manifests itself through the functional use of language.

1.2 Linguistic Variation in Discourse

Discourse can manifest itself in spoken and written communication mediums. The focus of attention of the present Chapter is the theoretical background underlying language variation in written, and in particular, in electronic written discourse.

The written language that has emerged after the spoken language, according to Halliday (1985), is the result of cultural changes which were brought about by new communicative needs that could not be readily met by the spoken language.

Linguists have been investigating the subject matter of written language in order to identify its linguistic peculiarities. The theoretical considerations of these peculiarities, which are highly diversified, historically fall into two trends: the overestimation and underestimation of the written language in comparison with the spoken language. The overestimation of written communication refers to the period until the nineteenth century, when writing was considered to be the true form of language. However, this 'privileged' attitude towards writing gradually declined in the nineteenth century when linguists, for example, Daniel Jones in Britain, started investigating phonetics as a separate discipline. Moreover, the phonetician Sweet (1964) claimed that phonetics should be the basis of theoretical and practical studies of language and that the spoken form of language should take precedence over the written form. The focus on phonetics encouraged linguists to research speech as the primary and true form of language. For example, Hall (1964: 8-9) sides with Bloomfield by maintaining that 'Speech is fundamental and writing [is] only a secondary derivative', Fillmore (1981: 153) shares Hall's view by stating that written communication is 'derivative of the face-to-face conversational norm'. The neglect of written language, as Matsuda (2003) argues that the neglect of written language was most prominent in the United States between the 1940s and the 1960s, when the view of language as speech was institutionalized through the work of Leonard Bloomfield. However, the written language started gradually regaining its position in applied linguistics in 1960s and soon obtained the status of a separate discipline. The renaissance of the status of written language in 1960s has even given rise to theories that overemphasize the literacy by assuming that logical thought is possible only in literate societies.

Both trends, the underestimation and overestimation of written language, are criticized by linguists who emphasize the functional role of both communication mediums and deal with dpecific features of language variation in written discourse. For example, Nuan (1993: 8) writes: 'Although spoken language emerged before written language,

written texts are much more than merely 'talk written down'. Linguists who argue against the underestimation or overestimation of the role of writing (Hymes, 1984; Halliday, 1985; Hatch 1992; Nuan, 1993; Hughes, 1996; Crystal, 1997; McCarthy, 2001) remind that regarding writing or speech as 'better' is one-sided view on communication. Instead they suggest a compromise approach by stating that each of these two modes is a different communication system that has originated to perform specific as well as overlapping functions that mutually complement each other in communication, and therefore they possess salient features that deserve profound investigation.

The departure point of the present study is the assumption that written discourse is one part of the two specific, but in communication equally important and complementary mediums - spoken and written – that are characterized by linguistic expression through lexical and grammatical variation. The theories on register-pertaining language variation are explored according to the systemic functional approach and considered from two angles: firstly, focusing on the peculiarities of written and spoken discourse and, secondly, focusing on the multi-dimensional analysis (MDA) that reveals both peculiarities and similarities of spoken and written discourse.

1.2.1 Linguistic Manifestation of Register

The notion of *register* is closely related to the concept *genre,* and for this reason both terms are used by linguists who focus on the investigation of language variation. The present chapter deals with the application of notions *register* and *genre* in the investigation of language variation in texts as discourse manifestations and elucidates the use of these terms in the present study.

The dependency of the lexico-grammatical resources on contextual factors underlie the language use variation across texts, which represent particular *genres*, and outline the interrelationship of *register* and *genre*. The latter is a macro level concept, an abstract notion dating back to Aristotle's *Poetics.* Aristotle (1959) has implied, when discussing differences among 'epic' and 'tragedy', that distinctions, for example, between 'epic' and 'tragedy' are based on the way they are organized, presented, and perceived. Likewise he holds the opinion that rules and conventions of these *genres* impose content-based constraints on their development. Accordingly, the term *genre* has been used for many years to denote, alongside tragedy and comedy, various styles of literary discourse such as sonnets, romances, novels and other discourses that can be identified by their overall

structure. *Genres* reflect differences in external format and situations of use, and are defined on the basis of systematic non-linguistic criteria, for example, the poem, business letter, newspaper article, advertisement, and guidebook.

With the development of functional linguistics, the term *genre* has been adapted to denote different types of communicative events (Martin 1984; Swales 1990) and the study of texts as *genres* 'how texts are perceived, categorized and used by members of community' (Swales, 1990: 92). For this reason, since the end of the nineteenth century, as Connor (1999: 100) points out, 'the concept of *genre* has become a significant issue in applied linguistics'. It is generally assumed by functional linguists dealing with *genre* analysis that it is defined as a linguistic realization of a communicative purpose. This general definition is elaborated in applied linguistics by the following three approaches (Hyland, 2002: 17): *Systemic Functional (or Sydney School), New Rhetoric and English for Specific/Academic Purposes (ESP/EAP).*

According to the *Systemic Functional* approach (Martin, Christie and Rothery, 1987; Martin 1992; Slade 1999), *genres* are staged, goal-oriented, social activities that involve the use of language in a conventional, step-wise structure. The emphasis in the *Systemic Functional Approach* is on reaching the communicative purpose by introducing steps or text development stages that can help writers to achieve their communicative goals. That is why the *Systemic Functional* approach correlates with Halliday's proposed relation of form, function and context.

The *New Rhetoric* approach, which is promoted by Miller (1984), Freadman and Medway (1994), Bazerman (1994) Berkenkotter and Huckin (1995), defines *genres* as 'typified rhetorical actions based in recurrent situations', (Miller, 1984: 159), thus emphasizing the action that 'it [genre] is used to accomplish' (Miller, ibid. 151). The role of typified actions in reaching the communicative purpose is also brought out by Bazerman (1994: 79):

> …systems of complex literate activity constructed through typified actions – typified so that we are all to some extent aware of the form and force of these typified actions … By using these typified texts we are able to advance our own interests and shape our meanings in relation to complex social systems.
> (Bazerman 1994: 79)

This approach shifts the emphasis from form to actions: the contextual features, the assumptions and aims of the communities who use the *genres*. Although texts that belong to particular *genres* have an ordered, unified form, they are at the same time ongoing processes of discourse production, reception and are tied to other texts. The *New Rhetoric* approach aims at a more open view on the diversification of communicative situations that,

for example, can arise due to the advent of technology-enhanced communication and contributes to the ongoing discussion about the structural stability of *genres*: how far the identical situations tend to recur and how far a particular *genre* might be subjected to bending.

The *ESP/EAP* approach, which, as Belcher (2004, 2006) asserts, has willingly embraced the insights of the two previously considered schools of genre theory, is proposed by Swales (1990) and Bhatia (1983) who define *genres* as structured, strategic communicative events carried out in specific discourse communities the members of which share broad communicative purposes. Discourse community and its members agree on acceptable, specific, structural conventions of *genres*. For example, research articles, presentations, books and other texts belong to different *genres* because their sets of communicative purposes or schematic structures are different. Consequently, the emphasis is placed on the consistency of communicative purposes and *genres* are framed according to their formal properties, namely, their typical schematic structures, which are recognisable to the users of a particular *genre*. This approach is concerned with the distinction of discourse forms, which, as Briggs and Bauman (1992) note, correlates with Aristotle's approach to *genres* as the *ESP/EAP* puts an emphasis on the elaboration of the structures of academic and professional *genres*. One of the most recognized approaches is Swale's (1990; 141-148) structural model 'Create a Research Space' that proposes an elaborated move structure for research article introductions. These moves are motivated by the rhetorical need to bring out the significance of a particular research within a discipline for which this introduction is written. Each move is subdivided into steps that conventionally realise the underlying purpose of the introduction. Similar researches within professional and academic *genres* in English have been conducted by other linguists, for example, the investigation of methods sections of research articles (Paul and Charney, 1995), grant proposals (Connor and Mauranen, 1999), abstracts (Hyland, 2000). Accordingly, *ESP/EAP* theorists have moved the *ESP/EAP* towards more socio-historical view of genre.

In spite of their diversification in the definition of the way how the communicative purpose is reached (through staged, goal-oriented actions, or structured strategic events, or rhetoric actions), all three approaches to genre are rooted in the functional approach to the language use in conventionalized, action-oriented settings that impose constraints on the choice of lexico-grammatical features. Apart from the conventionalized, structural aspect, *genres* as units are constituted by the variation of lexical and grammatical resources.

Accordingly, *genre* can manifest itself only through a specification of the contextual factors of discourse field, tenor and mode (presented further in this chapter) and accordingly a description of the linguistic features realized in the transactional, interactional and textual components of a particular *genre.*

Increasingly, during the recent years, analysts have penetrated beneath the *generic* staging of the texts in applied linguistics in order to investigate frequently occurring linguistic features and explain their functional role within texts as discourse manifestations, for example, the investigation of imperatives (Swales et al., 1998), theme (Gosden, 1993) and hedges (Hyland, 1996; 1998). This growing trend has been determined by the current tendencies to lay an emphasis on the investigation of language variation in applied linguistics.

Hyland (2002b: 19-20) argues that the functional investigation of the linguistic characteristics of texts can help avoid the oversimplified, mono-functional perception of these texts. Consequently, Bhatia (1997; 1999) points out that indirect purposes, or 'private intentions, may be expressed simultaneously with more public or 'socially recognized' ones. Therefore, as Hyland (ibid.) asserts, it is essential to carefully validate the moves 'in terms of linguistic features used in the texts'. Moreover, the importance of the functional investigation of linguistic features has become topical due to the diversification of the communicative contexts in relation of the development of new technology-based communicative mediums. Obviously, these trends encourage linguists to suggest a more open view on *genres* (Bhatia, 2004: 25):

> I would like to turn my back to the ideal world, as it were, and try to face the real world of discourse, which is complex, dynamic, constantly developing and often less predictable. There are regularities of various kinds, in the use of lexico-grammatical, discoursal and generic resources; there are rhetorical situations, which often recur, though not exactly in the same form and manner; there are expert and well-established users of language in specific disciplinary cultures who try to exploit, appropriate and even bend generic expectations in order to be innovative and effective in their use of language. All these factors make the real world of discourse complex and yet interesting'.

(Bhatia, 2004: 25)

Accordingly, genre researchers currently tend to see *genre* as contextual, dynamic and varied phenomenon.

Another aspect that urges the investigation of linguistic variation is the overlapping relationship of *genre* and *register*. This relationship is approached by linguists from two perspectives: on the one hand, from the angle of their distinguishing features and, on the other hand, from the angle *of genre* and *register* reciprocity. For example, Swales (1990) has summarized the unique aspects of *genres* and *registers* by stating that *registers* are

overall correlations of linguistic features in texts with appropriate contextual and situational factors on the formality and informality continuum, but *genres* set structural conditions on different parts of text such as its beginning, body, and ending. Ventola (1984) and Martin (1984, 2003) have attempted to provide the theoretical framework for the distinction of *register* and *genre.* They view *register* and *genre* as different levels: *genre* as an additional level of context beyond field, tenor and mode. Thus, *register* is the pattern of linguistic choice, but *genre* is the pattern of *register* choices. Ferguson (1994: 20-21) suggests the following distinction: *register* variation is associated with 'a communication situation that recurs regularly in a society', while *genre* variation is associated with 'a message type that recurs regularly in a society.

Lee (2001: 46-47) concludes that 'it is useful to see the two terms *genre* and *register* as really two different angles or points of view, with *register* being used when we are talking about lexico-grammatical and discoursal-semantic patterns associated with situations (i.e., linguistic patterns), and *genre* being used when we are talking about memberships of culturally-recognisable categories'.

However, researchers admit that *register* and *genre* connection, especially in the case of restricted *registers* (e.g. weather forecast) tends to be so overlapping that, as Biber (1995) argues, linguists' attempt to provide a clear theoretical framework for distinction of *genre* and *register* has remained unattained. Correspondingly, linguists (Trosborg, 1997; Hyland, 2002b; Lee, 2001; Bhatia, 2004) emphasize the *genre* and *register* reciprocity. R*egister* can cut across several *genres*, while the texts belonging to the same *genre* can show considerable *register* variation. For example, according to Bhatia, the scientific *register* can cut across the following *genres*: a research article, text book and academic essay. Similarly, an academic essay as a *genre* can display *register* variation (scientific, legal, business) depending on the discipline it belongs to: science, law or business. Consequently, various *register* choices can be more or less appropriate or more or less effective in the realization of particular *genres*. Thus, one *register* may be realized through various *genres*, in this sense *genres* are subordinated to *registers.* Conversely, one *genre* may be realized through a number of *registers* just as a *genre* constrains the ways in which *register* factors of *field, tenor, mode* can be combined. Lee (2001: 48) also provides the terminological clarification about the application of the terms *genre* and *register.* According to him the concept '*genre variation* makes no sense as a parallel to *register variation'*. Language (registers) can vary across genres, but it is tautologous to talk about genres (text categories) varying across genres or situations. In other words, the exploration

of the differences among genres includes the way the language varies because of social and situational characteristics and other genre constraints (registrar variation), but the way texts vary because of their categorisation.

Genre and *register* reciprocity shows that there can be a wide range of language variation patterns across texts that represent a particular discourse. Therefore, the functional investigation of language variation across registers can shed light on the *genre*-pertinent choices of lexico-grammatical features across texts developed in context embedded communicative situations across disciplines.

This assumption correlates with the multi-purpose needs of the students who have intended to qualify for *BA/MA* degree in *English Philology* at the *University of Latvia*. The extensive study of Kramiņa (2000) has proved that the needs of these students range from teaching various target groups (young, adolescent learners, learners with specific needs and adult learners specializing in different fields of science, law, medicine, art, economy and other disciplines) to interpreting and translation (technical texts of various disciplines, fiction of various *genres*). These multi-purpose needs of students can be addressed through involving them in linguo-functional writing and research activities devoted to the application and analysis of the variation of linguistic features in electronic texts that range from transactional to interactional written communication envisaged in e-courses that are integrated in the *BA/MA* programmes.

The above investigations in the theory of genre in functional linguistics allow applying the notion of genre to label *NNSE* electronic texts arranged in groups according to their communicative purpose. As the present study focuses on linguistic peculiarities of students' electronic texts, arranged according to their communicative purpose, these peculiarities are explored from the angle of their register variation. Therefore the present chapter further focuses on the notion of register and its dimensions.

The concept of register is rooted in the notion of varieties of language, grouped by linguists into user-related and use-related varieties. User-related varieties or dialects are associated with different groups of speakers, namely, language users living in a particular location or their belonging to a particular demographic group. Use-related varieties or *registers* stem from communicative circumstances and purposes. Biber (1995: 17; 2004: 20 – 21) argues that variation of linguistic features across *registers* are more extensive than across dialects: 'regardless of any dialect differences, speakers using the same *register* are doing similar communicative tasks; therefore, in most basic respects the characteristic grammatical features of a *register* are relatively constant across speakers and dialects'.

The research in use-related varieties has a comparatively long history. It dates back to anthropological linguists who investigated non-western languages in various social contexts in the first half of the twentieth century. Malinowski (1923/46) was one of the first researchers who claimed that language became intelligible only when it was placed within its context of situation. Firth (1935, 1950, 1951), influenced by Malinowski's investigations, extended the notion of context of situation by pointing out that a description of a context allowed predicting what language would be used, thus pursuing functional-semantic tradition. This trend was continued in the second half of the nineteenth century by a range of studies dedicated to the exploration and description of the ways in which linguistic forms were influenced by the purpose and the context of the communicative situations (Hymes, 1984; Ferguson, 1994). Hymes (ibid. 44) maintains that the investigation of *register* should be one of the main concerns within the research in linguistics: '[the] sociolinguistic perspective […] has the possibility of taking the lead in transforming the study of language, through developing and consolidating the systematic study of verbal repertoire [*register* variation]'. Ferguson (ibid.16) brings out the essence of *register* by stating that it 'is the linguistic difference that correlates with different occasions of use.' However, the major contribution to register theory has been made by Halliday's (1985) research in systematic correlation between the organization of language itself and specific contextual factors.

This correlation underlies the purpose of the present study which is to investigate the linguistic elements in electronic texts developed in a specific communication medium - online environment. The peculiarities of online communication require that the elicitation of linguistic variation must be based on the research of contextual factors. Therefore, the contextual factors and dimensions that determine the choice of linguistic features in a discourse have been considered in detail.

The theory of contextual factors is based on the researchers' investigation of the elements of communication, language functions and their interrelation (*Table 1*).

Table1: **Elements of communication and language functions**

Elements of communication	**Language Functions**	
	Functions related to elements	**Macro functions**
the addresser: the person who originates the message	the emotive function: communication of the inner states and emotions of the addresser	interactional (Brown & Jule) social-expressive (Lyons) interpersonal (Halliday)
the addressee: the person to whom the message is addressed	the directive function: seeking to affect the behaviour of the addressee	
the channel/mode: the medium through which the message travels	the phatic function: maintaining social relationships	
the message form: the particular lexico-grammatical choices of the message	the poetic function: the choice of the particular form(s) in the message	
the topic: the information carried in the message	the referential function: carrying information by description of persons, things, events, circumstances, activities etc.	transactional (Brown & Jule) descriptive (Lyons) ideational (Halliday)
the code: the language or dialect used	the meta-linguistic function: focusing on the code itself to clarify or to negotiate it	
the setting: the social or physical context	the contextual function: creating a particular kind of communication	textual (Halliday)

Buhler (1999) has proposed three functions, emotive, directive, and referential, Jakobson (1960) has expanded this initial three-function model by adding phatic, metalinguistic and poetic functions of language. Moreover, Jakobson has matched the elements of communication with functions. Hymes (1984) has suggested adding, the setting or the contextual function to Jakobson's model.

All these functions, as linguists emphasize, are involved in communication, yet a particular function tends to predominate in a certain communicative situation. For example, the purpose of referential function is to carry information. Accordingly, this function predominates in Text 1, the government report (Biber, 1995: 138) that aims at informing about the problem-points in oral training of pupils:

Text 1

> The initial oral training is too rarely continued and developed in the larger stages and many pupils do not progress beyond the standard of speech they had reached by the end of the second year. Many teachers feel that they cannot afford the time necessary for the development of oral work, but in most cases it is not additional time which is required so much as more systematic and purposeful training in the correct use of more difficult speech forms.

The purpose of the phatic function, however, is to maintain social relationships, and therefore it predominates in the dialogue represented in Text 2:

Text 2

L: Hi, Susie!
S: Hi Lucy!
L: How are you?
S: I'm very well, thanks. And you?
L: I'm fine, thanks.
S: How's John?
L: He's fine, thanks. How's Pete?
S: He's very well, too.
L: Bye, Susie. Nice to see you.
S: Nice to see you, too. Bye.

Due to the variation of these functions in particular communicative situations, linguists have summarised them in macro functions (see Table 1), that have been named differently by various researchers. Thus, Brown and Jule (1983: 1) point out that interactional functions, or social expressive according to Lyons (1997) and interpersonal by Halliday (1991) are 'involved in expressing social relations and personal attitudes'. Whereas transactional functions, as Brown and Jule (ibid.) denote, or descriptive according to Lyons (ibid.) and ideational by Halliday (ibid.) 'serve in the expression of content'. Textual function, according to Halliday (ibid.) is a function of the organizing of information.

On the basis of the three macro functions of a language, Halliday (1985), has proposed his major contribution to the theory of context, namely, the systematic correlation between the organization of the semiotic system, language, and specific contextual factors, as they are the three kinds of meanings language is structured to make field (function for relating experience), tenor (function for creating interpersonal relationships), mode (function for organizing information). Thus, Halliday has claimed that each of these factors have linguistic consequences, namely, significant impact on the choice of lexico-grammatical features in a particular manifestation of discourse, a text. On the basis of a detailed analysis, he has demonstrated that it is possible to identify parts of a language system that are connected with the realization of each of the three types of contextual factors. Namely, the three types of meaning or macro functions can be related upwards to context and downwards to lexico-grammar. Halliday, Hasan (1985) and other linguists (Quirk et. al., 1985; Greenbaum, 1996; Valdmanis, 1999, 2002; Hyland, 2002; Bhatia, 2004, Eggins, 2004) label the mentioned contextual factors – *register*.

The first factor, discourse field, refers to the type(s) of the communicative activity and includes the description of the participants, process, and circumstances that these activities involve by focusing on 'what is happening' (Halliday, Hasan, 1985: 12). For

instance, in ’complaining/asking about the terms of delivery’ the factor is ’complaining’ or ’asking’. Discourse field is a transactional (ideational, descriptive) component that covers the corresponding linguistic content. Thus, in the case of ’complaining/asking about the terms of delivery’, the information about the circumstances and outcomes of the situation that has caused the complaint has been included. Eggins (2004: 103) notes that some researchers label field as topic. It has to be noted that there is a similar terminology variation in the Latvian language.

Martin (1992: 536), however, suggests even a broader definition, namely, the discourse field according to him captures an institutional focus or social activity type. Eggins (2004) concludes that the discourse field varies along one dimension of the situational context – technicality. This dimension ranges communicative situations from technical to everyday situations. Technical situations are characterised by a more specific knowledge that interactants are supposed to possess regarding the activity, whereas in everyday situations the interactants are supposed to possess common knowledge. Eggins (ibid. 107) claims that knowledge constituting a discourse field can be represented in taxonomies, which reveal a significant difference between the complexity of technical and common sense situations that manifest themselves in the language use. The way the linguistic features vary depending on the type of the communicative activity, namely discourse field, is presented in Text 3 (Biber, 1995: 152), the discourse field of which is the performance of down-draught systems, and Text 4 (Biber, 1995: 132), the discourse field of which is Christmas shopping. Thus, the technical situations, as shown in Text 3, are characterised by the specific knowledge of an activity, deep taxonomies, which manifest themselves, for example, in an excessive use of specific technical terms (*thermal currents, velocities, moulds*), abbreviations (*ft., in.*) as represented in Text 3. Verbs tend to reflect technical processes (*require, exceed*). The common sense situations, exemplified in Text 4 about Christmas shopping, are characterized by common knowledge or little assumed knowledge of an activity expressed by everyday terms or words that are commonly understood (*shopping, things*) full names and standard syntax. Accordingly, discourse field, as summarised by Eggins (ibid. 110) is realised through transitivity patterns, namely, the patterns of processes (the use of verbs), participants (the use of nouns) and circumstances (prepositional phrases of time, manner and place).

Text 3

> It follows that the performance of down-draught systems can be improved by the influence of cross draughts only if the thermal currents are blown into exhaust air streams at higher velocities than the cross draughts, so that the resultant direction of all dust-bearing air streams is towards the grid.
>
> The exhaust air volume required by the 6-ft. x 4-ft. grid with the 8-in. deep hot and cold moulds and the 16-ins. Deep cold moulds tested in the absence of appreciable cross draughts exceeded the volumes required by the 4-ft. 6-in. x 3-ft. 6-in grid by between 25 and 40 per cent.

Text 4

> How are you doing? I'm here at work waiting for my appointment to get here, it's Friday. Thank goodness, but I still have tomorrow, but this week has flown by, I guess because I've been staying busy, getting ready for Christmas and stuff. Have you done your Christmas shopping yet? I'm pretty proud of myself. I'm almost finished. Me and L went shopping at Sharpstown last Monday and I got a lot done, I just have a few little things to get. Thanks for the poster, I loved it, I hang it in my room last night, sometimes I feel like that's about right.

The second factor, discourse tenor, according to Halliday, Hasan (1985), characterises the social relations, shared knowledge and roles of the addresser and the addressee covering the communicative functions that relate to interactants' roles. These roles in detail include the communication setting (the place and time of the communication) and interactants' relationships (the level of acquaintance, friends, acquaintances, colleagues, etc). Thus, for instance, in 'customer/accountant asking/complaining about the terms of delivery', the factor is 'customer' or 'accountant'. In the given example, the customer is an individual, addressing another individual that representing an institution that determines the choice of particular linguistic resources. The discourse tenor is an interactional (social-expressive, interpersonal) component. Poynton (1984) has suggested a sub-classification of discourse tenor that according to her comprises three dimensions: power, contact and affective involvement. The dimension of power, positions communicative situations in terms of roles the interactants play, namely, of equal or unequal power (for example, roles of friends and roles of boss/employee). Contact denotes the frequency of interactancts' contact (for example, roles of family members or acquaintances). The dimension of affective power reveals the level of interactant's affective involvement, namely, high or low involvement (for example, the emotional involvement of friends or the involvement of colleagues). The considered roles of interactancts provide a direct link between language and context. According to the considered tenor dimensions, Eggins (2004) brings out the contrast between two situation types: the informal and formal situations. Eggins (2004: 100) emphasizes that a formal situation is the one where the interactants' power is unequal, contact is infrequent and it displays low affective involvement, whereas informal situation would involve interactants

of equal power, who frequently see each other and who are affectively involved. For example, the communicative situation of Texts 3 does not envisage interactants' frequent contact and represents their low affective involvement. The communicative situation of Text 4 shows that there is an occasional contact between the interactants, who are of equal power and show some affective involvement in their communication. Accordingly, in the exemplified communicative situations, the three tenor dimensions (power, contact and affective involvement) display variation of linguistic consequences that are realized through the following mood patterns. Firstly, there is a difference in sentence structures (Text 3 contains only declarative sentences, but Text 4 apart from declarative sentences includes also interrogative sentences.). Secondly, in Text 3 a greater degree of certainty is expressed than in Text 4 (I guess). Thirdly, Text 4 contains attitudinal words, expressions of intensification (I love, pretty). Finally, the lexis in Text 3 is more neutral than in Text 4 and Text 3 contains only full forms, whereas Text 4 displays such abbreviated forms as I'm, I've. Poynton (1984) exemplifies the correlation between all three tenor dimensions (power, contact and effective involvement) with the help of terms of address. She shows that in the case of equal power, frequent contact and affective involvement first names, also nicknames and diminutive forms are used. However, in communicative situations involving unequal power, infrequent contact and low affective involvement titles, in some cases no names or full forms of names are used in addressing a person.

Halliday, Hasan (1985) exempliefy the third factor, discourse mode or the channel of communication through which the activity (*field*) is undertaken via enacting particular social relations (*tenor*), in the following way: 'customer/accountant asking/complaining about the terms of delivery in speech or in writing'. Discourse mode is a textual component and distinguishes between 'speech' or 'writing'. Discourse mode, as emphasized by Martin (1984), involves two dimensions: spatial/interpersonal distance and experiental distance. The first continuum, spatial/interpersonal distance, embraces the situations ranging from immediate feedback between the interactants, for example, face-to-face discussion, to the situations excluding immediate feedback, namely, visual or aural contact between the writer and the reader(s), for example, writing a book. The second dimension, experiential distance, ranges situations depending on the distance between the language and the social process. In the case of close distance language is used to accompany social process for example, playing a game, whereas in the case of a more significant distance language is used to constitute social process, (for example, writing a piece of fiction). These two dimensions of discourse mode, as emphasised by Eggins (2004), determine the language

use and reveal the contrast between spoken and written discourse, the linguistic consequences of which are considered in detail in chapter *1.2.2 Contrasting of Written vs Spoken Discourse.*

Each of the three factors and the dimensions comprised by these factors are responsible for variation of lexico-grammatical features within a particular manifestation of discourse – a text. Field, tenor and mode, which are collectively referred to by researchers as *register*, determine the level of formality in a particular instance of communication. The formality levels are perceived as continua ranging from highly formal to highly informal *register*.

Linguists, however, tend to vary the application of the terms to the contextual factors. For example, Greenbaum (1996: 15-16) labels them in the following way: field – *register*, tenor – formality, and mode – medium. Moreover, he notes that some researchers base *register* variation solely on the discourse field, and therefore the field is labelled as *register*. In this connection, Halliday, and Hasan (1985), (1995: 7) have pointed out that *registers* 'can be identified at any delicacy of focus [at any level of generality]'. Correspondingly, there could be distinguished very general *registers* such as speech and writing or highly specified *registers* that are defined in terms of multiple dimensions correlating with the situational factors. However, such diverse application of terminology and narrowing of the notion of *register* to one dimension can constrain the perception of the details of each contextual factor (field, tenor, and mode) and the dimensions within the contextual factors as well as favour the implementation of one-dimension based generalizations to a particular *register*. This a tendency, according to Bhatia (2004: 32), was observable 'in early *ESP* (*English for Specific Purposes*) work, which primarily focused on such field-dominated *registers* as science or business, functional variation in them was seen primarily in the terms of specialist lexis and some surface-level syntactic features alone.' Trosborg (1997) is convinced that it is inappropriate to apply the term *register* to the linguistic variation that is adjusted to only one contextual factor, for example, 'employer register' (focusing on tenor), 'written *register*'(focusing on mode).

Within the framework of the present study, the notion of *register* is applied to characterise the context of a particular communicative situation, namely, the situational factors (*field, tenor and mode*) and their dimensions, thus embracing transactional, interactional and textual macro functions of students' electronic discourse. The theories of language variation in discourse and, in particular, in written and electronic written

discourse are based on the analysed theories regarding the situational factors and dimensions in the mentioned discourse.

1.2.2 Contrasting of Written vs Spoken Discourse

Distinctive features of written discourse are revealed by linguists through contrasting language variation in spoken and written discourse on the basis of two dimensions of discourse mode: spatial/interpersonal distance and experiental distance. The unique features of written discourse originate from the main peculiarity of writing: the manner of production. Writing is a semiotic system that uses 'graphical substance' in the form of marks on surface: letters, punctuation, capitalization, spatial organization. Cook (1989: 8, 10) holds the view that due to the rich graphical possibilities, writing, like spoken communication, possesses paralinguistic features: handwriting (font), colours, underlining, italics, boldface, tables, diagrams that contribute to the clarity of information. Crystal (1997: 181) even claims that the peculiar features of writing have originated vocabulary items (longer names of chemical compounds) that appear only in writing as well as the graphical characteristics of writing has developed the genres that cannot be structurally conveyed by reading aloud (timetables, graphs).

Specific features of written discourse have been revealed by linguists who have contrasted spoken and written situations of language use. This contrasting is based on the end points of two dimensions of the discourse mode, namely, spatial and interpersonal distance (distance between interactants) as well as experiential distance (distance between language and social process), as considered in chapter Register and Genre. Linguists (Goody, 1977; Tannen, 1982; Chafe, 1982; Brown and Yule 1983; Chafe et.al.(1986), Cook, 1990; Olson et al., 1985; Hatch, 1992; Hughes, 1996; Crystal, 1997, 2001, McCarthy 2001) who have explored the specific features of spoken and written language, focusing on discourse mode have investigated the stereotyped communicative situations, for example, the contrasting of expository texts and face-to-face conversations. The above linguists have considered spoken and written communication by contrasting them according to the two dimensions of discourse mode: spatial/interpersonal distance and experiental distance. Table 2 contains the summary of the specific features of spoken and written communication grouped according to the mentioned dimensions.

Table 2: **The characteristic features of spoken/written language based on the discourse mode dimensions: spatial/interpersonal distance and experiential distance**

Discourse mode	Characteristic featues
Spoken (close spatial/interpersonal and experiential distance)	interactive (Crystal) involved (McCarthy) time-bound, spontaneous (Crystal) aural (Hughes) prosodically rich (Crystal) fragmented, real-time (McCarthy) loosely constructed (Crystal) unplanned (Hatch) contextualized (Tannen, Chaffe) implicit (McCarthy) immediately revisable (Crystal)
Written (significant spatial/interpersonal and experiential distance)	uninteractive, decontextualized (Crystal) contrived (Crystal) visual, static (Hughes) integrated (Chaffe, McCarthy) explicit , lapsed-time (McCarthy) factually communicative (Crystal) planned (Hatch) elaborately structured (Crystal) space-bound (Crystal) detached (Tannen, McCarthy) repeatedly revisable (Crystal)

The summary of the characteristic features (Table 2) shows that communicative situations in which people use spoken language are aural, spontaneous, time-bound and thus typically interactive, involving two or more interactants. In such communicative situations people produce prosodically rich, fragmented, loosely constructed, contextualized and immediately revisable discourse. In contrast, the characteristic features of the communicative situations in which people use written language reveal that these situations are uninteractive, decontextualized, lapsed-time, visual and static. Language is used to reflect or provide information on a more specific topic, and for this reason the writer produces a factually communicative, planned, elaborately structured text.

Contrasting of the said two communicative modes, serves the purpose of exploring the specific features of written discourse, as the language variation based on the end points of discourse mode reveal the spoken/written language divide. Specific linguistic features are illustrated by examples of lexico-grammatical elements found in a written text (Text 5) and conversation (Text 6). Although the presented examples are too small for detailed statistical calculations, they can exemplify the function of linguistic features.

In order to illustrate language variation discussed above, the term lexico-grammatical features is introduced in the present study. The term has been suggested by Biber et.al. (2004: 12) and Halliday (2004: 43), who point out that lexis and grammatical

forms closely coexist in communication. They emphasize that in authentic communication grammar and lexis, which are often treated as independent components of the English language, are brought together by people's knowledge of word behaviour (lexis) and grammatical patterns. For example, there is a particular set of verbs that commonly occur with a *to clause* (want, like, seem etc) and a different set of verbs that commonly occur with a *that clause* (think, say, know etc).

Text 5: Expository text

> In such studies, it is sometimes helpful to examine some of the articulatory movements, in addition to the acoustic or spectrographic signal, in order to understand changes due to prosodic conditions, given the same phonemic materials or word sequence. Generally, mandible position varies greatly depending just on the degree of prominence attached to the word: the more prominence, the lower the mandible. Due, at least in part, to the complex physical and physiological interaction among articulatory organs, other articulatory gestures such as tongue body position also change due to prominence in the particular utterance situation in a rather complex way. Some researchers assume that it is primarily the acoustic or auditory properties rather than articulatory states that are inherent to phonemic properties. However, quite often, formant frequencies also are considerably different for the same vowel in the same phonological context when prominence varies. (143 words)
>
> (Fujimura, O., Erikson, D. (1997) In: W.J. Hardcastle, J.A. Laver (eds.) *Handbook of Phonetic Sciences,* Blackwell Publishers p. 87)

Text 6: Face-to-face conversation

'I've been upstairs,'she said.
'Oh yes?'I said.
'I found a hair, 'she said.
'A hair?' I said.
'In the bed,' she said.
'From a head?' I said.
'It's not mine,' she said.
'Was it black?' I said.
'It was,' she said.
'I'll explain,' I said.
'You swine,' she said.
'Not quite,' I said.
'I'm going, 'she said.
'Please don't ,' I said.
'I hate you!' she said.
'You do?'I said.
'Of course,' she said.
'But why?' I said.
'That black hair, 'she said.
'A pity,' I said.

'Time for truth,' she said.
'For confessions ?' I said
'Me too,' she said.
'You what ?' I said.
'Someone else, 'she said.
'Oh dear, 'I said.
'So there!'she said.
'Ah well,' I said.
'Guess who?'she said.

‘Don’t say,‘ I said.
‘I will, ‘she said.
‘Your fried, ‘she said.
‘Oh damn, ‘I said.
‘And his friend, ‘she said.
‘Him too?’I said.
‘And the rest, ‘she said.
‘Good God, ‘I said.

‘What’s that?’ she said.
‘What’s what?’ I said.
‘That noise?’she said.
‘Upstairs?’ I said.
‘Yes,‘ she said.
‘The new cat, ‘ I said.
‘A cat?’ she said.
‘It’s black,‘ I said.
‘Black?’she said.
‘Long-haired,’ I said.
‘Oh no,‘ she said.
‘Oh yes,’ I said.
‘Oh damn,’ she said.
‘Goodbye, ‘I said.

‘I lied,’ she said.
‘You lied?’ I said
‘Of course,’ she said.
‘About my friend?’ I said.
’Y-ess,’ she said.
‘And the others? ‘I said.
‘Ugh, ‘she said.
‘How odd, ‘I said.
‘I’m forgiven?’she said.
‘Of course, ‘I said.
‘I’ll stay?’ she said.
‘Please don’t,’ I said.
‘But why?’ she said.
‘The cat,’ I said.
‘Was white,’ I said.
(157 words without reporting words) (B.Patten Hair Today, No Her Tomorrow)

1 The expository text (*Text 5*), which is transactional, is detached, non-interactive because there is a considerably low potential of immediate response and possibility to rely on particular communicative situation. Goody (1977: 78) is convinced that ’written communication permits words and sentences to be examined out of their original contexts […] where they appear in very different and highly abstract context’. Conversation (*Text 6*) which is interactional, on the contrary, is contextualized and interactive. Thus, the major difference between spoken and written language is that in spoken language action processes are performed by human actors, whereas written language predominantly deals with abstract ideas linked by relational processes, which leads to heavily nominalized language in such texts.

1.1 The information in *Text 5* is placed in 5 sentences. Each of these sentences contains two or more clauses that are linked with such connectives as in addition, in order to, due to, however. This text presents high nominal style, as there are 36 nouns, which means that actions are expressed in a form of nouns (physical and physiological interaction), but this text does not display personal pronouns apart from 2 cases of pronoun it. The conversation, *Text 6*, is split into 66 utterances. Most of the utterances describe concrete actions (I've been upstairs, I found a hair, I'll explain, I'm going, you lied etc.) presented by verbs and performed by human actors (I, you), thus showing the direct reference to the conversation participants. For this reason Text 6 contains considerably small number of nouns, namely 20, in comparison with *Text 5*, but considerably high frequency of pronouns (23).

1.2 Each sentence in the expository text contains clear structural framework: subject, verb and complement (Some researchers assume that it is primarily the acoustic or auditory properties rather than articulatory states that are inherent to phonemic properties.), but the conversation displays a range of elliptical sentences (In the bed.) (That black hair.).

1.3 Text 5 contains only full grammatical forms (It is sometimes helpful …,), but the exemplified spoken text, conversation, presents a frequent use of contracted forms (I've been, it's not mine, what's that).

2 Text 5 is explicit and informative, as the message sender/recipient cannot rely on prosodic and paralinguistic features that are characteristic of implicit and interactive spoken communication. Thus, Chafe, et.al (1986) hold the view that precise lexical choice is a very difficult production task that is rarely accomplished in speech. Tannen (1982: 78) maintains that writing requires addressing the reader only through content; therefore, writers are forced to encode meaning only through lexis, thus reaching high lexical density.

2.1 The written text contains specific vocabulary (sequence, properties, prominence, to examine), but the vocabulary of conversation is rather generalized (to find, to go, to stay, upstairs) including slang (damn).

2.2 Comparatively low incidence of the verb ‘to be’ (4) in written text in comparison with the face-to-face conversation that contains 7 instances of the verb ’to be’.

2.3 Conversation contains also such interjections as oh yes, oh dear, oh damn, yes, ugh, which are not found in the transactional text.

3 Written text (*Text 5*) is integrated and planned because its space-bound and contrived character allows the elaboration of the text without the reader knowing that the initial drafts have contained any inadequacies: correct errors, improve the rhetorical organization of the text by rearranging the parts of it. Conversation (Text 6), in contrast, is fragmented and unplanned, reflecting the real-world sequencing of actions by the interactants. These peculiarities lead to the differences of the rhetorical patterns of spoken and written language.

3.1 The clauses of *Text 5* tend to begin with nominalizations (mandible position, formant frequencies), whereas the clauses in the conversation begin with human interactants (I, you).

3.2 Written text follows standard grammatical conventions, so only full grammatical forms (e.g. it is sometimes helpful) are used in it, but conversation shows simplification of grammatical structures (e.g. I’ll, I’m, don’t).

3.3 Written text contains detached lexical marking of stance that shows the detached expression (e.g. mandible position varies greatly); however, the lexical stance marker in conversation (e.g. I hate you!) is directly evaluative.

The investigation of the two idealised written and spoken communication situations shows that such contrasting can highlight the most specific features of writing that are displayed through functional characteristics of two stereotyped situations.

The value of contrasting of spoken and written texts, as Brown and Yule (1983), Cook (1989), Nuan (1993), Olson et al. (1985), Hatch (1992), Hughes (1996), Crystal (1997, 2001) have pointed out, is that, on the one hand it reveals that the communication medium can strongly determine the functional variation of language in idealised situations, and therefore show the language use differences in speech and writing. On the other hand, it demonstrates that the language functions can overlap across both communication mediums. For example, Hatch (1992:236) argues that some written genres can share particular sets of lexico-grammatical features which are characteristic of spoken language. This assumption confirms the linguists’ claims (Nuan ibid. 8) that written language can

perform similar broad functions as spoken language (to get things done, to provide information and to entertain) as both language modes can be used for the same purposes. It means that some written texts can be more like spoken texts than others, but some spoken texts can be more like written texts than others.

With these assumptions, linguists draw attention to the fact that the investigation of language variation through the contrasting of idealized situations of spoken/written communication demonstrate essential, however, one-sided and limited insight in the actual versatility of written discourse. Therefore the analysis examined above that underlies the skills based approach to language studies is a departure point for a more elaborate and context-bound approaches to functional variation of lexico-grammatical features in written discourse.

1.2.3 Multi-dimensional Analysis (MDA) of Discourse

The above contrasting of spoken and written texts from the angle of end-points of the dimensions of one register factor, communication mode, is rightly labelled by Street (1995) as the 'great divide'. He argues that such one-factor view from the angle of discreet poles conceals more than it reveals as it neglects the considerable diversity of contexts in which written communication is embedded. The relations among registers are too complex to capture the similarities and differences among them through the discreet poles, because the dimensions of register factors overlap and work simultaneously in discourse. McCarthy (2001) has concluded that linguistic variation in written discourse should be explored by the investigation of a range of texts and taking into account all situational factors that determine their register. McCarthy (ibid.) also has concluded that linguistic variation in written discourse should be explored by the investigation of a range of texts taking into account all contextual factors. In order to reveal the complexity of the linguistic realisation of spoken and written texts representing various genres, linguists have proposed (Biber, 1988, 1989, 1995), elaborated and applied (Tribble, 1999; Biber and Conrad, 2003; Xiao and McEnery, 2005) the multi-dimensional analysis (MDA) in their investigation of the linguistic characteristics of discourse.

The purpose of the MDA, initially put forward by Biber (1988), has been to provide comprehensive characteristics of register variation in discourse. It includes two stages: firstly, the identification of the dimensions of variation and, secondly, the specification of the linguistic similarities and differences of registers in relation to these dimensions.

Dimensions represent the empirically determined and quantitatively identified (by a quantitative, namely, statistical analysis of dimensions or factors) groupings of linguistic features that co-occur with significant frequencies in texts. Biber's (ibid.) MDA synchronic study of discourse comprises the exploration of 67 different linguistic features in 500 spoken and written texts representing 23 genres. Biber has distinguished six dimensions, interpreted in terms of communicative functions that are represented by the co-occurring linguistic features and ranged along the continuums of the six dimensions. The notion 'multi-dimensional' implies that several dimensions of variation have been applied to the analysis material – discourse. Biber (1988) has found that the relations among texts according to the six dimensions are determined by a complex assessment of the frequency of linguistic features that perform particular functions in discourse. The MDA is a corpus-based analysis of a collection of naturally-occurring texts by using automated computational techniques to perform the distributional analysis of linguistic elements across these texts in order to reveal relations among these texts. This approach combines quantitative and functional research techniques, as the quantitative analyses are considered in functional terms to identify the underlying communicative functions with the co-occurring linguistic features in the texts.

A concise overview of the six dimensions, linguistic features and their relation to language functions is presented in Table 3. The register dimensions formulated by Biber include linguistic features with positive loadings as well as some features with negative loadings, which are respectively designated with a plus or minus symbol (see Annex 1 for the summary of the applied dimensions, linguistic features, and the examples of the author of the present thesis from students' corpus). Biber has proposed six dimensions and the the following communicative functions.

Dimension 1, *involved versus informational production*, comprises 28 linguistic features. 23 features with positive loadings are more frequently used in involved non-informational discourse. However, 5 linguistic features with negative loadings are more frequent in informational discourse. Thus, according to Biber's findings, along Dimension 1, higher scores are presented by such involved discourse as face-to-face conversations, personal letters, spontaneous speeches and interviews, but lower scores by more informational discourse, namely, official documents, academic prose, press reviews.

Dimension 2, *narrative versus /non-narrative concerns*, contains 6 positively loaded linguistic features. These linguistic features are more frequently used in narrative discourse. Accordingly, romantic, science, adventure fiction, according to Biber's findings

show higher scores, whereas academic prose, professional letters, telephone conversations and face-to face conversations display lower scores.

Dimension 3, *explicit versus situation dependent reference*, clusters 4 features with positive loadings and 3 features with negative loadings. The features with positive loadings are more frequent in such discourse as official documents, professional letters, academic prose, which display higher scores on this dimension. However, the linguistic features with negative loadings are more frequent in the discourse that shows lower scores on this dimension, for example, adventure, romantic fiction, personal letters, face-to-face conversations, telephone conversations.

Dimension 4, *overt expression of persuasion*, contains 6 positively loaded linguistic features, which, as Biber has identified, are more frequent in discourse that aims at more overt expression of an addresser's viewpoint. Although the difference of the scores among the texts is less marked on this dimension, such texts as professional letters, personal letters present higher scores than face-to-face conversations and academic prose.

Dimension 5, *abstract versus non-abstract information*, comprises 6 positively loaded linguistic features that are more frequently used in such abstract discourse as academic prose, official documents, but less frequent in professional letters and considerably infrequent in personal letters and face-to-face conversations.

Dimension 6, *online informational elaboration*, clusters four positively loaded linguistic features used in discourse aiming at informational elaboration under strict real-time conditions. Accordingly, Biber (1988: 156-157) has explored that the highest scores are displayed by prepared speeches, interviews and spontaneous speeches, which possess an informational focus, but the speaker should address real-time production constraints. Accordingly, academic prose, professional letters display considerably lower scores.

Table 3: **Summary of lexico-grammatical features, functions and dimensions (Biber: 1988, 1995)**

Dimensions	Functions	Lexico-grammatical features
Dimension 1 Involved vs informational production	Discourse with interactional, affective, involved purposes associated with strict real-time production and comprehension to discourse with highly informational purposes.	28 features: (+) private verbs, that deletion, contractions, present tense verbs, 2nd person pronouns, do as a pro-verb, analytic negation, demonstrative pronoun, emphatics, 1st person pronouns, pronoun it, be as main verb, caus. subordination, discourse markers, indefinite pronouns, hedges, amplifiers, sentence relative, wh questions, possibility modals, non-phrasal coordination, wh clauses, final prepositions, (-) other nouns, word length, prepositions, type/token ratio, attributive adjectives

Dimension 2 Narrative vs non-narrative concerns	Discourse with primary narrative purposes to discourse with non-narrative (expository, descriptive) purposes.	6 features: (+) past tense verbs, 3rd person pronouns, perfect aspect verbs, public verbs, synthetic negation, present participle clauses
Dimension 3 Explicit vs situation-dependent reference	Discourse that identifies the referents explicitly through relativization to discourse that relies in nonspecific deictics and reference to an external situation.	7 features: (+) wh relative clause, pied-piping constructs, phrasal coordination, nominalization, (-) time adverbials, place adverbials, adverbs
Dimension 4 Overt expression of persuasion	Discourse with the overt expression of message senders' own point of view to argumentative discourse designed to persuade the message recipient.	6 features: (+) infinitives, predication modals, suasive verbs, conditional sub., necessity modals, split auxiliarities
Dimension 5 Non-abstract vs abstract information	Discourse with a highly abstract and technical information to discourse with non-abstract focus.	6 features: (+) conjuncts, agentless passives, past participial clause, by-passives, past participial whiz deletion, other adverbial subordination,
Dimension 6 Online information elaboration	Discourse that is informational, but produced in real-time conditions to the disocurse without real time constraints.	4 features: (+) that clause as verb complements, demonstratives, that relative clauses, that clauses as adj. complements

Thus, the *MDA* can reveal the way texts can be characterized as more or less international or transactional from the perspective of each of the above-considered dimensions, spoken and written discourse can be characterized as more or less international or transactional from the perspective of each of the above-considered dimensions. Bibber concludes that texts are multi-faceted, as they can be more similar with respect to one or several dimensions, but considerably different with respect to other dimensions. The *MDA* opens a way to elaborate, functional characteristics of texts as their linguistic characteristics can vary or overlap across spoken and written discourse. Bibber (1995: 36-37) with his extensive study has proved the following: 'There is no linguistic or situational characterization of speech and writing that is true of all spoken and written genres.'

The variation of linguistic features proposed by Bibber can be illustrated by the comparison of three texts: an expository text (*Text 5*), a face-to-face conversation (*Text 6*) and a professional letter (*Text 7*). Although these texts are too small to present the calculation of dimension scores, they can be used to exemplify the way texts can differ according to one of the dimensions, but overlap along the other dimension. These above three texts are compared according to 9 linguistic features clustered in Dimension 1, Dimension 2, and Dimension 3.

Text 7: Professional letter

> We felt that we needed a financial base on which to work, but the goals which we indicated for I. are also included in the goals of L., including of course the occasional papers. In the meantime, we are going ahead with plans to establish three language resource institutes resource centres with a reading library of ESL materials and directors who are competent ESL professionals, (2) as a funner for consultant activities both outward using local expertise needed in other areas where we have L and inward bringing into the area needed expertise and including workshops, mini-conferences, and seminars, and finally (3) to offer educational programmes.
>
> (Biber, 1995: 132)

On **Dimension 1**, *involved versus informational production*, the three texts have been compared with respect to the following linguistic features: the first person pronouns, the second person pronouns, contractions, nouns and word length. As conversation (Text 6) is highly involved, interactive and affective it displays a comparatively high frequency of the first and second person pronouns (*I*, *you*), simplification of grammatical structures (*isn't, don't*,), low nominal style (20 nouns out of 157 words), and comparatively short words (mean word length in characters 3.33). The expository text (Text 5), however, can be characterized according to this dimension as informational without interpersonal and affective content. The author of this text has refrained from the use of the first and second person pronouns, has used only full grammatical forms (*In such studies, it is sometimes helpful to examine some of the articulatory movements …*) and to reach higher informational density, has used 36 nouns (out of 143 words) as well as comparatively long (mean word length in characters 5.64) and specific words (*articulatory*, *acoustic*, *prominence*). As concerns the professional letter, it can be characterized as more affective, but less informational than the expository text. It is written by one person to another who represent a particular institution. That is why the letter contains only personal pronoun *we*. The limited range and number of personal pronouns emphasize the institutional affiliation and signals that apart from interpersonal relations, the informational purpose also takes an important place, which manifests in comparatively high information density and lexical variation that is, however, lower than in the expository text. The professional letter does not display contracted forms, but contains a considerably high proportion of nouns (23 nouns out of 100 words) and diversified, more specific vocabulary (*resource, competent, professionals, and consultant*) than the conversation. The mean length of which in the characters is 4.44. On this dimension, as illustrated in *Figure 1*, conversation is the most involved, the professional letter is less involved, whereas expository prose is the least involved and thus the most informational text.

Alongside **Dimension 2**, *narrative versus /non-narrative concerns*, the examples of the three texts have been compared according to two linguistic features: third person pronouns, and past tense verbs. All three texts exibit a considerable overlap as their overall purpose is non-narrative. Consequently, these texts display less diversified linguistic variation along this dimension. Conversation contains an occasional use of the third person pronouns and the professional letter includes the reference to the third person(s) (s/he) by providing a name of a person(s). The authors of all three texts have refrained from the use of the past tense verbs. The texts predominantly display the present tense forms: expository text contains only present tense forms, professional letter displays present tense forms and two instances of past tenses, whereas conversation predominantly displays present tense forms and only three instances of the future tense forms.

According to **Dimension 3**, *explicit versus situation dependent reference*, the three texts have been compared according to two linguistic features: phrasal coordination and nominalizations. The expository text (Text 5) is referentially the most explicit and informational as it displays frequent nominalizations (*movements, prominence, situation, interaction, addition*) and occasional phrasal coordination (*physical and psysiological interaction*). The professional letter (Text 7), which is more situation dependent and less informational, displays the use of phrasal coordination (*ESL materials and directors*), but the author of the letter has refrained from nominalizations. Conversation (Text 6) is the least integrated and informational, and it does not display the above two linguistic features.

The comparison of the three texts (Texts 5,6,7) can be illustrated by ranging three texts (expository text○, conversation △ and professional letter□) along three exemplified dimensions as shown in Figure 1.

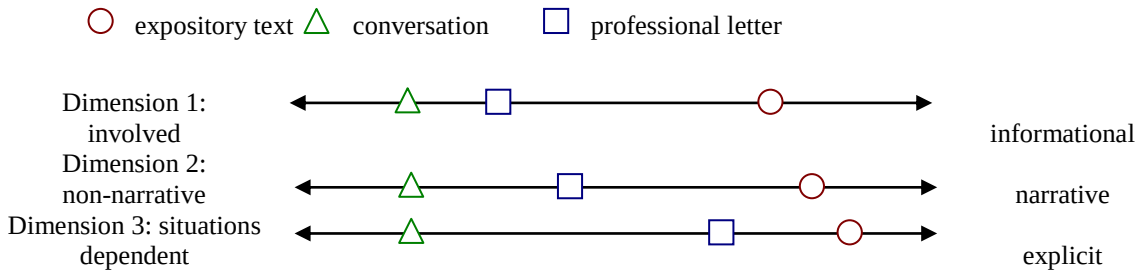


Figure 1: **Ranging of conversation, expository text and profession letter according to linguistic features of three dimensions**

Linguists (McCarthy and Carter; 1994: 9; McCarthy, 2001: 94) see the value of the *MDA* in the identification of the key areas that distinguish written and spoken communication without oversimplifying of their differences as the spoken/written divide

cannot concede the texts to be grouped into particular registers. The *MDA* that aims to bring together contexts with their linguistic expression goes along the lines of the widening the scope of the concept discourse (hence written discourse) by extending it from the textual to social interpretation level of the texts. The widening of the discourse scope as well as the current tendency of mode-mixing favoured by the advent of technology-enhanced written communication strongly diversifies the contexts of written discourse. Hyland (2002b: 53) reminds that the current investigation of language variation requires more delicate, socially inferred and genre-sensitive frameworks that go beyond the stereotyped situations and the clustering of lexico-grammatical features. Moreover, technological changes originate such new discourse forms as chat, email, discussion forums, which violate the traditional division of texts into genres.

The methodology of the MDA, apart from above-considered synchronic discourse investigation, has been applied in contrastive discourse analysis (Biber 1995) and learners discourse analysis (Xiao and McEnery et. al. 2005) as well as in the research of native and non-native students' transactional academic writing in order to identify the over/underuse of particular linguistic features. For example, the comparative investigation of connector usage in English essay writing of native and non-native *EFL* students (Milton, 1993; Granger's and Tyson's, 1996; Bolton et al., 2002). Another trend in the investigation of transactional academic texts is the application of the *MDA* to identify the register-pertinent changes manifesting themselves in the variation of frequency rates of a range of linguistic features in non-native students' transactional writings before and after full-time academic writing courses (Shaw and Liu, 1998). Accordingly, the *MDA*, as Johnstone (2002) notes, is suggestive in the investigation of language variation especially across the texts that represent different genres. This approach, therefore, has recently been introduced in the research of academic discourse in two communication modes: spoken and written (Csomay, 2002; Bibber et.al 2004). However, there are comparatively few studies which have included the MDA of language variation within non-native students' electronic discourse that comprises transactional as well as interactional electronic texts. The studies devoted to students' electronic discourse have focused on the linguistic characteristics of interactional texts (Davis, Thiede, 2000). It is, therefore topical to explore the linguistic characteristics of *NNS* interactional and transactional texts. Thus, on the basis of Biber's MDA findings regarding register variation of written texts, the author of the present study concludes that it is relevant to implement the research methodology of *MDA* in the investigation of the present study, namely, research of *NNS* language variation in the

corpus of their electronic texts. Such extension of the application of *MDA* can lead to understanding of the language variation differences among the texts of *NNS* corpus containing their interactional and transactional texts.

1.2.4 Linguistic Variation in Written Electronic Discourse

Linguistic peculiarities of electronic texts as discourse manifestations have been investigated by linguists since the time when computers and relevant software, first of all, the word processor and then computer networks spread among language users. Human-to-human communication via computers or interactive networking was designed in the United States in the late 1960s to enable the transmission of computer programmes and data among remote computers in the interests of national defence. The application of computer networks spread rapidly as a means of interpersonal communication among computer scientists in the early 1970s, then among a limited number of academic and business users in the 1980s and grew into popular use in the 1990s. Around the same time linguists Ferrara, Brunner and Whittemore (1991), drew attention to the linguistic characteristics of interactive electronic written discourse in their publication *Interactive Written Discourse as an Emergent Register*. The term *interactive written discourse* was gradually replaced by the notion *computer- mediated discourse* that according to Herring (2001) came into use in 1995. Crystal (2001) points out that there are other synonymous notions that denote interactive written discourse, and one of the most widespread terms is *computer-mediated communication* (CMC), defined by Coffin et.al. (2003: 132) as 'communication between individuals and groups using any form of information technology, for example, email, chat, or a computer conferencing system'.

The advent of computer technology during the past sixteen years has resulted in the expansion of writing that takes place on the computer screen and is communicated to other individuals electronically. Moreover, online environments, used in language studies, for example *WebCT, Moodle*, apart from the tools for interactional texts (emails, chats and conferences, discussions) provide also tools for writing, editing and submission of transactional texts, for example, students' essays. Thus, with the production of more subtle, yet at the same time easy-to-use online environments, (for example, as Beatty [2003] notes that more and more current email programmes provide rudimentary spelling checkers), and their increasing penetration in various domains, transactional word processed documents are embedded into online environments by using various techniques. For

example, such transactional texts as students' essays (Biesenbach-Lucas and Weasenforth, 2001) can be written in the text-box of online messaging systems (emails), attached to online messages or stored in the file bank of a particular online environment and, if necessary, made available for readers. Therefore, the author of the present study concludes that students' transactional texts, which have been presented to their audience using online environments, constitute part of students' electronic computer-mediated discourse that as Coffin et.al (2003) have claimed, includes written communication using information technology. The classification of students' texts is presented in chapter *Arrangement of the Study Corpus* containing the author's proposed classification of students' interactional and transactional electronic texts.

Texts of computer-mediated discourse, like conventional texts, use a semiotic system in a form of letters, punctuation, capitalization and spatial organization. However, according to researchers (Herring, 2001; Whittaker, 2003), the texts developed with the help of various messaging systems can be of a wide variety of registers, namely, display functional variation realised through the linguistic features characteristic for written as well as spoken communication.

Current theories on the linguistic characteristics of *CMC,* therefore, aim at the investigation of the influence of various messaging systems on the context-based language use in the texts developed with the help of these messaging systems and the comparison of the linguistic peculiarities of these texts with the functional peculiarities of spoken and written language as presented in chapter *Contrasting of Written vs Spoken Discourse*. The departure point for linguists is time/place matrix suggested by Dix et al (1993). This matrix is based on the distinction of texts according to the dimensions of one situational factor, namely, discourse mode: spatial/interpersonal distance and experiential distance. Spatial/interpersonal distance refers to the distance of visual and aural contact between interactants, whereas experiential distance denotes the distance between the language use and the occurring social process. This matrix is applied to characterize various electronic communication situations and is recognized by researchers, for example, Levy (1997) as a helpful background for a more subtle insight in *CMC* and the messaging systems used in this communication. The author of the present study has summarised the affordances of messaging systems and the examples of these systems (Table 4), which have been considered by linguists.

Table 4: **Summary of the examples of the messaging systems and their affordances**

Synchronity (Dorning, 1998; Crystal, 2001; Coffin et al, 2003; Herring, 2001; Whittaker, 2003)	**Transmission type** (Cherny, 1999; Herring, 2001)	**Number of Participants** (Dorning, 1998; Crystal, 2001)	**Anonymity** (Selfe & Meyer, 1991; Herring, 2001)	**The overall purpose** (Crystal, 2001; Coffin et al, 2003)
Asynchronous Email, email-based systems (listserv, discussion list, usenet newsgroups, bulletin boards, asynchronous discussion in online learning environments	**One-way transmission** Email, email-based systems (listserv, discussion lists, usenet newsgroups, bulein boards, asynchronous discussion in online learning environments, most of Chat, MUDs, MOOSs,teleconferencing messaging systems	**Two participants** Email private chat, ICQ, specific online learning environments	**Anynymous messages** Chat (IRC, Webchat etc.), discussion systems	**Various exchanges between private mailbox holders** Emails, private chat
Synchronous Chat(IRC, webchat, etc.); MUDs and MOOSs, teleconferences in institutional online environments	**Two-way transmission** ISQ (I seek you)	**A groups of participants** Mailing lists, institutional discussion forums, IRC, teleconferencing systems, online learning environments	**Authorised messages** Chat, discussion systems, email, enail-based systems, ISQ	**Continuous topic-related** Discussion and teleconferencing systems, ISQ
-	-	**Multiple Participants** Newsgroups, MUDs, MOOs	-	-

The first medium variable or affordance, *synchronity,* goes along the time axis proposed by Dix et.al (1993) and is recognized as an important variable by all linguists who investigate electronic discourse. According to this variable, Kiesler et. al. (1984); Crystal (2001); Coffin, et.al. (2003); Herring (2001); Whittaker (2003) and other linguists distinguish *synchronous* and *asynchronous CMC*. Hopper (1994) and Whittaker (ibid.) in parallel with synchronity suggest a more subtle characterization of messaging systems by introducing the notion *interactivity* (interactive/non-interactive messaging systems). In asynchronous or non-interactive *CMC* mode users do not need to log on at the same time in order to send and receive messages, as the messages are stored at the addressee's site until they can be read. Therefore, the messages of asynchronous/non-interactive *CMC* are

relatively permanent. The examples of such messaging systems are email, email based systems, asynchronous discussion forums and conferences in institutional online environments. Synchronous or interactive discussion requires that the message sender and the addressee are logged on simultaneously, which means that the messages depending on the peculiarities of a particular synchronous messaging system tend to be more ephemeral than the messages of asynchronous *CMC*. The examples of synchronous messaging systems are *Chat, MUDs, MOOs* and, as distinguished by some researchers (for example, Bierman, 2005), teleconferencing systems that are used in educational setting. Although the researchers point out that synchronous or interactive messaging systems allow providing instant feedback, it has to be noted that the majority of these systems do not enable the users to get an immediate feedback in the same way as in face-to-face conversation. Messages become visible to the addressee only when they have been sent, but during face-to-face communication the message reaches the addressee word-for-word.

In order to emphasize that there are some messaging systems that provide the possibility to see the message senders posting word-for-word as it is typed, Cherny (1999) and Herring (2001), distinguish one more affordance, *transmission type*: *one-way* and *two-way transmission*. *One-way transmission* means that a message is transmitted as a single unit and recipients can view that message only when it has arrived. All asynchronous as well as most of the synchronous *CMC* messaging systems, according to Cherny and Herring, belong to one-way transmission. Two-way transmission allows providing an immediate text-based feedback, namely, during communication participants' screens split into two or more sections and the words of each participant appear keystroke by keystroke in the respective sections as they are typed. An example of such messaging system is *ICQ* 'I seek you'. Consequently, synchronity/interactivity and transmission type of messaging systems emphasize the level of persistence of the message as well as the potential speed of the message exchange and, therefore, correlates with the discourse field.

Another affordance proposed by linguists is the number of participants allowed by a particular messaging system. Dorning (1998) and Crystal (2001) distinguish messaging systems that allow two participants, for example, email and private chat and systems that allow a group of participants, namely, mailing lists, institutional discussion forums, teleconferencing systems, *IRC*. In addition, Dorning (ibid.) proposes a third group, multiple participants, and mentions as an example, newsgroups and discussion forums open to multiple users. This affordance refers to the message sender/recipient relationships and shared knowledge, hence the discourse tenor.

The next affordance that is suggested by Selfe and Meyer (1991); Herring (2001), anonymity, divides the messaging systems into two groups: systems that allow anonymous postings, for example, IRC, webchat, discussion forums that are open to multiple participants, as well as the systems that show the authorship of the posting, for example, email and email related systems, private chat, thematic discussion forums, teleconference systems for a particular group of people. Anonymity relates to message sender/recipient relationships and refers to discourse tenor.

Alongside with the previously considered affordances, Crystal (2001) and Coffin et.al. (2003) mention one more affordance, the overall purpose of a messaging system. According to them, there are systems that envisage exchanges between private mailbox holders, for example, email, private chat, and there are messaging systems that are meant for continuous topic related discussion for a particular group of people. This affordance contributes to the discourse field of text-based electronic communication.

The previously considered affordances or the communication channel specific features of various messaging systems invest in the context shaping of the language use in various CMC modes. For example, it is questionable whether the participants of the discussion forum that is open to multiple anonymous users would post the messages containing overtly hostile content (also labelled as flaming) targeted at a particular addressees if the messaging system would provide such tenor-related restriction as authorised postings.

The above overview on peculiarities of the affordances gives grounds for the author of the present study to draw two conclusions. Firstly, the summative overview of the peculiarities of messaging systems shows that their impact on the communicative context can go beyond the two dimensions comprised by register factors, discourse *mode*, and leaves an impact on interactants' relationship, namely, discourse tenor. Secondly, the overview of the affordances of messaging systems is a significant; however, only initial insight leading to a more sophisticated and detailed analysis of context-based peculiarities of linguistic expression in the texts of electronic discourse.

Moreover, Kern (2006) considers that the three earlier mentioned electronic discourse types range along the continuum from product oriented forms resembling paper-based writing (for example, web sites, hence electronic texts incorporated in web sites, and most emails) to more process-oriented and interactive discourse that shares many features of speech (for example, chat, instant messaging). A detailed analysis of communicative situations of electronic texts (transactional texts on web, emails, synchronous conferences,

asynchronous discussion forums) that lead to the peculiarities of their linguistic expression are further explicated in the present chapter. This analysis is based on the survey of the affordances of the messaging systems and Crystal's (2001: 42) written language criteria that he has applied in his comparison of conventional transactional texts with the transactional texts on the web, emails and chatgroups that include synchronous and asynchronous discussion (Table 5).

1.2.4.1 Texts on the Web

The first electronic discourse that Crystal subjects to comparison with the characteristic features of transactional written language is texts on the Web. According to Crystal (2001), web sites can incorporate several electronic situations one into another by embedding texts of various formats (*word processed, pdf, web pages*) into online environments. Therefore, the word processed essays by *NNS*, which are integrated into e-course *English Academic Writing III* can be included into this group. Crystal (ibid.) concludes that texts on the Web strongly resemble the traditional communicative situations of conventional transactional writing according to all seven written language criteria presented in Table 6. For example, web sites that contain reference publishing, advertising and other texts that belong to a range of genres reflect the same communicative situations as in conventional written communication. Accordingly, most of the genres of written language that can be found on the Web show insignificant differences of lexico-grammatical features apart from those changes that are caused by the adaptation to the electronic medium. Accordingly, as it is seen in Table 6, electronic texts on the web possess extra space-bound, visual and graphical properties. Web-based texts are less permanent than their paper-based counterparts. In the case of repeated reference to a particular text on the Web or in online environment, it may happen that certain changes are encountered, for example, the changes referring to the factual information, visual appearance or graphical properties. In addition, the text may have been replaced by another one or it might have expired. Another specific feature is that a page, once downloaded to a particular user's screen, may have its text cut, added to, revised, annotated, even totally restructured. Moreover, versatile and sophisticated visual and graphical characteristics of web-based texts can substantially differ from the visual and graphical characteristics of paper-based texts. Texts on the web or texts embedded in online environments often provide visual aids to support text, in the form of photographs, maps, diagrams, animations and other visual aids. The visual and graphical potential of

electronic texts on the Web and in online environments is currently investigated by researchers, for example, Kress and Leeuwen (2001). However, the visual aspect of electronic texts posted on the Web or online environments goes beyond the scope of the present study. Since the linguistic characteristics of Web pages and texts embedded into online environments correlate with linguistic characteristics of formal communicative situations in writing according to the dimensions of register factors which are analysed in the chapter *Language Variation in Written Discourse*, the proceeding analysis is devoted to the linguistic characteristics of such electronic discourse modes as emails, synchronous conferences and asynchronous online discussions.

Table 5: **Written language criteria applied to CMC (Crystal, 2001: 42)**

Written language criteria	**Web**	**Email**	**Chat groups (asynchronous/synchronous discussion)**
Space-bound	**yes**, with extra options	**yes**, but routinely deleted	**yes**, but with restrictions
Contrived	**yes**	variable	no, but with some adaption
Visually decontextualized	**yes,** but with considerable adaption	**yes**	**yes**
Elaborately structured	**yes**	variable	no
Factually communicative	**yes**	**yes**	variable
Repeatedly revisable	**yes**	variable	no
Graphically rich	**yes,** but in different ways	no	no

1.2.4.2 Emails

Email is considered to be the most widely used CMC mode not only in personal, but also in institutional communication, especially in academic and business institutions (Baron 2000; Crystal, 2001). As a messaging system, it possesses the following characteristics: it is asynchronous, one-way, predominantly two-participant messaging system of authorized postings the overall purpose of which is various exchanges (opinion-based as well as factual) between private mailbox holders. Email as an electronic communications system, possesses one more peculiarity, fixed discourse structure: headers (e-address, a brief description of the topic of the message, the date and time of the sending of message), greetings/farewells and the body of the message. In this respect email resembles, for example, a research paper that also is supposed to contain particular discursive elements:

title, authorship, abstract, introduction and other elements. The messaging system peculiarities, according to Crystal (2001), tend to bend email away from written language characteristics towards spoken language according to four written language criteria (Table 5): space-bound, contrived, elaborately structured and graphical characteristics of email communication. Firstly, email messages are less space-bound than conventional writings. Although emails can be permanent, the routine textual deletion is encouraged by a particular email management system. This peculiarity interrelates with contrived characteristics of emails and the flexibility in the elaboration of linguistic structures in the messages. The second criterion that also contributes to the contrived characteristics and linguistic elaboration is the flexible communication speed allowed by email messaging systems. Thirdly, email messages can be repeatedly revisable, which is far from possible in conventional writing. Finally, email messages lack graphical richness that can account for more intensive application of, for example, punctuation marks (multiple exclamation or question marks), capitalization or emotions (☺) in less formal messages, whereas lexical complexity in more formal messages.

The communication channel-provoked written/spoken language variation in email discourse show that the possible linguistic characteristics of this CMC mode are far from straightforward (Davis & Brewer, 1997; Baron, 1998; Murray, 1995). A number of researchers (Murray, 1995; Biesenbach-Lucas & Weasenforth, 2001) have contrasted and investigated two text-types, students' word-processed essays and emails in educational setting. Biesenbach-Lucas & Weasenforth (2001) have investigated the linguistic characteristics of essays that a student group developed as word-processed documents, while the other group developed them as email messages. The given investigation shows that variation of cohesive devices in word-processed and email-based essays are statistically insignificant and thus point to the role of all three contextual factors (*field, tenor* and *mode*) in the choice of lexico-grammatical features in email communication. On the other hand, the same study shows that email-based essays differ from the word-processed essays with the respect to the linguistic features that express the overtness of writers' position. Namely, students' email-based essays contain a larger number of lexical items that express writers' stance more overtly than in the word-processed essays. For example, in email-based essays the use of personal pronoun *I* and such pronoun-based phrases as *I think, I agree, in my opinion* is more frequent. These results, however, imply that the students have contextualized the communicative situation in essay writing differently when developing it as an email message. The findings point to the role of

messaging systems as well as contextual factors (field, tenor and mode) in the choice of lexico-grammatical features in email communication.

These studies support Crystal's conclusion that the perception of email as a typically informal communication mode that stands close to spoken discourse is too straightforward. Moreover, Davis and Brewer (1997) point out that email is a relatively new medium which continues to change; therefore, it is too early to identify it as a genre. The email medium is maturing and the register variation tend to cut across a range of particular instances of email communication. For example, considerable register variation are seen in two email messages presented in *Text 8* and *Text 9*. *Text 8* contains email that resembles a personal letter to a person the author of the message obviously knows very well and shares a particular knowledge about the event. *Text 9* shows an email that contains the lexico-grammatical features that are predominantly used in professional letters (see the example of the comparative analysis of a personal letter and a professional letter in the chapter *Language Variation in Written Discourse*). Consequently, both texts (*Text 8* and *Text 9)* are email messages that have been transmitted through the same or very similar messaging systems, but the linguistic characteristics of these two texts considerably differ.

Text 8

> Dear colleagues,
>
> At the same meeting the decision was also taken to found a Nordic association of language technology, which would also be responsible for organizing the NoDaLiDa conference. However, although the process of founding the association is well under way, with drafts of bylaws currently being circulated, it is unlikely that it will be completed before the fall. Since preparations for the next NoDaLiDa conference need to be initiated before that, I instead turn directly to the future members of the association, through the NoDaLiDa mailing list, to seek guidance in organizing the technical program for our next conference.
>
> Best regards, A.

Text 9

> Hi again!
>
> I've just had a quick note from Knut, who says it will be easiest to pick out your students, since they all put "other" for the language -- nice! We'll give you new numbers and adjust what's in the database – I'll talk to him again tomorrow, but it sounded like nobody would have to retake anything, which is a relief.
>
> Cheers, B.

Email messages can display even subtler language variation in-between and beyond the two exemplified situations. Chen (2006) draws attention to the difficulties non-native students of English might face in the development of email messages in institutional

contexts; as such emails require a subtler analysis of the communicative situation than the emails to close friends. Consequently, email writers are supposed to make serious decisions concerning their choice of lexico-grammatical features depending on the peculiarities of the communicative situation. Linguists direct the attention to the comparatively wide register variation in email communication by providing the guidelines for students to draw their attention to field and tenor in email communication. Swales (2001) emphasizes that there are situations in which informality is irrelevant. For example, when a student sends a request to a faculty member or an administrator, informal language may be perceived as too direct, hence insufficiently polite. A similar conclusion is drawn by Bloch (2002) on the basis of his investigation how non-native students have used email initiatively in order to interact with their instructor. He concludes that writing a successful email requires more than simple fluency in the target language. It requires the ability to express oneself using a variety of linguistic forms most importantly, to know when it is situationally appropriate to use particular linguistic forms.

Therefore, Crystal advises the message senders to be cautious about, for example, the following email writing guidelines:

> A well-written electronic missive gets to the point quickly, with evocative words, short graphs, and plenty of white space. Spelling and punctuation are loose and playful.
>
> Hale and Scalon (1999: 3)

Crystal (2001) notes that emails in educational and business settings are routinely seen as professionalism (one that can speed up decision-making and build strong daily working realtionships) rather than just an opportunity to chat. Such approach to email communication is supported in more recent email writing manuals. For example, to draw attention to the possible register variation in emails, Trask (2005: 5) writes in his manual that 'some business firms instruct their employees that business emails should be prepared with the same care as business letters'. Emmerson (2004) in his book *Email English* has devoted the first section to the choice of the vocabulary/phrases according to two formality levels: informal and neutral/formal. He points out that the neutral formality level is the most common in professional/work emails, in which, 'the writer and reader are both busy, so the language is simple, clear and direct, the language is more personal than presumably in transactional writing, however, it is not similar to speech.'

Over the past years, email messaging systems have been maturing and their applicability has been expanding. Accordingly, the perception of the notion *emailing* as

predominantly informal electronic communication is misleading in the context of the current and presumably the future email communication needs in various domains. Baron (2000) even considers that two styles of email communication (more formal and less formal) might emerge, one edited and the other unedited. Crystal outlines the future of email communication along similar lines:

> …The result will be a medium which will portray a wide range of stylistic expressiveness, from formal to informal, […] and where the pressure on users will be to display stylistic consistency, in the same way that this is required in other forms of writing. Email, then will take its place in the school curriculum, not as a medium to be feared for its linguistic irresponsibility, but as one which offers a further domain within which children can develop their ability to consolidate their stylistic intuitions and make responsible linguistic choices.
>
> Crystal (2001: 128)

Accordingly, the context-pertinent linguistic variation are among the current needs of NNSE.

1.2.4.3 Synchronous and Asynchronous Online Discussion

Synchronous and asynchronous discussion, which Crystal (2001) in his comparative analysis labels – chat groups, according to the affordances of the messaging systems are synchronous chat and/or conferences as well as asynchronous discussion forums, are one-way, online communication among people with the overall purpose to exchange opinions or argumentation. The communicative situations of these CMC modes show a considerable deviation from the formal writing situations, namely, transactional texts, and strongly bend synchronous and asynchronous discussion towards spoken communication with respect to six out of the seven written language criteria put forward by Crystal (Table 5): space-bound, contrived, elaborately structured, factually communicative, repeatedly revisable and graphical characteristics of this *CMC* mode. Firstly, online discussions are less space-bound than conventional writings. Erikson (1999) emphasizes that messages are comparatively persistent: they stay on the screen for a period of time (before the arrival of other messages that replace it or makes it scroll out of sight), which depends on the system that is used. Although there are systems providing an archiving log for all messages; however, in practice there are few systems that correlate with contrived characteristics of emails. The second criterion that also contributes to the linguistic characteristics is the high communication speed in synchronous discussion and the flexibility of the communication speed in the case of asynchronous discussion that according to Sotillo (2000) causes a greater syntactic complexity of asynchronous than the synchronous discussion. Thirdly,

due to the specific features of the communication speed and the overall purpose of discussions, the discussion postings tend to be considerably less elaborately structured than transactional writing and more formal email messages. Fourthly, discussion postings tend to be far less factually communicative as their overall purpose is the opinion exchange. However, Crystal (2001) points out that more academic and/or professional asynchronous discussions tend to be more factual in their overall purpose, but more social and lucid chat groups tend to contain the sequences with negligible factual content. Finally, discussion postings are not repeatedly revisable and lack graphical richness that in the same way as in informal email communication can account for a more intensive application of, for example, punctuation marks (multiple exclamation or question marks), capitalization or emotions (☺) in less formal messages, but lexical elaboration in more formal messages.

As it can be seen from the comparison of the chat (synchronous and asynchronous discussion) with formal writing criteria they, like email communication, also represent a hybrid communication medium tends to lean more towards the linguistic characteristics of spoken language than email communication (Warschauer 1996; Smith 2005; Coffin et. al 2003; Fitze, 2006). That is why a number of studies are devoted to the investigation of the similarities and differences between synchronous and asynchronous discussion or the differences of these two online communication modes with the face-to-face communication. Collot & Belmore (1996), for example, have explored bulletin-board messages using Biber's *MDA* and found that these messages have some linguistic characteristics of written genres and other characteristics of oral genres. The messages that they have analysed are, consequently, more like Biber's public interviews as well as personal and professional letters. According to Fitze's (2006) findings of the analysis of text-based electronic conferences and face-to-face communication of the students, text-based electronic conference messages contain a wider range of lexical variation and higher type/token rates than the transcripts of face-to-face communication. Smith (2005), Yates (1996) on the bases of their studies also conclude that online discussion has been found to exhibit particular characteristics that resemble spoken communication, but other characteristics that are similar to written interaction. For example, lexical density is similar to transactional written texts, whereas the frequency of first person pronouns is close to spoken language. The researchers doubt whether such linguistic features of online discussion messages are only due to the peculiarities of this electronic communication mode. It has to be emphasized that paper-based personal letters or notes exchanged by close friends contain the same language use peculiarities that are found in the online

discussion postings. Moran & Hawisher (1998) claim that *CMC* (synchronous and asynchronous communication) includes a wide variety of more or less formal and informal genres in the same way as other forms of writing (for example, the genres ranging from a shopping list to a formal essay).

The study of Davis & Thiede (2000) convincingly shows that lexical complexity in asynchronous discussion forums goes beyond the peculiarities of the communication mode and depends on the discussion topic (field) and how far writers 'accommodate to each other in online medium' (tenor). Their investigation of lexical complexity demonstrates that students choose lexis depending on their social status and discourse conventions: belonging to the institutional communication has brought in the changes in the lexical variation in discussion postings.

Coffin et al. (2003) conclude that an important feature of writing as part of computer conferencing is that it is not a homogenous entity. As in face-to-face contexts, writing used in electronic conferencing varies according to the purpose of the writing, the subject matter and the addressee/addresser relationship. For example, chats in virtual cafes, coffee bars and other informal sub-conferences, where the relationship between peers is generally equal, are more likely to simulate face-to-face conversation than thematic, course related online conferences. Contextual heterogeneity that determines linguistic characteristics of asynchronous online discussions is seen in two examples (*Text 10* and *Text 11*) of online discussion postings.

Text 10: Thematic, course related asynchronous discussion among a particular group of students

> I'm afraid that I can hardly imagine these two essays combined. The approaches of the authors differ too much to make one out of the both essays. But it doesn't mean I disagree. If one could combine the best features of the first author's approach with the best features of the second author's approach that would make a golden alloy for the future pieces of writing. So these are the approaches we need to revise and retain the best we have had.

Text 11: Asynchronous discussion forum open to all students

> oooh, yeah.....exams are coming quite in a fast run... i wonderrrr.....but we will see... however - we all will going to survive that storm, so.. don't you worry! *thumbs up* =))

Text 10 contains the posting that belongs to a thematic discussion within a particular student group in an academic setting. This posting is less formal than the email text included in *Text 8*. The discussion posting in *Text 10* is more interactive as its overall

purpose is to exchange opinions and arguments about a particular topic. The interactive character of the posting is emphasized by the use of the first person pronouns, contractions and the overtly expressed stance of the writer ('I'm afraid that I can hardly imagine, ...but it doesn't mean I disagree'). On the other hand, this posting contains thematically pertinent vocabulary (the keywords are *essay, writing*) and syntactically elaborate sentences. In contrast, the second posting (*Text 11*), which is a part of anonymous, asynchronous institutional discussion forum that is open to all students, strongly differs from the text included in *Text 8*. *Text 9* contains predominantly elliptical sentences and correlates with the language use and sentence patterns in informal face-to-face conversation presented in *Text 6* (see 1.2.2 *Language Variation in Written Discourse*).

Coffin et.al. (2003) rightly notes that electronic conferencing requires the use the written, rather than spoken language as the chief medium of communication. Although this form of written exchange may differ from more formal styles of academic writing (for example, academic essays), it is, nevertheless, an important vehicle for learning to express abstract ideas and exchanging views on disciplinary knowledge. Therefore, it is useful to consider the nature and role of such exchanges and to explore their relationship with more traditional, individually produced, academic text types such as essays.

The comparative analysis of emails, synchronous and asynchronous online discussions with written language criteria reveals that linguistic peculiarities of the considered *CMC* modes go beyond the contrasting of written versus spoken discourse. Neither emails nor online synchronous and asynchronous discussions can be characterized as predominantly possessing only those lexico-grammatical features that are expected in idealized situations of written or spoken communication. Instead, these *CMC* modes display a wide range of language variation that include the elements of both written and spoken language criteria, which according to Crystal (2001) show that text-based *CMC* tends to be more like written language that has been 'pulled some way in the direction of speech [but not as] spoken language which has been written down'. Register variation, therefore, cut across these two *CMC* modes and the genres that they can potentially incorporate. Davis and Brewer (1997: 19) claim that: 'Writing in the electronic medium, people adopt combinations of oral and written discourse to their own, individual communicative needs'. Accordingly, the text-based electronic communication mode through which a particular writing activity (field) is undertaken via interpreting particular message sender/recipient relations (tenor) are interrelated and overlap. These conclusions are reinforced by Carter et al (2001) who emphasizes that, for example, chatrooms can be

used not only for playful interactions, but also for online business meetings, scientific conferences and forums.

The above-considered findings regarding the linguistic characteristics of text-based electronic discourse correlate with Coffin's et.al. (2003) conclusion that it is important to continue the investigation of language variation in non-native students' discourse. Matsuda et al (2003) in their survey *Changing Currents in Second Language Writing Research* also draw attention to the topicality of more detailed and in-depth research in the investigation of particular non-native students' lexical and syntactic features of writing in a range of different genres as well as professional and institutional contexts. Biber et al (2002) conclude that past research of students' classroom academic discourse mostly focused on the linguistic characteristics of academic prose and academic lectures with the majority of the research on the technical research article (in science or medicine). However, it can be concluded that non-native students' electronic discourse has been insufficiently investigated.

Biber's *MDA* is acknowledged and widely used by linguists as a research method in the investigation of register variation in a range of spoken and written genres. Thus, Johnstone (2002) emphasizes that the *MDA* can be applied in the research of language variation across the texts that belong to different communication modes. The *MDA* can help detect the context embedded language variation that range texts along register dimensions. This approach has recently been introduced in the contrastive research of the academic discourse in two communication modes: spoken and written (Csomay, 2002; Biber et.al 2004). However, there has not been much research applying *MDA* in the investigation of *NNSE* written discourse ranging from interactional to transactional electronic texts. Moreover, such studies of register variation are in their initial stage in Latvia. Consequently, it is topical to explore the context-driven language variation in non-native students' electronic academic discourse that embraces academic texts (statements and essays), e-letters and asynchronous/synchronous messages that are incorporated in the e-environment of the project *IDEELS* and e-course *English Academic Writing III.* These e-environments belong to the online venues, which, as Crystal (2001) puts it, can combine all elements 'or where one situation is used within another.' For example, an e-course contains email and online discussion facilities that allow incorporating of word-processed documents in attachments as well as storing or posting them in e-environment and adding a thematic discussion for online analysis of these essays. Accordingly, such multi-functional Internet environments can be used for involvement of students in writing activities that

range from interactional to transactional text-based electronic written communication followed by lingua-functional analysis of this communication discourse.

The author of the present study, therefore, considers that the application of Biber's *MDA* can reveal the variation of linguistic features in *NNSE* electronic discourse that comprises transactional and interactional electronic texts.

The present study aims at a two level linguistic investigation based on the MDA, revealing, firstly, the variation of lexico-grammatical features across students' transactional and interactional electronic texts and, secondly, the variation in students' transactional electronic writings before and after full-time technology enhanced course of written communication. The preliminary research (Vinčela, 2006, 2007, 2008) allows to hypothesize that the frequencies of lexico-grammatical features across students' transactional and interactional electronic writing discourse tend to vary according to the situational factors alongside with the students' involvement in electronic writing activities requiring register variation by applying e-courses that are integrated in BA/MA programmes.

The next chapter *Models of written discourse development* outlines the current models of the development of students' written discourse in academic setting and proposes a model for the development of text-based electronic academic writing discourse in online language studies environment.

SUMMARY

Registers are configurations of lexico-grammatical features within texts according to the situational factors: *field, tenor* and *mode*. *Genres*, which are defined in applied linguistics as linguistic realizations of a particular communicative event, can manifest themselves only through the functionally adequate choice of lexico-grammatical features. The term *register*, therefore, is adopted and applied in the present research aiming at the functional investigation of students' electronic discourse that comprises transactional and interactional electronic texts developed in online environments.

Linguists explore language variation in written discourse from two perspectives: from the perspective of contrasting written and spoken communication and from the perspective of the MDA (multi-dimensional analysis) of spoken and written discourse. Contrasting emphasizes the unique features of writing that are identified by the juxtaposition of the idealized situations of these two communication modes: face-to-face

conversation and informative writing. The MDA seeks to bring out the linguistic characteristics of the multi-functional written discourse by the analysis of language variation in texts along the continuums of functional dimensions. The *MDA* has proved that linguistic characteristics of written discourse, which currently tends to diversify due to the emergence of various electronic communication situations, is situationally versatile and multi-functional. It comprises the linguistic features of spoken language as well as written language that are detached and informative. Hence it is relevant to apply the *MDA* in the present study to investigate the linguistic variation in *NNSE* electronic written discourse.

Written electronic discourse comprises transactional and interactional electronic texts and it is defined by linguists as a hybrid communication medium which envisages a wide range of language variation along the continuums of register dimensions. The register-pertinent language variation in transactional and interactional electronic discourse goes beyond the affordances of a particular messaging system and is determined by the virtual situational contexts. In order to detect context embedded language variation in interactional and transactional electronic discourse types, academic texts (statements and essays), e-letters and synchronous/asynchronous discussion messages written by *NNSE*, it is relevant to apply *MDA*.

1.3 Models of Written Discourse Development and Role of Information Technology in their Application

The analysis of models of text-based discourse development is undertaken in the present study on the basis of conclusions drawn in the first two chapters: *Text and Discourse,* which emphasizes the interrelationship of these notions, and *Linguistic Variation Discourse,* which brings out the register pertinent language variation across the texts thus revealing the linguistic peculiarities of electronic discourse. According to the theoretical considerations represented in these chapters, a text is part of a discourse and its belonging to discourse is established through writers' use of register pertinent lexico-grammatical features. Such perception of *text* and *discourse* has heralded the emergence of the theories underlying the models of text-based discourse development in the *S/FLW* (*second/foreign language writing).* The purpose of these models is to bring out the knowledge and skills that writers require in order to develop a text that meets writer and reader needs.

Bjork et.al (2003) summarize the diversity of the departure points of the *S/FLW* in tertiary education over the last thirty years and conclude that they reflect the process of changing from normative instruction and writing process towards the constructivist perspective to writing. It is a move from teacher-centred *Presentation, Practice, Production* models towards student-centred *Hypothesizing, Exploration, Construction* models. Such shift, which marks the broadening of the paradigm of *S/FLW*, is revealed by the researchers in their diachronic distinction of the overlapping stages of text-based discourse development. Raimes (1983: 237-260) and Tribble (1996: 37) suggest the following stages and terminology: focus on form, focus on the writer (or writing process approach), and focus on the reader (genre approach). Hyland (2002b) proposes text-oriented, writer-oriented and reader-oriented approaches to research and teaching of writing. Coffin et.al (2003) put forward a similar grouping: text approaches, process approaches and writing as a social practice. According to Bjork et al. (2003: 10), the diverse *S/FLW* perspectives can be arranged in the models that focus on the text, the writer, and the discourse community. Although the above groupings of written discourse development models are labelled differently by various researchers, their underlying theories strongly overlap and reveal the gradual paradigm shift in *S/FLW*. Moreover, it is important to note that these approaches co-exist and overlap within the recent models of *S/FLW,* as, in practice; there are no clear boundaries and dividing lines. Accordingly, the

origin and elaboration of *S/FLW* models proposed by researchers tend to go along the lines of linguistic theories of discourse contextualization levels put forward by Bhatia (2004): discourse as text, genre and social practice (see chapter 1.1 *Text and Discourse*). Therefore, it is relevant to consider the perspectives and theoretical frameworks underlying discourse development models, thus revealing how they interact within the current *S/FLW* instruction and draw on them in the shaping of the strategy for *NNS* electronic discourse development in academic setting that integrates virtual written communication. The text-based discourse development approaches are considered in the following sequence: text-oriented, writer-oriented and reader-oriented models.

1.3.1 Text-oriented Models

Text-oriented models comprise, according to Hyland (2002b) two broad perspectives: texts as autonomous objects and texts as genres. Both perspectives focus on the product – text; however, each of them substantially differ in the perception of *text* and *discourse* interrelationship.

The departure point for the perception of texts as autonomous units is the linguistic theories that dominated in the 1960s and the early 1970s. These theories are recognized by Bhatia (2004) as the textualization of lexico-grammar and based on the narrowing of the term *discourse* to the notion of *text.* Accordingly, linguists in the 1960s and 1970s were influenced by frameworks of formal linguistics and tended to focus on the surface-level statistical significance analysis of lexico-grammatical features, refraining from the investigation of their functions in discourse.

This approach, according to Tribble (1996) and Zamel (1982), has been put into practice by the implementation of the linear product-centred model (*Figure 2*). This model sequences three stages: presentation of authorative texts to students, writing practice by imitation or adoption of authorative texts and then the production of a text by the students themselves. The first stage envisages a teacher-led presentation of a text that is considered to be ‘a model of good writing’. A teacher analyzes and explains grammatical features of the model text that deal with the textual context. The second stage is devoted to the imitation of the examples of ‘good writing’, and the third stage envisages a teacher-conducted independent production of a text. Accordingly, this approach proposes models that are based on language system knowledge and the accuracy of the text. These models,

from the perspective of register, envisage writer's knowledge of discourse field, namely, the knowledge of content requiring specific lexis used in a particular domain.

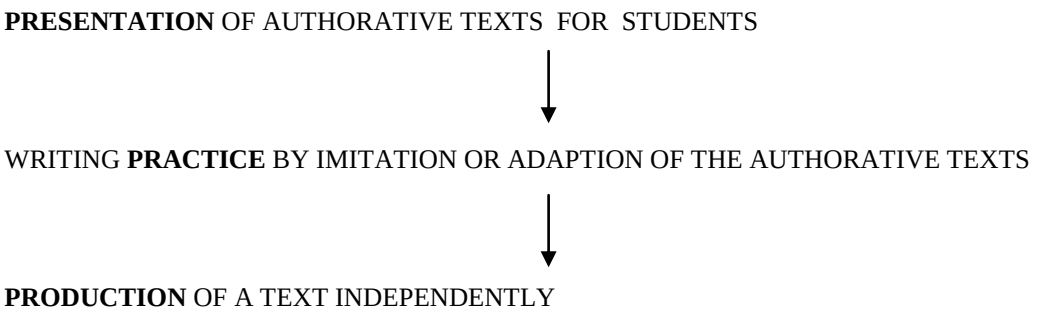


Figure 2: **Text-oriented, linear model of writing instruction**

Tribble (1996) asserts that this approach has manifested itself in the widespread application of guided composition as the main teaching method in writing, thus considering writing as an extension of grammar aiming at the development of a general understanding of language. Zammel (1982, 1985, 1987), who has extensively advocated the need of the paradigm shift towards more student-centred writing, reveals similar limitations of decontextualized text-oriented *S/FLW* instruction:

> It is important to note that students in these classrooms had been accustomed to modes of instruction that reinforced narrow and limited notions about the function of writing. For example, they had drilled the rules and formulaic principles and had little awareness of audience and reader perception. They assumed that writing was done in response to tests or homework assignments that were evaluated by the teacher.
> (Zammel. 1987: 708)

These conclusions correlate with Hyland's (2002b) assumption that a grammatically correct text is not necessarily the main indicator of successful writing, because, apart from accuracy of the language forms, it is essential to vary these forms in accordance with the relevant communicative context. The importance of moving beyond the linguistic accuracy is exemplified by Tribble (1996) in his book *Writing*, in which he presents two examples of the letters of complaint (*Letter of Complaint 1* and *Letter of Complaint 2*) written by intermediate level students of English. He explains that both examples contain language use mistakes; nevertheless, *Letter of Complaint 1* has been recognized as more acceptable by most readers. In the first sentence of this letter, the author has attempted to establish writer/reader relationships and only then introduces the

negative adjective *disappointed,* characterizing the idea of the complaint. However, the text included in *Letter of Complaint 2* starts directly with the complaint containing the negative adjective *unhappy,* which is reinforced by the adverb *very*.

Letter of Complaint 1 (Tribble, 1996)

After waiting for two weeks for a reply about the letter of complaint I send to you. I thought it was ncessary for me to write you again in order to let you know how **disappointed** I am.

Letter of Complaint 2 (Tribble, 1996)

I am **very unhappy** with the accomodation you have arranged for me. I have already argued in person but in vain; here the bathroom is dirty and the shower doesn't work and more over a security system is inexistent.

These examples strive to demonstrate the limitations of the perception of texts as autonomous units and show the topicality of broadening *S/FLW* paradigm towards functional and more student-centred writing instruction, namely, the implementation of the genre approach to *S/FLW*.

Some linguists, for example, Hyland (2002b), Coffin et.al (2003), Bjork (2003) consider that the genre perspective belongs to text-oriented approaches, as genre approach views text from the angle of product. However, other linguists, for example, Tribble (1996, 2002), Bhatia (2004), distinguish it as a separate perspective dealing with the genre space of discourse. In the present study the genre approach is viewed in relation to text-oriented approaches that explore linguistic characteristics of discourse from the point of a product.

The genre approach, as proposed by Tribble (1996, 2002), Bhatia (2004) in *S/FLW* instruction, draws on genre theories in applied linguistics that have been outlined in the chapter *Linguistic manifestation of register*. According to these theories, genre approach is based on the perception of discourse as genres: the functional language use in conventionalized communicative settings, namely rhetorical contexts that are identified on the basis of shared communicative purposes. Accordingly, researchers promoting and exploring genre approach in *S/FLW* have developed models explaining the structural moves that according to Swales (1990) incorporate the content schemata (organization of facts and concepts) and formal schemata (organization of rhetorical elements) of texts representing various academic and professional genres. Hyland (2000) has developed the move structure for abstracts, Bhatia (1993) for sales letters, whereas Swales (1999) has created his distinguished *CARS* (Create a Research Space) model for the rhetorical structure of introductions of research articles. Moreover, Bhatia (1997a) has worked out a

four-stage model embracing the knowledge required in order to acquire expertise in development of texts that explicitly display their belonging to a particular genre:

> 1 Knowledge of the code – of lexico-grammatical, semantic and discoursal resources of the genre.
> 2 Acquisition of genre knowledge – of the communicative goal-oriented purposes of the genre.
> 3 Sensitivity to cognitive structures – awareness of genre-specific values of features in specialized contexts.
> 4 Exploitation of generic knowledge – ability to manipulate conventions for pragmatic effects.
>
> (Bhatia 1997a: 136-138)

These models present the rhetorical structures that realize communicative purposes as well as show how these structures correlate with the knowledge writers require (language system, content knowledge and knowledge of their readership) in order to develop a text that represents a particular genre. However, *ESP/EAP* researchers (Bhatia, 1999; Hyland, 2002b, 2004), who also draw on the two other schools of genre in applied linguistics (systemic functional and new rhetoric), express the concern of too conventionalized approach to genre development in *S/FLW*. Bhatia (2004), Swales (2004), have acknowledged the importance of the adoption of a more open view on genres; otherwise, they, as Hyland (2002b) puts it, are perceived as monofunctional units that are 'uncritically regarded as naturally superior forms of communication'. Overemphasis on conventions contradicts the current text-based communication reality: diversification of communicative contexts due to the advent of technology-enhanced communication modes.

Accordingly, alongside with the genre approach models outlined earlier, it is essential to explore register-pertinent clusters of lexico-grammatical features that distinguish texts belonging to a particular genre. This examination has been emphasized by the model rooted in the systemic-functional approach to genre teaching. The said model sequences three stages (*Figure 3*), which bring out student-centred linguistic and functional exploration and subsequent construction of the texts belonging to a particular genre.

Figure 3: **A systemic-functional approach to genre teaching (Hyland, 2002b)**

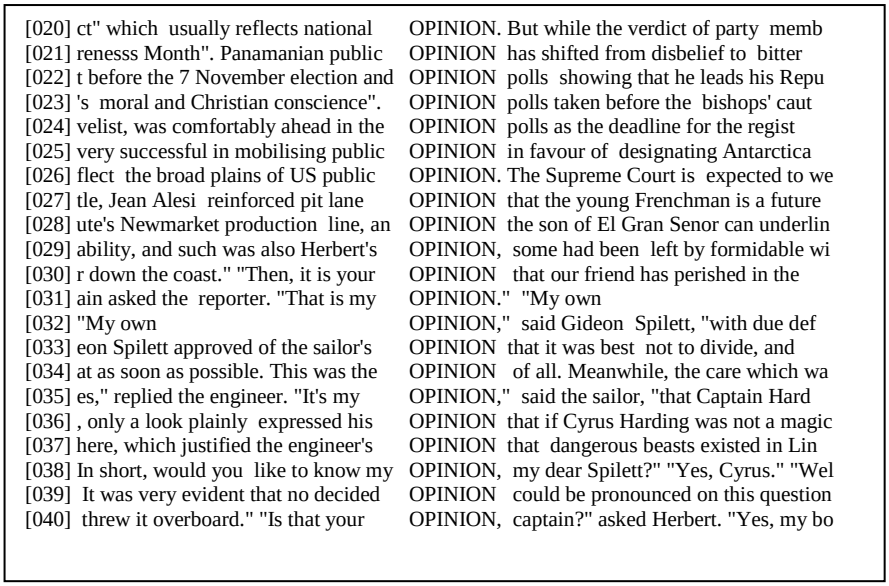
The *Modelling* stage offers a teacher-led presentation of the genre, namely, numerous texts that belong to this particular genre, as well as the stages, purposes and lexico-grammatical features that realize these purposes. The description of the requirements of different text types highlights to students how such 'universal' texts as, for example, the essay can vary in the purpose or function that determines the choice of the language resources in these texts. The next stage, *Negotiation*, focuses on exploration, namely, students' discussion and composition that is based on student-teacher and student-student collaboration through questions which might help shape the text and provide a scaffold for composition stages. The final stage, *Construction*, envisages students' construction of a text belonging to a particular genre by implementation of the notes and summaries they have made during the *Exploration* stage as well as developing and improving several drafts. This model confirms the tendency of the merging of *S/FLW* approaches, in this particular case the genre and writing process approaches, *as* the composition during the *Negotiation* stage envisages drafting and redrafting.

Moreover, the above presented model can be substantially enhanced by the implementation of IT. Thus, Flowerdew (2000) has proposed and Tribble (1996, 1998, 2002) elaborated the application of concordance, namely, corpus data, in the students' exploration of the functional characteristics of the lexocigrammatical features of genres. This approach is recognized as data-driven learning, defined by Johns and King (1991) as 'the use in the classroom of computer-generated concordances to get students to explore regularities of patterning in the target language, and the development of activities and exercises based on concordance output'. Concordance is based on the use of data from a corpus or corpora, which is/are defined by Meyer (2002: XII), 'as any collection of texts (or partial texts) used for purposes of general linguistic analysis'.

One of the concordance activities that can support writing process in the previously discussed model is the exploration of the concordances of particular words within genres or genre groups. It has to be noted that the choice of corpus and processing tools depend on writers' or institutional access to technology, corpora and tools.

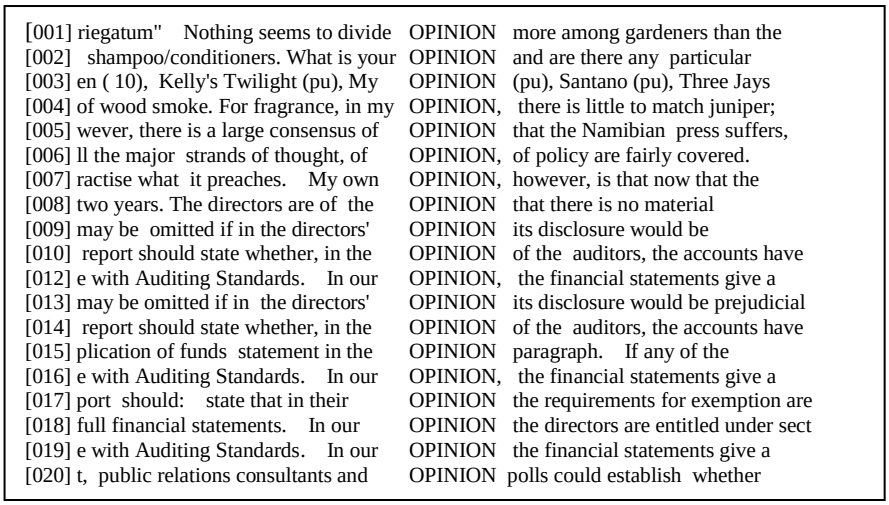
For example, the collocations of the word *opinion* can be comparatively explored within *AWL* (academic writing) mini corpus and *BNC* (British National Corpus) of written texts (*Figures 4* and *5*) with the help of *Compleat Lexical Tutor* (Cobb, 1998; Diniz, 2005). The output of *BNC* written texts (*Figure 4*) shows that the word *opinion* collocates with such personal pronouns as *my* and *your*, whereas the output of *AWL* corpus (*Figure 5*) shows that the word *opinion* collocates with these two pronouns only within the quotations

and direct speech that are incorporated in the texts of *AWL*, as this corpus contains the texts of specialized academic writing genres. These results imply that direct expression of writers' stance is far from a characteristic feature of academic genres or according to Biber's terminology 'academic prose', which by the application of data-driven functional investigation of lexico-grammatical features can be explored by students themselves.



[020]	ct" which usually reflects national	OPINION. But while the verdict of party memb
[021]	renesss Month". Panamanian public	OPINION has shifted from disbelief to bitter
[022]	t before the 7 November election and	OPINION polls showing that he leads his Repu
[023]	's moral and Christian conscience".	OPINION polls taken before the bishops' caut
[024]	velist, was comfortably ahead in the	OPINION polls as the deadline for the regist
[025]	very successful in mobilising public	OPINION in favour of designating Antarctica
[026]	flect the broad plains of US public	OPINION. The Supreme Court is expected to we
[027]	tle, Jean Alesi reinforced pit lane	OPINION that the young Frenchman is a future
[028]	ute's Newmarket production line, an	OPINION the son of El Gran Senor can underlin
[029]	ability, and such was also Herbert's	OPINION, some had been left by formidable wi
[030]	r down the coast." "Then, it is your	OPINION that our friend has perished in the
[031]	ain asked the reporter. "That is my	OPINION." "My own
[032]	"My own	OPINION," said Gideon Spilett, "with due def
[033]	eon Spilett approved of the sailor's	OPINION that it was best not to divide, and
[034]	at as soon as possible. This was the	OPINION of all. Meanwhile, the care which wa
[035]	es," replied the engineer. "It's my	OPINION," said the sailor, "that Captain Hard
[036]	, only a look plainly expressed his	OPINION that if Cyrus Harding was not a magic
[037]	here, which justified the engineer's	OPINION that dangerous beasts existed in Lin
[038]	In short, would you like to know my	OPINION, my dear Spilett?" "Yes, Cyrus." "Wel
[039]	It was very evident that no decided	OPINION could be pronounced on this question
[040]	threw it overboard." "Is that your	OPINION, captain?" asked Herbert. "Yes, my bo

Figure 4: **Concordance of the word *opinion* from *AWL* mini corpus retrieved by *Compleat Lexical Tutor***



[001]	riegatum" Nothing seems to divide	OPINION more among gardeners than the
[002]	shampoo/conditioners. What is your	OPINION and are there any particular
[003]	en (10), Kelly's Twilight (pu), My	OPINION (pu), Santano (pu), Three Jays
[004]	of wood smoke. For fragrance, in my	OPINION, there is little to match juniper;
[005]	wever, there is a large consensus of	OPINION that the Namibian press suffers,
[006]	ll the major strands of thought, of	OPINION, of policy are fairly covered.
[007]	ractise what it preaches. My own	OPINION, however, is that now that the
[008]	two years. The directors are of the	OPINION that there is no material
[009]	may be omitted if in the directors'	OPINION its disclosure would be
[010]	report should state whether, in the	OPINION of the auditors, the accounts have
[012]	e with Auditing Standards. In our	OPINION, the financial statements give a
[013]	may be omitted if in the directors'	OPINION its disclosure would be prejudicial
[014]	report should state whether, in the	OPINION of the auditors, the accounts have
[015]	plication of funds statement in the	OPINION paragraph. If any of the
[016]	e with Auditing Standards. In our	OPINION, the financial statements give a
[017]	port should: state that in their	OPINION the requirements for exemption are
[018]	full financial statements. In our	OPINION the directors are entitled under sect
[019]	e with Auditing Standards. In our	OPINION the financial statements give a
[020]	t, public relations consultants and	OPINION polls could establish whether

Figure 5: **Concordance of the word *opinion* from *BNC* written corpus retrieved by *Compleat Lexical Tutor***

This technology-enhanced genre approach to discourse development promotes the idea of function and presents to students the recurrent lexico-grammatical features of the

texts that are grouped into genres. Moreover, Bhatia (2004), whose multi-perspective model of discourse analysis promotes the bottom-up approach, considers that it is profitable in *S/FLW* instruction to deal with the exploration of the textual space and then move to genre and social space of discourse. The application of the previously considered technology in writing process models requires a special knowledge – operation of queries for particular data extraction from corpora.

The paradigm shift towards student-centred writing is reinforced by writer-oriented models of text-based discourse development. These models, like genre perspective, view texts as a part of discourse; however, from the angle of the individual discourse.

1. 3. 2 Writer-oriented Models

Writer-oriented models, alongside with the genre perspective, which belongs to text-based approaches, have developed as another *S/FLW* instruction trend that moves beyond the perception of texts as self-contained units. The mentioned models, focus on a text as discourse manifestation from the angle of the individual discourse of a writer and explores writing as an educational activity. They emphasise writer's knowledge that relates to discourse tenor, namely, writers and readers roles, frequency of their contact and affective involvement.

The overview of writer-oriented approaches is presented by Hyland (2002b), who has arranged them into three broad sub-approaches: writing as a personal expression, writing as a cognitive process and writing as a situated act. These sub-approaches reflect a shift from the perception of writing process as a self-discovery to the context in which a particular writing process takes place.

The view of writing as a personal expression is based on the work of Elbow (1998), Murray (1985), Moffett (1982), who hold the opinion that writing is a creative process of self-expression and self-discovery. The second trend is proposed by Zamel (1983, 1987), White & Arndt (1991), Raimes (1983, 1985), who explore writing as a cognitive process and put in the foreground the idea that writing is an exploratory and generative process during which writers discover, explore and reformulate their ideas. These researchers emphasize that the writing process comprises a number of steps, which reveal the cognitive complexity of the activities underlying each step. The third trend of writer-oriented approaches emphasizes the role of 'task environment' (Hyland, 2002: 30), namely, the local context of writing and its impact on writers. Kantor (1984), Newkirk (1984) have

investigated the role of the classroom as a community of writers and the encouraging stance of the writing teacher. Flower (1989) has added that the writing process incorporates also the prior knowledge of the writer, which is the result of, for example, such activities as reading, talking, observing, thinking, feeling, transcribing words on paper.

The writing process specific features brought out by the three sub-approaches manifested in the stages proposed by the researchers who view the writing process as a cognitive activity. Tribble (1996) emphasizes that the first models of writer-oriented approaches distinguished three linear stages: prewriting (specifying the task/planning and outlining, collecting data, making notes), composing and revising (reorganizing, shifting emphasis, focusing information and style for the readership), editing (checking grammar, lexis, surface features, for example, punctuation, spelling, layout, quotation, conventions, references). However, researchers are convinced (Zamel 1982, Raimes 1985) that the linear approach fails to give a full picture of the actual writing process. Accordingly, more recent writer-oriented models (*Figures 6* and *7*) differ from the initial three-stage linear model in two ways. Firstly, the recent models emphasize the recursive characteristics of composition process, and, secondly, they propose a more detailed staging of the writing process activities. For example, Tribble (1996) proposes five (*Figure 6*), whereas Coffin et.al (2003) seven (*Figure 7*) recursive stages. Thus in the process of developing transactional and/or more extended texts, writers tend to revisit the writing process stages, for example, return to pre-writing, which can result in the introduction of substantial changes within the text as well as doing additional research. Moreover, all writing process models incorporate the focus on language forms in a text, as these approaches envisage the editing stage before publishing.

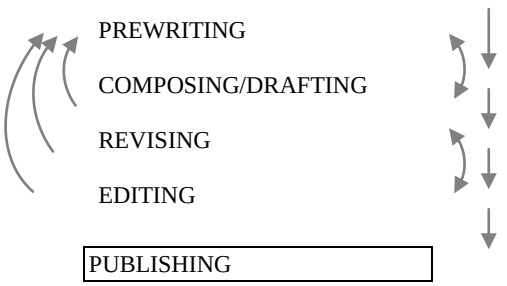


Figure 6**: Dynamic and recursive approach to the stages of writing process Tribble (1997: 39)**

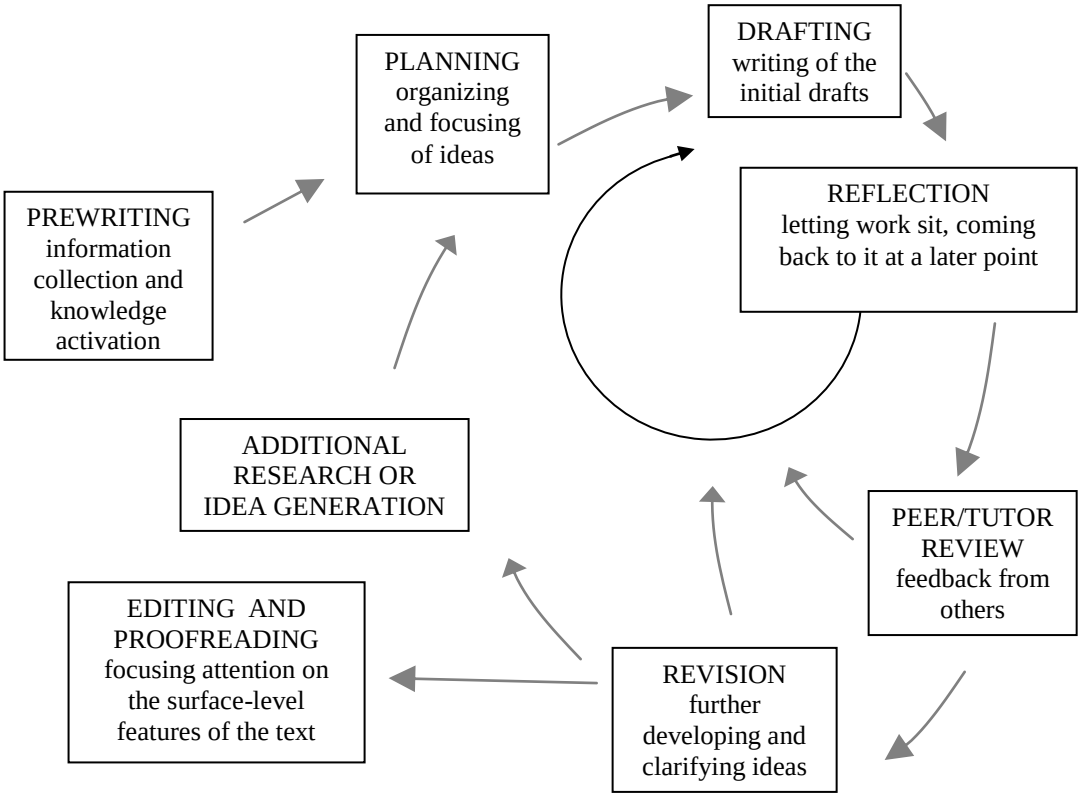


Figure 7: **Dynamic and recursive approach to the stages of writing process Coffin et al (2003: 34)**

In recent decades, with the increasing application of technology in written communication, namely, word processors and computer networks, the researchers (Bruce and Rubin, 1993; Chadwick and Bruce, 1989; Pennington, 1996, 1999, 2003; Warschauer, 1997, 1999) have explored the impact of technology on the writing process and concluded that the technology enhanced writing process is more circular, continuous, dynamic and collaborative than the conventional pen and paper-based text generation process.

Chadwick and Bruce (1989: 18) emphasize that text manipulation features of word processors 'allow writers to jump backwards and forwards in their texts, revise and rephrase, delete and insert […] make any number of changes without the […] fear of spoiling the presentation of the text'. These features of word processor enhance planning, drafting and revision stages of text composition. As a result, Pennington (2004: 76) concludes that word processor-enhanced writing process stages become so compact that they grow into 'continuous revision […] in which changes are made physically and temporarily near the point of generation of text, rather than being made in a separate step following the generation of a draft'. Technology-enhanced writing process stages,

therefore, merge into continuous generation and revising of a text or as Owston et.al (1992: 272) put it 'cumulative alteration'. The continuous revising promotes two-level improvement of texts: firstly, editing of language forms that is facilitated by spelling and grammar checkers, and, secondly, as Susser (1993), Pennington (1996) point out, meaning-level revision of texts, namely, correlation of the language forms with their functions within a particular text. Meaning-based revision is often combined with word-processed peer feedback and, according to Farneste (2006) can be supported by the implementation of track changes.

Apart from the word processor, which changes writing into a more circular and continuous process, the application of computer networks, namely, *CMC* messaging systems can make the composition process more dynamic. Stages of the writing process that envisage student and teacher-student collaboration, for example, peer and teacher feedback exchange, can be carried out through various *CMC* messaging systems, email, synchronous/asynchronous conferencing. The affordances and characteristic features of these messaging systems are presented in table 4 in chapter 2.2.3 *Language Variation in Written Electronic Discourse*.

Warschauer (1997, 1999) asserts that individual students' participation level in *CMC*-based peer feedback adds a new dimension to *S/FLW* by opening additional opportunities for written collaboration. Sullivan and Pratt (1996) have found that the feedback given by *ESL* student peers over the computer network is more focused, specific and reinforced by a greater number of students than the oral feedback. These features of *CMC* feedback exchange contribute to the creation of writers' community for sharing with knowledge and the development of a unified voice. Thematic, teacher-led feedback exchange via *CMC*, therefore, as Pennington (2004) asserts, has a value in helping student writers to develop objectivity, build their arguments and create an individual writer's voice. These theories give grounds for the author of the present study to conclude that *CMC*-enhanced collaboration during the writing process enables the students to practise two different types of written communication: firstly, *CMC*-based interactional texts, and, secondly, the development of transactional texts. Consequently, during such technology-enhanced writing process the students can experience and explore the discourse-tenor based linguistic variation across electronic discourse that ranges from interactional to transactional written communication.

Tribble (1996: 45), however, points out that writing process approaches, which focus on the individual discourse and the context of its production, tend to refrain from

'wider social and institutional influences which help construct writer's intentions and readers' expectations'. In order to expand the perception of communicative *context* and *readership* in written communication, Tribble suggests the integration of the previously explicated text and writer-oriented approaches.

1.3.3 Reader-oriented Models

The refinement and elaboration of text-oriented and writer-oriented models have led to their cumulative merging within reader-oriented models in *S/FLW* instruction or, according to Bjork (2003), writing as a social practice that focuses on social context of communication. The departure point of reader-oriented approaches is the perception of a text as the manifestation of discourse and its contextualization which includes: discourse as text, genre, individual writing process and social practice. Accordingly, this approach correlates with the linguistic theories that emphasize the broadening of the notion of *text* towards the concept of *discourse* (Widdowson, 2004; Bhatia, 2004). Therefore, as Bjork (2003) points out, an important aspect in reader-oriented approaches is the writer/reader contextualization framework. Consequently, according to the reader-oriented approach, *text* is a pragmatic process of discourse realization and as such comes from a particular contextual reality or *discourse* and through writers' variation of register pertinent lexico-grammatical features in a text enables the reader to co-construct the discoursal implication of that text.

Hyland (2002) distinguishes three sub-approaches within reader-oriented *S/FLW* instruction: writing as a social interaction, writing as a social construction and writing as a power and ideology. The social-interactionalist orientation aims at the broadening of the notion of context beyond the individual discourse of the writing process to the wider social contexts that the particular text represents. Consequently, writing as a social interaction emphasizes that meaning is created between the message sender and the recipient(s), and, as Hyland (2002b), Nystrand et.al. (1993) point out this approach emphasizes the adoption of texts to the needs, beliefs, and understandings of a particular audience. Thus, a writer attempts during the composition to understand the identity of the audience in order to apply relevant lexico-grammatical features. The audience, however, is not necessarily a concrete reality, but rather the projected audience that takes the writer into a wider pragmatic context. Thus, the perception of writing as a social interaction proceeds from individual message senders to the projection of the audience.

The second trend in reader-oriented teaching is writing as a social construction, which emphasizes writers' belonging to the socially and rhetorically constituted community of writers and readers. Hyland (2002b) states that a text conveys certain meanings and gains its force as a particular kind of action only within the community for which it is written, by exhibiting the patterns and conventions which reflect the sociocultural understandings of that community. Current *ESP/EAP*, therefore, on the one hand, acknowledges the importance of disciplinary conventions, but, on the other hand, denotes that writing is produced and mediated through writers' experiences of prior discourse rather than explicit knowledge of conventions and rhetorical patterns. Therefore instead of modelling the practices of experts, this approach promotes a guiding framework for producing texts in order to raise students' awareness of the connections between language forms, purposes and participant roles in specific social contexts.

The third trend – writing as a power and ideology, is based on *CDA (Critical Discourse Analysis)* that views discourse as a social practice and recognizes that there is a dialectical relationship between a particular event and the social structures within which it occurs (Dijk, 1997). From a pedagogical perspective, a major task of *CDA* is to help students become aware of how writing practices are grounded in social structures. *CDA* means revealing particular groups and contexts, and help students explore the requirements of their target communities. Thus, dominant and superior forms of writing can, according to Coffin et.al (2003) be viewed as simply another writing practice requiring to take up a different identity: learning to write as 'academics ', 'news reporters' or 'language teachers'.

Reader-oriented approaches to the written discourse development, like text and writer-oriented approaches are enhanced by technology applications, for example, corpora and other data bases, concordance, word processors and computer networks. There are two specific features concerning technology-enhancement of reader-oriented approaches. Firstly, although the above considered technology applications can be used selectively, depending on the goals of a particular *S/FLW* course or writing activity, the overall tendency of reader-oriented perspective is the integrative implementation of various technology applications by using multi-functional online environments. The second characteristic feature is the application of technology for the expansion of the communicative context from classroom or institutional to cross-cultural perspective. Thus Kern et. al (2006) emphasizes that the recent trend of *CMC* application in *S/FLW* includes the focus on both language learning and intercultural awareness, namely, socially broader

focusing on the audience. Cross-cultural written communication can be directed to the practice of the choice of lexocigrammatical features in the development of the texts for an authentic as well as presumable audience, which represents a particular socially and rhetorically constituted group. The construction of the audience might be practised by involving students in extended cross-cultural online writing projects and simulations. For example, the project *IDEELS (Intercultural Dynamics in European Education through Online Simulation)* aims at the involvement of students in authentic, collaborative, problem-solving online writing activities that require taking responsibility for the situational language use that varies from interactional to transactional writing. *Figure 8* contains the fragment of the transactional text, *Policy Statement*, which has been collaboratively developed by one student group and then posted in the online environment of the project *IDEELS*, so that other student groups from the other institutions could get acquainted with this document and proceed with subsequent written communication activities envisaged during the simulation session.

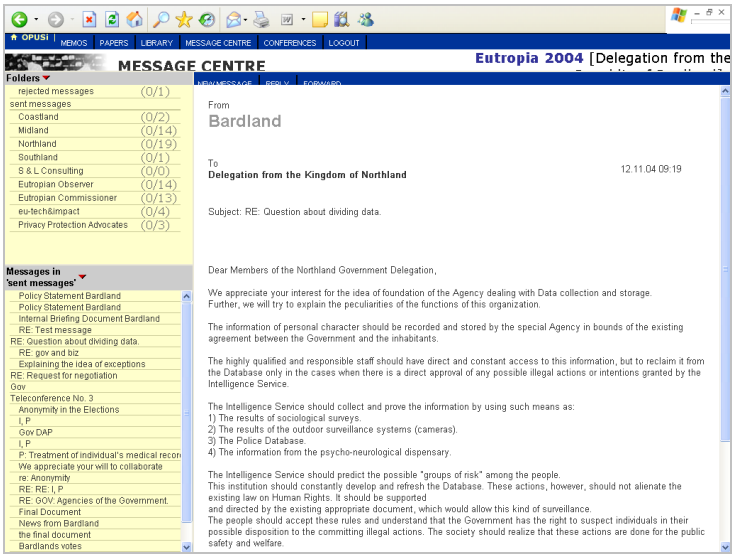


Figure 8: ***Policy Statement* posted in the online environment of the project *IDEELS***

The three above-considered groups of approaches to *S/FLW* (text, writer and reader-oriented) overlap, coexist and merge into integrated, technology-enhanced student-centred writing instruction that brings out the knowledge that writers need for written communication in diverse communicative situations. Matsuda (2003) also sides with the researchers, who consider that the pedagogical approaches, which are based on differing conceptions in *S/FLW,* are mutually interrelated. Tribble (2002) has summarized four knowledge areas (Table 6) that are brought out by each of the above illuminated

approaches to *S/FLW* and that writers require in order to develop a text that meets writers'/readers' needs in diverse communicative contexts:

Table 6: **Writers' knowledge (Tribble, 2002: 131)**

content knowledge	knowledge of the concepts involved in the subject area
writing process knowledge	knowledge of the most appropriate way of carrying out a specific task
context knowledge	knowledge of the social context in which the text will be read, and the co-texts related to the writing task
language system knowledge	knowledge of those aspects of the language system for the completion of the task

The author of the paper concludes that these knowledge areas correlate with the three contextual factors put forward by Halliday (1964); Halliday and Hasan (1995) which, according to them, determines register variation, namely, the choice of contextually pertinent lexico-grammatical features in a text that possesses distinct discoursal implication. The interrelation of the knowledge areas and the underlying situational factors are presented in Table 7.

Table 7: **Writers' knowledge and the underlying register factors**

<table>
<tr><td>content knowledge</td><td>discourse field</td></tr>
<tr><td>writing process knowledge and ICT skills</td><td rowspan="2">discourse tenor</td></tr>
<tr><td>context knowledge</td></tr>
<tr><td>language system knowledge</td><td>discourse mode</td></tr>
</table>

The author's analysis of the theories on the knowledge areas that writers need to know in order to vary the choice of lexico-grammatical features, the integration of *S/FLW* instruction approaches, and their enhancement with technology application, have enabled the author of the thesis to propose a technology-enhanced model (*Figure 9)* for students' involvement in writing and research activities aiming at the exploration and practice of register variation across the texts that range from interactional to transactional writing. This model aims at meeting the multi-purpose needs of the students of *BA/MA* programmes, namely, language variation in diverse written communication situations that these students might address in their future professional domains.

The model comprises three stages and each of these stages contains sub-stages. The first stage proposes research and writing activities aiming at the collaborative writing process of transactional texts (essays, research articles or papers within the framework of cross-cultural online simulation) as well as interactional texts (*CMC*-based communication

that supports the work at transactional texts). The texts developed during the collaborative writing process are stored in the messaging systems of e-course and the e-environment of the project IDEELS. The second stage, therefore, is devoted to the students' *CMC*-based collaborative research activities: the linguistic and functional analysis of the collaboratively developed transactional and interactional texts during the first stage of the writing activities. The research activities focus on the salient and dominant lexico-grammatical features and the exploration of their relevance in relation to their contextual peculiarities from the point of field, tenor and mode. The first two stages, collaborative electronic writing and linguistic and functional research of the electronic texts are a spring-board towards the third stage: the individual writing of transactional, electronic texts for the projected audience, emails and asynchronous discussion messages for a particular audience within the framework of *CMC*-based feedback exchange during the work at transactional texts. All three stages are interlinked: on the one hand, the first stage leads to the second and the third stages, but on the other hand, during the third stage activities, students draw on the experience, knowledge, conclusions as well as the texts developed and explored during the first two stages. Moreover, all three stages enable students to experience language variation across the texts that range from interactional to transactional written communication. This model underlies the e-course developed by the author of the present paper.

The structure of the e-course, the description of the collaborative writing and research activities that lead to the students' individual texts that include essays, emails and asynchronous discussion postings embedded within the above mentioned e-course as well as the arrangement of these texts in the students' electronic corpus, which is the research object of the present study, are considered in the chapter *Research Methodology and Procedure*.

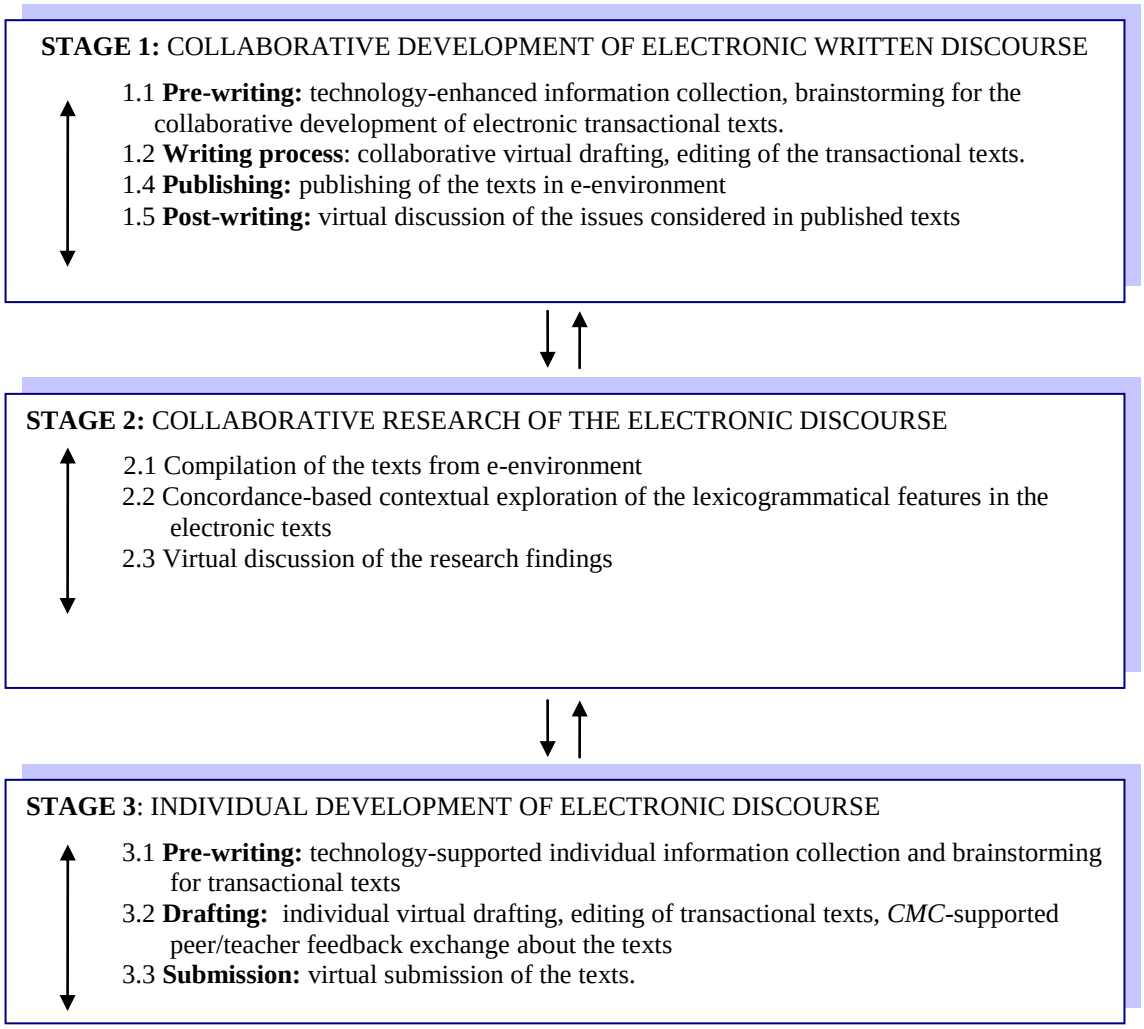


Figure 9: **Model of students' technology-enhanced writing process**

SUMMARY

The three groups of *S/FL* writing models, text, writer and reader oriented models reveal a paradigm shift from teacher to student-centred written discourse development, as text and writer-oriented models have developed and merged within the current technology-enhanced reader-oriented approaches. These approaches accumulate the knowledge writers require for bringing together language forms and their functions across the electronic texts that range from interactional to transactional written communication. The paradigm shift towards student-centred, technology-enhanced instruction allows the author of the paper drawing on researchers' findings and proposing a three stage technology-enhanced model for *NNS* writing instruction that aims at students' involvement in register pertinent language variation in order to meet their multi-purpose needs.

2 RESEARCH ON LINGUISTIC VARIATION IN STUDENT-COMPOSED ELECTRONIC DISCOURSE

The second part of this study includes a discussion on the research methodology and an analysis of the results of the study on linguistic variation in student-composed electronic discourse.

2.1 Research Methodology and Procedure

The present empirical research of the electronic discourse of *NNSE* is based on the pairing of research methods used in two linguistic areas: corpus linguistics and discourse analysis, which combine quantitative and qualitative investigation. The merits of such novel approach in the research of discourse have been claimed by Baker (2006), who asserts that using complimentary methods of analysis or forms of data, labelled by Newby (1977) as triangulation, substantially invest in the validity and reliability of the obtained results. The corpus-based quantitative research includes the application of *MDA* proposed by Biber (1988) in order to identify the frequency variation of linguistic features in the texts of *NNSE* corpus. The obtained results of variation are compared with the variation in texts that have been grouped into registers and studied by Biber (1988). The qualitative analysis aims at the exploration of the functional characteristics of electronic texts, which goes beyond the quantitative analysis.

The present empirical investigation was enabled by the implementation of e-studies at the *Faculty of Modern Languages* (*FML*), the *University of Latvia (UL)*. E-studies were initiated in 2001 by the decision of the *Senate* to launch *E-university* project at the *University of Latvia*. The goal of this in-process project is the promotion of innovative, technology-enhanced studies by creation, implementation and integration of e-courses within the conventional studies. This goal correlates with the objectives set out by *eEurope - An Information Society for all - Progress report for the Special European Council on Employment, Economic reforms and social cohesion towards a Europe based on innovation and knowledge* (2000) and *E-Latvia Programme* (2005).

Within the framework of *E-university* project, the students of *FML* were involved in the approbation of the model of technology-enhanced written electronic discourse development (*Figure 9*) during five academic years: 2001/2002-2005/2006. The

approbation included students' technology-enhanced collaborative writing during their participation in the project *IDEELS (Intercultural Dynamics in European Education through Online Simulation)* simulations and technology-enhanced individual writing envisaged by the activities within the e-course *English Academic Writing III*. Moreover, in 2002, when the e-course *English Academic Writing III* was designed, the writing activities of the project *IDEELS* simulations were integrated within the flow of the activities of the mentioned e-course. During the five-year period of approbation, 402 full-time students of *English philology BSP* were involved in the electronic text-based discourse development. The electronic texts of 102 randomly chosen students out of 402 students were selected for the inclusion in the study corpus.

The five year approbation results have been used to design new courses: *English Spoken and Written Communication I* and *English Spoken and Written Communication II*. Accordingly, the e-courses of the same titles have been developed and included in the *BSP* of *English Philology*. Since the bulk of the students' electronic texts, which have been selected for the study corpus of the present research, were developed during the approbation when the e-course *English Academic Writing III* was implemented, this title of the e-course has been used within the present study.

The goal of the research is to investigate the register-pertaining variation of lexico-grammatical features in the study corpus containing students' electronic texts developed during the previously mentioned five academic years (2001/2002-2005/2006). This discourse, therefore, comprises the study corpus compiled, arranged and analysed within the framework of the following investigation stages:

1. Investigation of e-environments and the design of e-course;
2. *NNSE* writing process of electronic texts by their involvement in technology-enhanced writing activities;
3. Extraction of *NNSE* electronic texts from e-environments and their arrangement in the corpus;
4. The choice of the linguistic features to be explored in the *NNS* corpus;
5. The structural and part-of-speech annotation of the texts included in the corpus;
6. The retrieval of the linguistic features from the electronic texts;
7. Descriptive statistics of the frequency of linguistic features.

The present chapter further provides stepwise outline of the outlined research methodology and procedure.

2.1.1 Electronic Environments

The first stage was devoted to the exploration of the electronic environment of the project *IDEELS* (*Intercultural Dynamics in European Education through Online Simulation)* and the design of the e-course *English Academic Writing III.* Both electronic environments have been used to implement the three-stage, technology-enhanced model (Figure 9 see chapter *Models of Written Discourse Development*) of students' written electronic discourse development.

The project *IDEELS* (2006) is located at the *University of Bremen* and *BSP* students of *English Philology* participated in the simulations within the framework of the co-operation between the *University of Bremen* and the *University of Latvia.* The project *IDEELS*, as it is stated in its homepage (*Figure 10*), is a virtual laboratory in which the participants (educators and students) of tertiary institutions throughout Europe are linked via online collaborative studies. Accordingly, one of the objectives of the project *IDEELS* simulations is the involvement of students in authentic, collaborative, problem-solving writing activities in order they could learn to recognize differences in register in written English. This objective served as a significant precondition for the integration of the collaborative writing activities in the flow of the e-course *English Academic Writing III.*

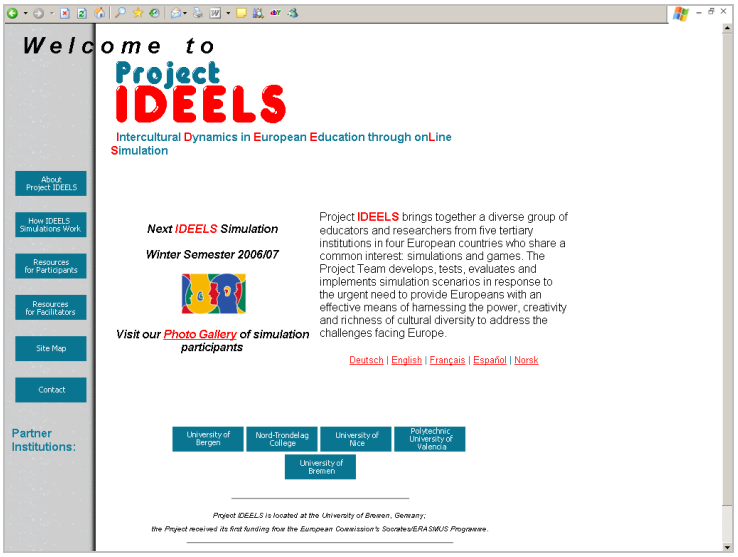


Figure 10: **Homepage of the Project *IDEELS***

The simulation world of the project *IDEELS* is a fiction country *Eutropian Federation* (*EF*). This fiction *Federation* incorporates countries that draw on real world

statistics and events in the countries of contemporary *European Union* (*EU*), which means that simulation scenarios deal with topical issues in *EU*.

During simulations, each country of *EF* is represented by a student group from a particular tertiary institution. These student groups act as high-level negotiators, who represent their simulated countries at the *Parliament* of *EF*. The delegations of the simulated countries collaboratively develop papers that contain valid arguments in order to promote the policy of their countries, however, aim also at the correlation of their interests with interests of the whole *Federation*.

The collaborative electronic texts were developed during the project *IDEELS* simulations by the implementation of the following messaging systems of the Project *IDEELS* interactive electronic environment *OPUSi*. This electronic environment provides a variety of functions, including individual, password protected user accounts for the access to the collaborative writing tools: *Papers* (see Figure 11) for the collaborative development of papers (*Policy Statement* and *Internal Briefing Document*), *Message Centre* (see Figure 12) for the exchange of letters and papers among the student groups that participate in the simulation, and *Teleconferences* (see Figure 13) that were implemented for the synchronous virtual negotiations aiming at the consensus building among the student groups who represent their simulated countries of *EF*.

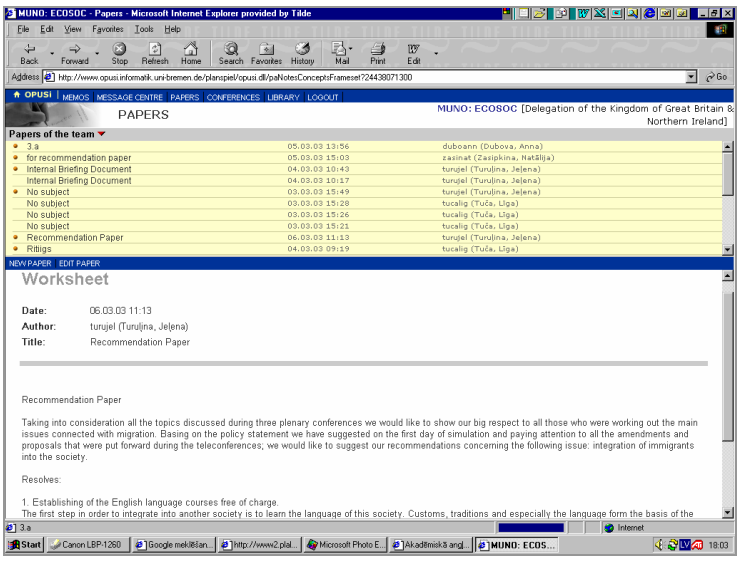


Figure 11: **A tool *papers***

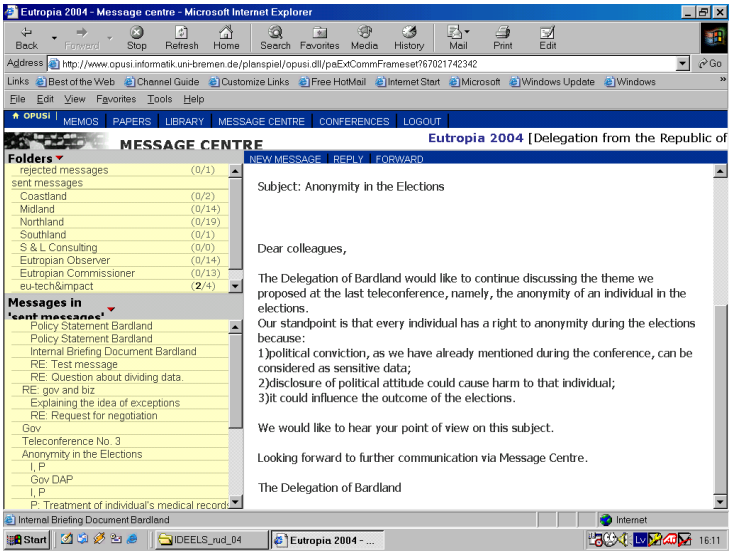


Figure 12: **A tool *Message Centre***

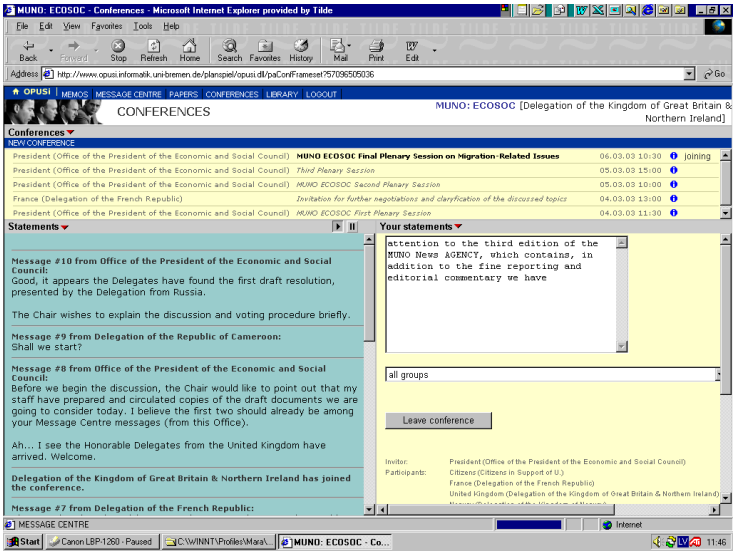


Figure 13: **A tool *Teleconferences***

The above-explicated arrangement of e-environment of the project *IDEELS OPUSi* enabled the first stage of writing process (Figure 9), namely, collaborative virtual writing.

The second two stages of writing process (see Figure 9 see chapter *Models of Written Discourse Development*) have been technology enhanced by e-course *English Academic Writing III*. It was designed on the basis of the course *English Academic Writing III* (*English Philology BSP*), worked out by Professor Ingrīda Kramiņa. The e-course has been designed by two lecturers of the *Department of English Studies*, one of which is the author of the present study. The e-course *English Academic Writing III*, which has been acknowledged by the annual e-course competition board of the *UL* as one of the most

successfully designed courses in 2002, was integrated within the *BSP* of *English Philology* and underwent approbation.

This e-course contains six structural parts that are reflected in its homepage (*Figure 14*): *Syllabus, Calendar, Course Content, Studies & Assessment, Reference Materials, Communication.* Each of the mentioned parts provides specific scaffolding for technology-enhanced, longitudinal collaborative and individual writing and research activities.

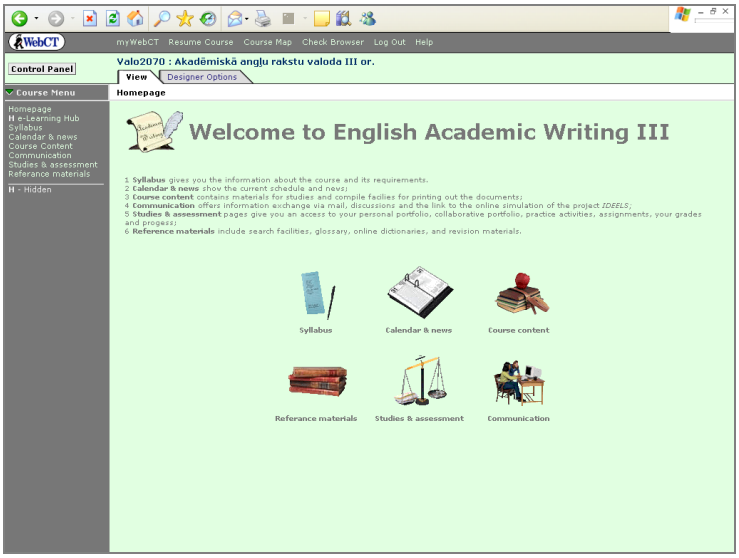


Figure 14: **Homepage of e-course '*English Academic Writing III*'**

The aim of *Syllabus* and *Calendar* is to inform the students about the content and flow of the writing activities. *Syllabus* contains the information of the course goals and overall outline of the activities, whereas C*alendar* displays updated schedule of each particular activity.

Course Content (*Figure 15*) contains the collection of the materials, developed and arranged by the authors of the e-course for the students to use during all stages of electronic discourse development and investigation by them.

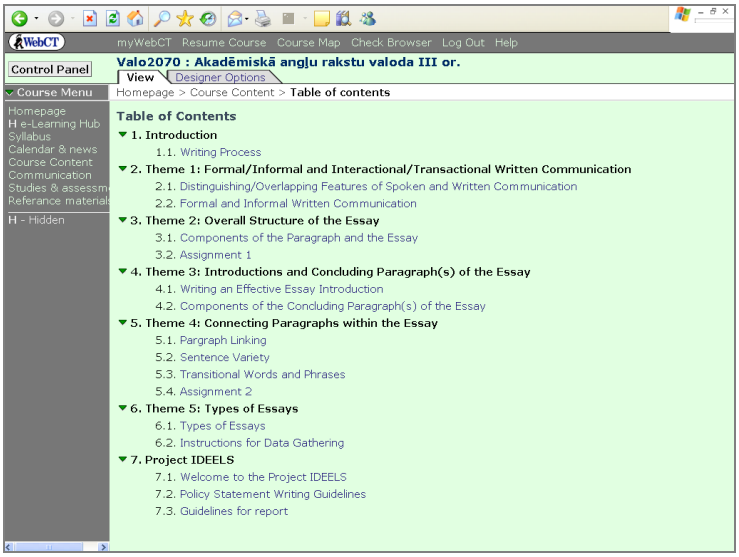


Figure 15: **E-course *English Academic Writing III*: *Course Content***

The part labelled *Reference Materials* (*Figure 16*) incorporates links to online dictionaries and the text retrieval programme *Complete Lexical Tutor*. These tools enhance the students' choice of lexis in their electronic texts and facilitate their research activities directed towards exploration of their collaboratively developed texts.

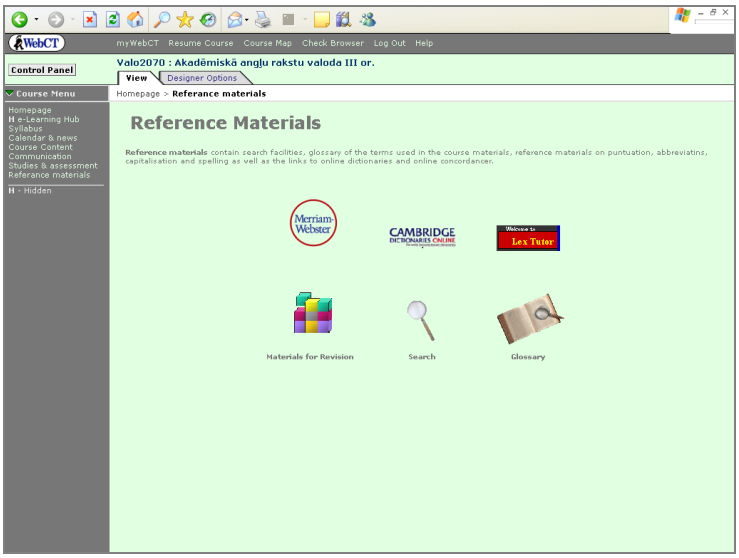


Figure 16: **E-course *English Academic Writing III*: *Reference Materials***

The part *Studies and Assessment* (*Figure* 17) offers two types of scaffolding. Firstly, the practice activities, *Self-tests* and *Practice tests*, which alongside with the students' research activities bridge up their collaborative and individual writing. The second type of scaffolding is the provision of the sites labelled *My Files*, which have been created for each student enrolled in the e-course, by the adoption of the presentation tools

of *WebCT* (version CE 4.1). This idea of the author of the present study has been put into practice to overcome the limitations of *WebCT* (version CE 4.1) *My Files* can be used by students for storing their drafts and other documents generated during individual writing. Apart from the mentioned practice activities and the sites *My Files*, the part *Studies & Assessment* also includes *Assignments* for the submission of the documents for grading (the final drafts and the selection of drafts reflecting the students' writing process) as well as *My Grades* and *My Progress* that show the students' achievement throughout the course.

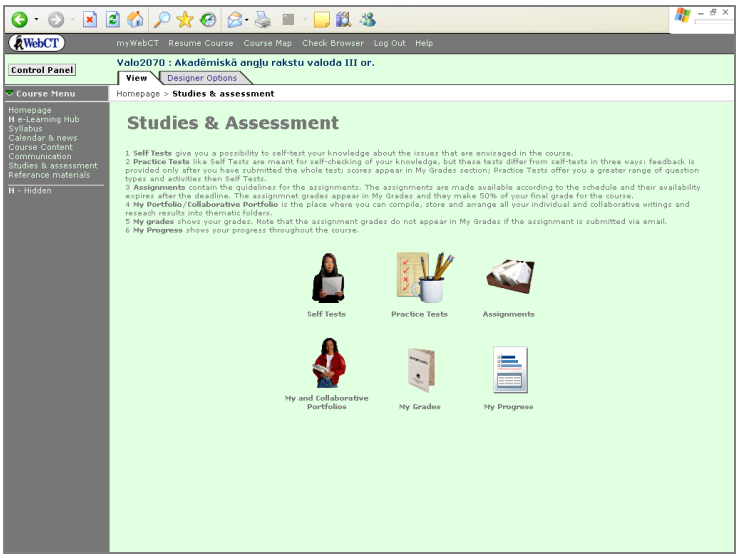


Figure 17: **E-course *English Academic Writing III*: *Studies & Assessment***

The structural part *Communication (Figure 18)* provides interactive scaffolding: messaging systems for asynchronous discussions and email communication. Asynchronous discussions are envisaged for teacher-moderated online feedback exchange among the students on their essays, whereas student/teacher emails are envisaged for the facilitating students' individual writing process and research. Apart from the mentioned interactive scaffolding, *Communication* page of the e-course contains a link to the project *IDEELS* and its messaging system *OPUSi.*

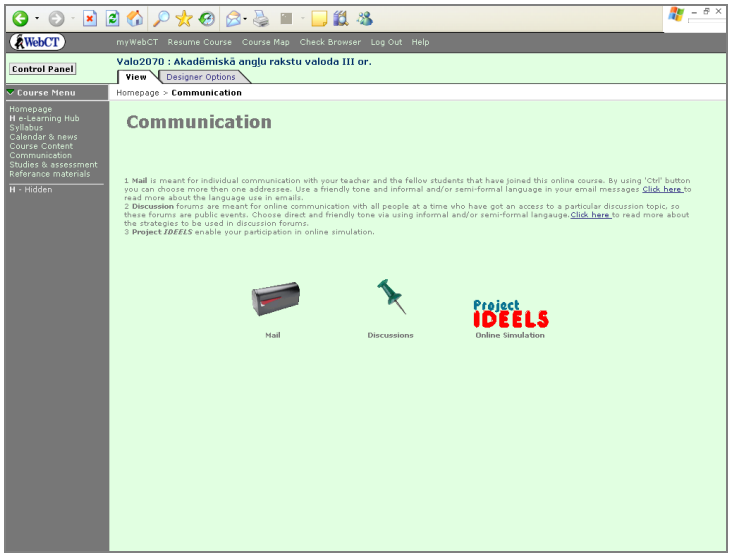


Figure 18: **E-course *English Academic Writing III*: *Communication***

The above-considered e-course *English Academic Writing III* provides scaffolding for students' involvement in technology-enhanced writing of interactional and transactional electronic texts. Moreover, the presented technology-enhanced approach is transposable to other online educational environments, for example, *Moodle*, which incorporate a range of tools that can be adapted and tailored according to the mentioned approach. Therefore, this approach is applicable irrespective of the variation of the affordances embedded in the e-environments currently applied to enhance language studies.

2.1.2 Technology-enhanced Writing Activities

The texts that have been included in the study corpus, which the students developed during collaborative and individual writing, were enabled by the above presented two e-environments. This chapter presents the situational characteristics and the flow of collaborative and individual writing activities.

2.1.2.1 Collaborative Writing Activities

Collaborative writing activities, according to the model presented in *Figure 9*, are the activities during which the students collaboratively developed electronic texts and took shared responsibility for them. These texts were written during the students' participation in the simulations of the project *IDEELS* (*Intercultural Dynamics in European Education*

through Online Simulation): *Human Rights* (the academic years 2001/2002; 2002/2003) and Data *Access and Protection* (the academic year 2004/2005).

The elaborate simulation world, scenarios and the roles of the participants contribute to the authenticity of the communication during these simulations. The idea of the authenticity of the communication during simulations in *ELT* (English language teaching) is explained by Jones (1990: 5). He defines simulations as complex, longitudinal activities that possess a great degree of authenticity because 'they are the reality of function in a simulated environment'. The reality of function of the project *IDEELS* simulations is that the participants act, react, communicate and even think as real representatives of the simulated countries.

The characteristics of the simulation world, the roles of the participants and affordances of the messaging system used during simulations shaped the contextual framework of collaborative writing activities and determined what knowledge writers required in order to meet the communicative goals of these activities. *Table 8* contains a summary of the situational factors that, as shown by the author of the paper, underlie the knowledge writers needed during the collaborative writing activities of the above-mentioned simulations.

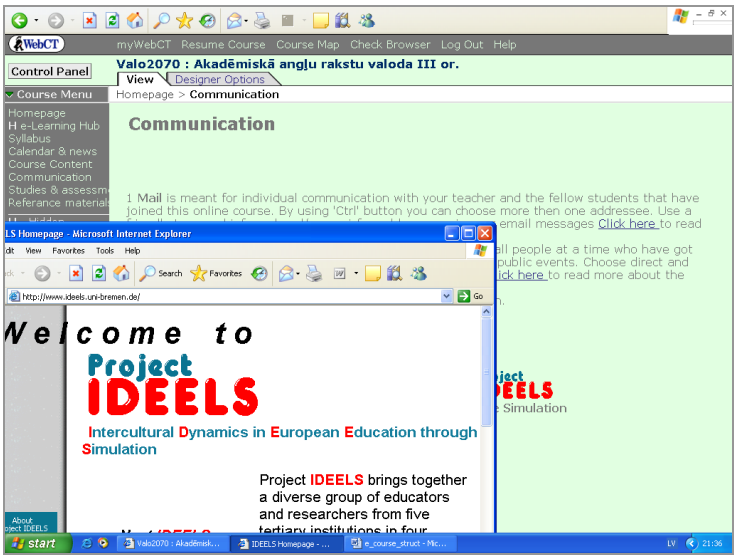


Figure 19: **The link to the project *IDEELS* web site from the course *English Academic Writing III***

Collaborative writing activities of simulations included the following stages: pre-writing, writing and post-writing activities. During all three stages the students used the

project *IDEELS* web site as well as the materials included in e-course *English Academic Writing III.* Therefore, e-course incorporated the link to the *IDEELS* web site (*Figure 19*).

The pre-writing stage was devoted to brainstorming and information collection for writing of transactional texts that were expected to be professionally written, formal documents: *Internal Briefing Document* and *Policy Statement. Internal Briefing Document* should explicate the policy goals and negotiating strategies of the delegations of *EF*. The *Policy Statement* in many ways was meant as a public version of the *Internal Briefing Document*. In it, students were also supposed to state the policy of their countries on the issues of the current simulation, without revealing their negotiation strategies. Pre-writing, therefore, envisaged a detailed exploration of the simulation world and documents about the simulated countries and the whole *Federation* as well as the e-course materials on the linguistic characteristics of formal documents and pre-writing strategies. Pre-writing stage finalised with the development of the overall outline of the structural parts of these two documents.

The next stage was drafting and revising of *Internal Briefing Document and Policy Statement.* Firstly, all students split into small groups (3 to 5 students in each small group) and drafted one particular thematic section of both documents. Then, they put the section drafts together in each of the two documents and started the revising and editing of these documents. Their work was facilitated by the e-course *English Academic Writing III* materials on the peculiarities of register-pertinent variation of lexico-grammatical features in transactional texts. Apart from the e-course, the students applied the messaging system *OPUSi* tool labelled *Papers*. This tool (*Figure 11*) enabled online development and revision of multiple drafts by all students who belonged to the delegation of a particular country of *EF*. The functional characteristic of the tool *Papers* encouraged the continuous, circular revising based on dynamic, interactive online feedback exchange that was reinforced by all students who participated in the development of the *Policy Statement* and *Internal Briefing Document*.

The writing process was finalised by sending the *Policy Statement* to all delegations of the countries of *EF* and forwarding the *Internal Briefing Document* to the simulation coordinator – *Commissioner* of *EF*. These documents were forwarded through *Message Centre* of the messaging system *OPUSi* (Figure 12). The students who acted as representatives of different countries of *EF* could get acquainted with the policy revealed in the *Policy Statements* by the representatives of each country of *EF*.

Post-writing incorporated two text-based online communication activities: participation in teleconferences and the exchange of online letters. Both writing activities were devoted to the virtual, text-based negotiations the departure point of which was the arguments put forward in the *Policy Statements*. Post-writing activities, therefore, revealed how far the students had succeeded in the explication of their policies.

Each teleconference (in a form of synchronous online discussion), which was enabled by the tool *Teleconferences* (*Figure 13*) of the messaging system *OPUSi* (*Figure 19*), was scheduled and devoted to particular issues. Although the teleconferences had a detailed agenda, the exact advance planning was hardly possible due to the dynamic flow of the synchronous virtual negotiations.

The purpose of the second post-writing activity – the development of electronic letter – was the negotiations of the delegations between the scheduled teleconferences. These letters aimed at the thematically specific information exchange directed to the consensus building on particular issues. Electronic letters were exchanged via *Message centre* (*Figure 12*) of the messaging system of *OPUSi*.

The final post-writing activity that simultaneously functioned as a post-simulation activity was a post-simulation teleconference (in a form of synchronous online discussion) that aimed at the disclosing of the identities of the simulation participants and the reflection on their insights. These postings, therefore, considerably vary from the simulation teleconference postings due to the shift of the situational characteristics. During post-simulation teleconferences, the students, who had acted as *Members of Parliament* during the simulations, had left the simulation world and participated in the post-simulation teleconferences as the students of a particular European tertiary institution.

The goal of the previously considered collaborative pre-writing, writing and the post-writing activities of the simulations was to expose the students to the correlations of linguistic variation according to the contextual framework (field, tenor, and mode) summarized in *Table 8*. During collaborative pre-writing, writing and post-writing activities the students have practiced the development of the texts that have been arranged into three register groups (*Policy Statement* and *Internal Briefing Document*).

2.1.2.2 Individual Writing Activities

Individual writing activities, according to the model presented in *Figure 9*, are the activities during which the students individually developed transactional and interactional electronic texts and took individual responsibility for them.

These texts have been generated and embedded within the e-course *English Academic Writing III*. During individual writing activities the students wrote essays for the presumable audience that, according to Barton (2004: 71), has been labelled as 'a general academic audience', as well as for their peers and a teacher. Transactional electronic texts developed for a general academic audience included essays, whereas interactional texts comprised teacher moderated online feedback exchange among the students about their transactional texts as well as teacher/student email exchange about the flow of their writing process. These characteristics of the communication environment determined the contextual framework of individual writing activities, namely, the knowledge the students required for the communicative situation pertaining application of lexico-grammatical features within their transactional and interactional electronic texts.

The individual writing activities contained the following stages: pre-writing and writing. Pre-writing covered brainstorming, namely, information gathering when the students used the materials of the e-course *English Academic Writing III* on writing process peculiarities and brainstorming strategies as well as they saved their drafts in *My Files*: the students' individual sites created for storing of their documents.

The second stage, writing, was devoted to the development of the drafts and revising/editing of these drafts. This writing stage was enhanced by interactive scaffolding: teacher moderated, online peer feedback exchange on meaning level linguistic characteristics of the essays and the writing process itself. *Figure 20* shows the screenshot of a discussion thread. As the discussion postings were students' individual messages, the feedback on each essay was reinforced by a number of students. This feedback encouraged the refinement process of the essay drafts.

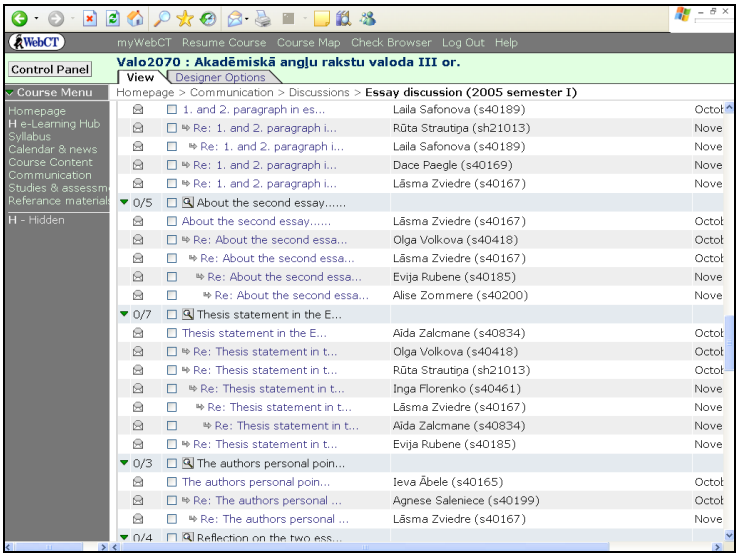


Figure 20: **E-course *English Academic Writing III*: Teacher-moderated online discussion of the essay drafts**

Moreover, during the individual essay writing process, the students communicated with the teacher using the e-course email (*Figure 21*) about their writing process management peculiarities and the linguistic characteristics of their essay drafts.

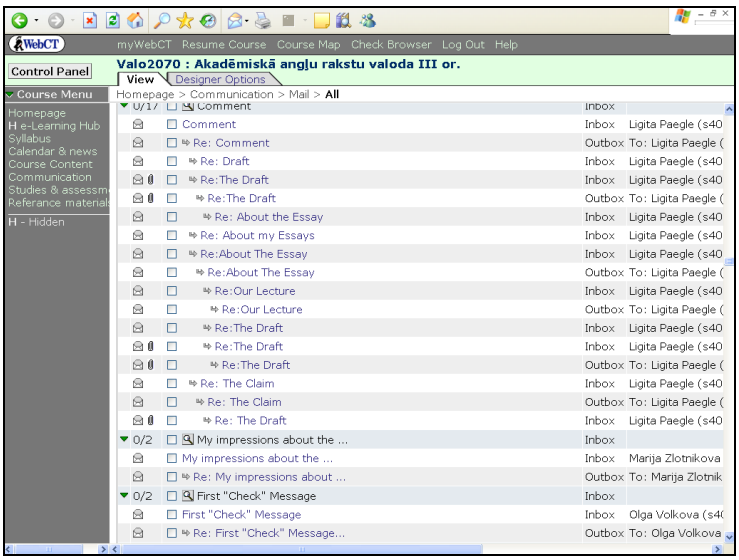


Figure 21: **E-course *English Academic Writing III*: student/teacher email communication**

After the completion of pre-writing and writing, the students submitted the final drafts of the essays by uploading them in *Assignments* and thus making them available for the teacher. The essays, apart from submission for grading, were published by creating the index page in the location *My Files* and a link was put to the document within *My Files* that

contained their final draft. Thus, the essays were made available for the whole group of the students who have been enrolled in the e-course.

During individual writing activities the students developed the electronic texts that have been arranged into three groups requiring register variation: academic texts (essays), e-letters (emails) and messages (synchronous and asynchronous online discussions). Collaborative and individual writing activities enabled the students to practise situationally diverse written communication that required variation of linguistic features according to the contextual factors. The additional challenge for the students was the overlapping of online writing situations. For example, during collaborative writing, synchronous online teleconferences overlapped with the exchange of electronic letters. During individual writing activities, teacher-moderated online discussion of the essay drafts overlapped with the students' work at the refinement of their own essays. The students, therefore, had to switch over from one virtual communicative situation to another to reach their communicative purpose of a particular writing activity. Such dynamic shift across the transactional/interactional continuum requiring register pertaining variation, revealed to the students the subtleties of language variation in diverse text-based communication situations that, according to their multi-purpose needs analysis (Kramiņa, 2000), they intend to address in their future professional domain.

Moreover, in order to explore the impact of the practice of register variation on the students' advancement of register-pertaining language use in transactional texts, before and after their participation in collaborative as well as individual writing and research activities, they were required to write timed pre-essays and post-essays. The aim of the pre-essays was to explore the students' initial proficiency in the choice of lexico-grammatical features in transactional texts. The post-essays aimed at the identification of the possible correlations of the lexico-grammatical features used by the students' after their participation in the technology-enhanced writing and research activities.

The timed essays as well as the texts of students' electronic discourse comprise the bulk of the investigation material, the students' corpus, which ranges from transactional to interactional electronic texts. Thus, the electronic environments used during the students collaborative and individual writing activities (Project *IDEELS* online environment *OPUSi* as well as *e-course English Academic Writing III*) have been used as both the written communication language study tools by students and the research tools, enabling the electronic acquisition of learner data for the compilation into the study corpus.

Table 8: **Situational factors and knowledge required in writing collaborative and individual electronic texts**

SITUATIONAL VARIABLES (Halliday, 1964; Martin, 2003)	**WRITERS' KNOWLEDGE** (Tribble, 2002)	**STUDENTS' KNOWLEDGE**	
		COLLABORATIVE WRITING ACTIVITIES	**INDIVIDUAL WRITING ACTIVITIES**
DISCOURSE FIELD	Content knowledge: knowledge of the concepts involved in the subject area	The knowledge of the issues to be considered in the interactional and transactional texts during simulations of the project *IDEELS*	Knowledge of the issues to be considered in the interactional and transactional texts
DISCOURSE TENOR	Writing processes knowledge: knowledge of the most appropriate way of carrying out a specific task	Writing process knowledge of transactional and interactional texts enhanced by the affordances of the messaging systems used by the project *IDEELS*	Writing process knowledge of transactional and interactional texts enhanced by the affordances of the messaging systems of the e-course *English Academic Writing III*
	Context knowledge: knowledge of the social context in which the text will be read, and the co-texts related to the writing task	Knowledge of the virtual simulation world and the simulation roles of participants	Knowledge of the expectations of the audience (general academic audience, peers and teachers)
DISCOURSE MODE	Language system knowledge: knowledge of the aspects of the language system for the completion of the task	Knowledge of language system for the development of transactional and interactional electronic texts within the framework of the project *IDEELS*	Knowledge of language system peculiarities for the development of transactional and interactional electronic texts through the e-course *English Academic Writing III*

2.1.3 Corpus of Student-Composed Texts

Researchers maintain (Meyer 2002; Kļaviņa, 1980) that any study corpus should be quantitatively and qualitatively representative. Namely, in order to reach a particular investigation goal, a corpus should contain a relevant range and magnitude of text samples. The quality and quantity of learner-data corpus depend on two factors: the goals of a particular study (choice of linguistic features to be investigated in a particular learner data) and access to the technology that is required for the acquisition and exploration of these data. De Haan (1992) emphasises that the optimum learner corpus size depends on the specific linguistic features subjected to the analysis. For instance, according to de Haan relatively small samples of words (for example, 20,000 words) might be sufficient for studies that involve high-frequency words and structures. Sinclair (1995) draws attention to the learner corpus compilation difficulties by stating that even in the most technically advanced countries, the learner corpus compilation is a painstaking process. In the light of these assumptions, Granger (1998) concludes that learner corpora of most *S/FLA* studies are far from the size of highly representative native corpora, for example, *British National Corpus* (*BNC*) due to the heterogeneity of learners and learning situations. The heterogeneity brings out the importance of the precise definition of the design criteria for a particular learner corpus.

The design of the specialized study corpus for the present investigation comprised the following two stages: the compilation of the initial corpus and the selection of texts from the initial corpus for the inclusion in the study corpus. The first stage included the compilation of the electronic texts developed by 402 full-time students (*Table 9*) who had completed the course *English Academic Writing III* during five academic years: 2001/2002 – 2005/2006 and participated in the technology-enhanced collaborative and individual writing activities.

All electronic texts of 402 students were manually extracted from the messaging systems of the online environments (e-environment of the project *IDEELS* and the e-course *English Academic Writing III*). These texts were grouped according to three broad groups of registers (academic electronic texts, electronic letters, and electronic messages) and holistically explored to identify the overall distribution of the lexico-grammatical features relating to the situational factors (or dimensions) as put forward by Biber (1988; 1995). For example, the holistic examination of the initial corpus revealed the application of the

personal pronouns *I, we, you* across the texts that differ according to the writing activity type (see chapter *Technology-enhanced Writing Activities*). The frequency of these personal pronouns, according to Biber can range the texts from transactional to interactional (see chapter 2.2.2 *Multi-dimensional Analysis of Language Variation*), depending on the message senders' involvement in the communicative situation. The detailed list of the lexico-grammatical features chosen for the present investigation is included in chapter 2.1.3 *Implementation of dimensions and linguistic features*.

Table 9: **Number of students involved in the writing of electronic texts**

Academic years	**Number of students whose electronic texts had been included in the initial corpus**	**Number of randomly chosen students whose texts had been included in the study corpus**	
		Number	*%*
2001/2002	78	17	21.8
2002/2003	76	22	28.9
2003/2004	91	21	23.0
2004/2005	84	18	21.4
2005/2006	73	24	32.9
Total	**402**	**102**	**23.37**

After the holistic exploration of the initial corpus, a range of the electronic texts representing the above-motioned three register groups were selected for the inclusion in the study corpus. The texts were randomly chosen from the discourse developed by the students who had participated in collaborative and individual writing activities. As a result, the discourse of 102 subjects (namely, 23.37 % of the total number of 402 students) was selected from the initial corpus for the inclusion in the study corpus. All subjects have been assigned codes (student 1, student 2, student 3 … student 102) according to the five academic years during which their electronic texts were developed.

Next, before the arrangement of the texts within the study corpus, its design criteria were worked out and its type defined. For the learner corpus design, Ellis (1994) and Granger (1998, 2002, and 2004) suggest two criteria groups. One criteria group discriminates the attributes that are common to all texts included in a corpus (shared features) and the attributes that vary across a corpus (variable features). The other criteria group aims at the distinction between the features that pertain to learners and the features that relate to learning situations.

The arrangement of the students' corpus is based on four shared and two variable features (see Table 10). Four shared features pertain to the students who had developed the

electronic discourse: age, language proficiency level, learning institution and the title of BS programme. All 102 students whose texts have been included in the study corpus were young adults (c. 19-22 years old), who have been studying the English language in a non-English speaking environment i.e. *EFL* students. All subjects were *BSP* students of *English Philology* in the second year of their studies at the same tertiary institution: the *University of Latvia*, the *Faculty of Modern Languages.* These students had completed the prerequisites required for the enrolment in the course *English Academic Writing III*: they had completed the courses *English Academic Writing I* and *II*. The fourth shared feature relates to the peculiarities of language study situations. Although the students' electronic texts have been developed by the application of various electronic tools, they all belong to written communication based on the semiotic system, and therefore have been labelled as written electronic texts. The two variable features underlying the corpus design (*Table 10*) relate to the language study situations, namely, the task type (for example, individual and collaborative writing activities) and the academic year when the particular texts have been written.

Table 10: **Corpus design criteria**

Design criteria	Shared features	Variable features
students	s students' age language proficiency level learning institution BA programme	–
study situations	communication mode	task type academic year

The structure and magnitude of this corpus is presented in *Figure 22*.

The number of tokens within the texts included in the students' corpus has been counted with the help of *WordSmith Tools* (Scott, 2004). Thus, the variable attribute, task type, splits all electronic texts into two broad groups: 290 collaboratively developed texts that contain 16, 386 tokens and 939 individually generated texts that contain 203,629 tokens. The situational peculiarities of collaborative and individual writing activities have been outlined in the chapter *Technology-enhanced Writing Activities.* These peculiarities of task setting account for the magnitude variation of the texts: their number and bulk. Such unbalanced corpus is relevant for the present study, as apart from quantitative analysis (frequency counts of linguistic features); a more detailed discourse analysis has been

performed on the exploration of complete texts of various magnitudes. However, these variation of magnitude have been taken into account in quantitative analysis; since the raw counts of the lexico-grammatical features have been normalized to the text lengths of 1000 words (see the chapter *Descriptive Statistics of Linguistic Features*).

Both collaborative and individual texts have been further split into transactional and interactional texts. Collaborative, transactional electronic academic texts include 8 academic texts (statements) that comprise 4096 words. Collaborative interactional texts comprise 25 collaborative e-letters containing 6253 tokens, and 257 messages of synchronous conferences that cover 6037 tokens. All messages have been further arranged in two smaller groups: 163 messages of simulation conferences (5226 words) and 94 messages of post-simulation conferences (811 tokens).

Individually developed transactional texts comprise 359 essays (138, 519 tokens). These essays have been arranged into two sub-groups: 155 untimed essays (69, 257 tokens) and 204 timed essays (69, 262 tokens). Timed electronic essays have been split into 102 pre-essays that contain 30, 440 tokens and 102 post-essays that contain 38, 784 tokens. Individually written interactional texts include 498 e-letters (emails) of 41, 528 words total, and messages of asynchronous discussions represented by 303 texts that comprise 23, 579 tokens.

The third variable feature is the academic year when the texts were developed. All collaboratively and individually developed texts that have been arranged into groups have been further grouped according to the academic years when they were developed (2001/2002 – 2006/2007) and arranged within corresponding files. For example, 102 electronic pre-essays have been arranged into five files and each file contains the essays developed during a particular academic year.

The text groups from the corpus that have been selected for *MDA* have been abbreviated and labeled in the tables and figures of the present paper as follows:

1 collaborative statements	c. statements
2 collaborative e-letters	c. e-letters
3 collaborative messages of synchronous simulation conferences	c. sim. messages
4 collaborative messages of synchronous post-simulation conferences	c- post-sim. messages
5 individual messages of asynchronous discussions	ind. messages
6 academic pre-essays	pre-essays
7 academic post-essays	post-essays

This arrangement aims at the revealing of the linguistic characteristics of the texts developed in contextually varying communication situations in language studies. The

proposed design of the corpus of student-composed texts allows exploration of the selected lexical and grammatical features across the texts grouped according to the task type.

Such specialized corpus containing complete electronic texts written by students allows the exploration of the lexico-grammatical features across the collaboratively and individually developed interactional and transactional texts as well as the pre-essays, written before the participation in the writing activities and the post-essays developed after the participation in the writing activities. The structure and magnitude of the study corpus is presented in *Figure 22*.

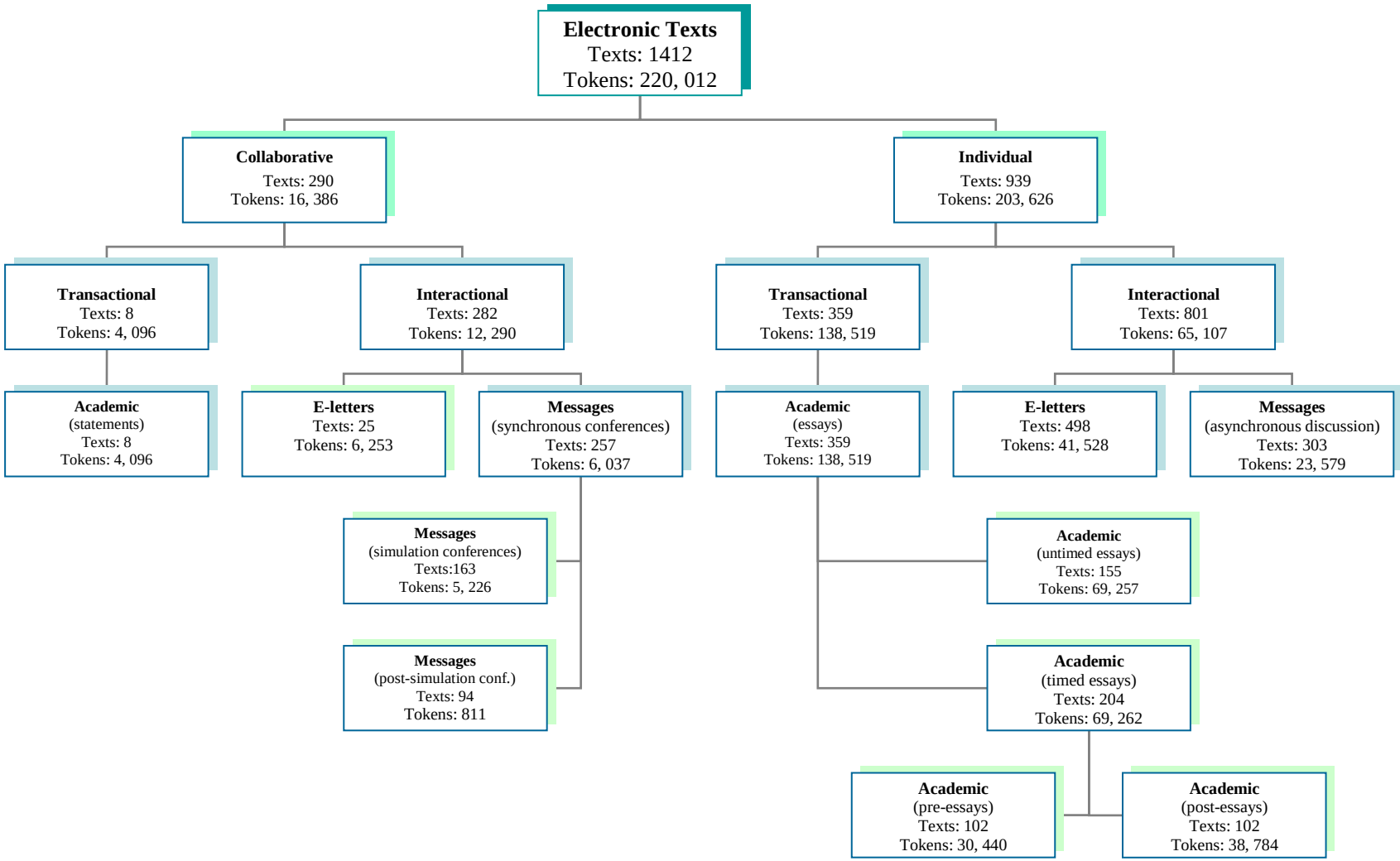


Figure 22: **Corpus of student-composed texts**

2.1.4 Implementation of Dimensions and Linguistic Features

The corpus of student-composed texts contains collaboratively and individually developed interactional and transactional texts that have been arranged in corpus according to the type of writing activities. In order to explore the language variation in these texts the linguistic features were selected that are related to the three dimensions (1, 3 and 5) of Biber's (1988) proposed six dimensions have been chosen for the present study. These dimensions have been selected, as Xiao and McEnery's (2005) findings show that the linguistic features clustered within them can specifically reveal the language usage variation in interactional and transactional texts of written discourse. In the present study, the selected three dimensions, like in Xiao and McEnery's study (ibid.), have been labelled Dimension A, Dimension B, and Dimension C. A more detailed theoretical background of the functional characteristics of these register dimensions that range the texts from more transactional to interactional have been presented in chapter 1.2.3 Multi-dimensional Analysis (MDA) of Discourse. The selected linguistic features within the framework of these dimensions have been listed below; however, Appendix 3 contains the examples of these linguistic features extracted from the texts of corpus.

Dimension A: Involved versus informational production

(+) private verbs, contractions, present tense verbs, second person pronouns, *do* as a proverb, the analytic negation *not*, demonstrative pronouns not followed by a noun, general emphatics, first person pronouns, the pronoun *it*, be as main verb, causative subordination because, discourse markers, indefinite pronouns, general hedges, amplifiers, sentence relatives/non-restrictive relative clauses, *wh* questions, possibility modals, non-phrasal coordination, *wh* clauses, final prepositions.

(-) other nouns, word length, prepositions, type/token ratio

Dimension B: Explicit versus situation dependent reference

(+) wh relative clauses, pied piping constructions, phrasal coordination, nominalizations,

(-) all adverbs of time, all adverbs of place, other adverbs.

Dimension C: Abstract versus non-abstract information

(+) conjuncts, agentless passives, by-pasives, past participal whiz deletions, other adverbial subordinators.

2.1.5 Marking and Annotation of Corpus

The whole study corpus (see Figure 22) was structurally and part-of-speech annotated. This annotation process aimed at enabling of the extraction of the linguistic features and examples from the corpus of student-composed texts..

2.1.5.1 Structural Marking

Structural marking aims at the provision of descriptive information about the texts included in the study corpus. Descriptive markup can be maintained even if the corpus is transferred from computer to computer. The texts of the present study corpus, therefore, have been structurally annotated according to the current standard for the markup of electronic documents *SGML* (*Standard Generalized Markup Language*) by using a marking system proposed by *Text Encoding Initiative (TEI)* (Burnard, Sperberg-McQueen, 2002). *TEI* is a recognised attempt to develop general guidelines for electronic text representation. The importance of the application of such comprehensive and consistent descriptive markup has been brought out by researchers (Granger, 1998; Johansson, 1994; Meyer, 2002; Spektors 2001).

TEI system has been developed to annotate a wide range of written corpora especially in humanities. Burnard (1995) distinguishes three types of tagsets that can be used in the electronic documents annotated with *TEI*-conformant markup: core tagsets (tags associated with file headers or paragraph divisions and other related tags), base tagsets (tags associated with particular kinds of texts, for example, drama, verse, speech) and additional tagsets (any tags that might be added to the core and base tagsets). The importance of the application of core tagsets in students' corpora has been emphasised by Granger (1998), who considers that descriptive information of the texts should be recorded and made accessible to the analyst.

Accordingly, all files included in the students' corpus were manually assigned *TEI* tags that belong to the core tagset. These tags reinforce the students' corpus structure based on the task type (or communicative purpose of a text).

The *TEI* header, as explicated by Johansson (1994) and further developed by Burnard and Sperberg-McQueen (2002), contains information about sources, text selection/sampling, and editorial principles. The *TEI* headers consist of four main parts: file description, encoding description, profile description and revision description. Each

of these parts may have their own internal structure. Granger (1998) claims that particular details about the attributes revealing the corpus design criteria should be recorded in the file header of the texts included in learner corpora. Therefore, pertaining tags were selected from all four parts of the *TEI* header. The first part, file description, is the mandatory part of the *TEI* header. It provides a full bibliographic description of a file. The file description tags that have been included in the files of the students' corpus are presented in *Figure: 23*.

```
<fileDesc>
        <titleStmt>
                <title>Title of the electronic text.</title>
                <author>The author of the electronic text.</author>
                <resp>
                        <role>Compiled by. </role>
                        <name> Name of Compiler</name>
                </resp>
        </titleStmt>
        <extent>Size of the electronic text(s)</extent>
        <publicationStmt>
                <publisher>Information of the publisher.</publisher>
                <pubPlace>Place of publication</pubPlace>
        </publicationStmt>
</fileDesc>
```

Figure 23: **File description tags**

The aim of these tags is to provide the source information of the texts that have been included in the corpus as well as denote the compiler and the size of the text included in the file. The tag referring to the title of the electronic text, discriminates among the simulation conference messages, or, for example, essays, whereas the tag of the author includes the students' codes that have developed these documents, which is followed by information about the corpus compiler and the size of the electronic texts included in the file. Further on, the file description includes the information of the text publisher and the place of publication, namely, in the case of individually developed texts, which have been embedded in the University hosted e-course, the publisher is considered to be *The University of Latvia* and the place of publication *Riga*.

The second part of the header presents the encoding description tags (Figure 24). Two encoding description tags have been chosen for the present study corpus. The first tag is meant to provide the aim of the compilation of the corpus, namely, the exploration of students' English language variation in electronic discourse. The second tag is essential in the present corpus, as it gives details about the editorial principles. The following editorial changes have been made within the texts. The texts have been spell-checked

using spell checker programme in order to eliminate such accidental misprints as the omission of a space between the words in online discussion postings. All end-of-line hyphens have been removed and the trailing part of a word has been joined to the preceding line.

```
<encodingDesc>
        <projectDesc>Information of the aim of the compilation of the electronic texts</projectDesc>
         <editorialDecl> Statement of editorial principles</editorialDesc>
</encodingDesc>
```

Figure 24: **Encoding description tag**

The third part of the header contains the profile description tags, which are displayed in *Figure 25*. The goal of the profile description tags in the present study is the provision of information about task setting. Namely, distinction between collaboratively/individually developed texts, whereas text description explicates the belonging of the texts to a particular register group: academic electronic texts, electronic letters and electronic messages. The participants' description contains information about the students who have participated in a particular communicative situation, namely, the codes of the students who have developed the electronic texts, thus reinforcing the distinction of the texts according to the academic years.

```
<profileDesc>
        <creation>Information on the creation of source text</creation>
        <textDesc>Description of a text in terms of situational paramethres.<textDesc>
        <particDesc>Description of the participants involved in the developmet of texts</particDesc>
</profileDesc>
```

Figure 25: **Profile description tags**

The final part of the header incorporates the revision description (Figure 26). The role of this description is to keep track of time and type of the changes that have been made within the present study corpus. This part of the header denotes the date of the change, the person who has made the change and the information of what particular changes have been made, for example, the information about part-of-speech (*POS*) annotation and the post-editing after the part-of-speech (*POS*) annotation.

```
<revisionDesc>
        <change>
                <date>The date of the change that has been made</date>
                <respStmt>
                        <resp>The role of the person who has made changes</resp>
                        <name>Name of the person who has made changes.</name>
                </respStmt>
                <item>Information of what changes have been introduced</item>
        </change>
</revisionDesc>
```

Figure 26: **Revision description tags**

Meyer (2002) suggests placing each text of a corpus within a separate file and creating file directories. However, the files of the present corpus contain more than one text, as the texts included in students' corpus are of various magnitudes ranging from 1 to 800 words. Accordingly, each file of the students' corpus contains a group of texts. For example, collaborative online simulation messages the project *IDEELS* developed during a particular academic year have been placed into one file.

2.1.5.2 Part-of-Speech (POS) Annotation

Linguistic features can be explored (McEnery, 2006; Meunier 1998; Meyer, 2002; Tribble, 2002), in two main corpora formats: raw and tagged corpora. A raw corpus contains machine-readable, structurally annotated *ASCII* texts without the extra annotation features added, whereas tagged corpus contains extra annotation features added to the texts comprising a corpus. Leech (1993: 275) defines corpus annotation in the following way:

> The practice of adding interpretative (especially linguistic) information to an existing corpus of spoken and/or written language by some kind of coding attached, or interspersed with, the electronic representation of the language material itself.

Although particular linguistic features can be studied in the raw corpus, for example, the type/token ratio, and word length, a more detailed investigation of the noun, personal pronoun, verb and other word class characteristics, can be explored only in the tagged corpus. Therefore, the raw corpus of the student-composed electronic discourse of the present study has been *POS* tagged by *CLAWS7 (the Constituent Likelihood Automatic Word-tagging System)* tagger (2007). This tagger has been developed at the

research centre of Lancaster University *UCREL* (University Centre for Computer Corpus Research).

The *CLAWS* has been trained on native corpora as it has been used to *POS* tag c. 100 million words of *British National Corpus (BNC)* and the tagged written texts of *BNC* have been widely used by researchers in their investigations. However, Meunier (1995, 1998) claims that native corpora trained taggers can be used on both native and more advanced non-native learner data without any specific adoption to learner language. Moreover, before *POS* tagging, the learner data of the present study have undergone the editorial changes explicated in the chapter *Structural Annotation*.

CLAWS tagger has been chosen for the annotation of the students' corpus, firstly, due to its comparatively easy access by using *UCREL* tagging service granted by *Lancaster University*. Secondly, and more importantly, it has been chosen after its exploration according to the tagger evaluation criteria put forward by Van Halteren (1999a, 1999b): the characteristics of the tagset, documentation, the tagging process and performance.

The exploration of the **characteristics of the tagset** reveals that *CLAWS* currently offers the tagsets of various refinements. Meunier (1998) emphasizes that the complexity of a tagset strongly determines the precision of further analysis of the learner data. If the required tags are not included in the system, the research process, as stated by Ball (1994), would be extremely laboured and might lead to low precision. *CLAWS* tagger currently offers the tagset *C5* that contains over 60 tags and the tagset *C7* that includes over 160 tags. Before *POS* annotation of the student-composed texts, the lists of both tagsets as well as the tagged samples of students' texts were explored and found that C7 tagset proposes a more detailed differentiation of the word classes that are vital for the present study. Thus, for example, as it is seen in *Figure 27*, which contains C7 tagged extract of student-composed email message, this tagset proposes differentiation of the personal pronouns. Namely, *PPISI* denotes first person singular pronoun *I*, but *PPH1* the pronouns *it*. Tagset C5, however, labels both these pronouns *PNP* without more subtle distinction. *Figure 29* presents a list of tags that have been used in the example included in *Figure 27*, whereas *Annex II* contains a full list of C7 tagset that has been used to tag the corpus of student-composed texts in the present study.

Apart from the comparatively refined tagset pertaining to the goal of the present research, the web site of *CLAWS* tagger provides detailed **documentation** about the tagsets, the tagging process and the assumptions underlying the various operations to

facilitate correct interpretation of the obtained results. *UCREL* of *Lancaster University* has developed *Wordclass Tagging Guidelines* that contain introduction to word classes, disambiguation guide by tag pair and disambiguation by word. These guidelines were explored and used in the interpretation of the results and disambiguation.

```
0000003 040 I                                          93 [PPIS1/100] ZZ1%/0 MC1%/0
0000003 050 know                                       93 VV0
0000003 051 ,                                          03 ,
0000003 060 it                                         93 PPH1
0000003 070 is                                         93 VBZ
0000003 080 possible                                   93 JJ
0000003 090 to                                         97 TO
0000003 100 use                                        97 VVI
0000003 110 personal                                   03 JJ
0000003 120 pronouns                                   93 NN2
0000003 130 in                                         93 [II/100] RP@/0
0000003 140 the                                        93 AT
0000003 150 essay                                      93 [NN1/100] VV0%/0
0000003 151 .                                          03 .
0000003 152 ------------------------------------------------------
0000003 160 But                                        96 CCB
0000003 170 the                                        93 AT
0000003 180 writer                                     93 NN1
0000003 190 should                                     93 VM
0000003 200 avoid                                      97 VVI
0000003 210 expressing                                 03 VVG
0000003 220 his                                        93 [APPGE/100] PPGE@/0
0000003 230 personal                                   03 JJ
0000003 240 opinion                                    03 NN1
0000003 250 too                                        93 [RG/99] RR@/1
0000003 260 straightforwardly                          06 RR
0000003 270 by                                         93 [II/78] RP%/22
0000003 271 ,                                          03 ,
0000003 280 for                                        96 REX21
0000003 290 example                                    96 REX22
0000003 291 ,                                          03 ,
0000003 300 writing                                    93 [VVG/71] NN1/28 JJ%/0
0000003 301 ,                                          03 ,
0000003 310 I                                          93 [PPIS1/100] ZZ1%/0 MC1%/0
0000003 320 guess                                      03 [VV0/100] NN1@/0
0000003 321 .                                          03 .
```

Figure 27: **A tagged unedited extract from the corpus of student-composed texts**

```
0000003 040 I                                          93 PPIS1
0000003 050 know                                       93 VV0
0000003 051 ,                                          03 ,
0000003 060 it                                         93 PPH1
0000003 070 is                                         93 VBZ
0000003 080 possible                                   93 JJ
0000003 090 to                                         97 TO
0000003 100 use                                        97 VVI
0000003 110 personal                                   03 JJ
0000003 120 pronouns                                   93 NN2
0000003 130 in                                         93 II
```

0000003 140 the	93 AT
0000003 150 essay	93 NN1
0000003 151 .	03 .
0000003 152 ----------	
0000003 160 But	96 CCB
0000003 170 the	93 AT
0000003 180 writer	93 NN1
0000003 190 should	93 VM
0000003 200 avoid	97 VVI
0000003 210 expressing	03 VVG
0000003 220 his	93 APPGE
0000003 230 personal	03 JJ
0000003 240 opinion	03 NN1
0000003 250 too	93 RG/99
0000003 260 straightforwardly	06 RR
0000003 270 by	93 II
0000003 271 ,	03 ,
0000003 280 for	96 REX21
0000003 290 example	96 REX22
0000003 291 ,	03 ,
0000003 300 writing	93 VVG
0000003 301 ,	03 ,
0000003 310 I	93 PPIS1
0000003 320 guess	03 VV0
0000003 321 .	03 .

Figure 28: **A tagged and edited extract from the corpus of student-composed texts**

APPGE	possessive pronoun
AT	article
AT1	singular article
CC	coordinating conjunction
CCB	adversative coordinating conjunction
CS	subordinating conjunction
CST	that
DAR	comparative after-determiner
DAT	superlative after-determiner
DD1	singular determiner
II	general preposition
JJ	general adjective
MC1	singular cardinal number
MD	ordinal number
NN	common noun, neutral for number
NN1	singular common noun
NN2	plural common noun
NNO2	numeral noun, plural
PPGE	nominal possessive personal pronoun
PPH1	3rd person sing. neuter personal pronoun
PPIS1	1st person sing. subjective personal pronoun
REX	adverb introducing appositional constructions
RG	degree adverb
RGR	comparative degree adverb
RGT	superlative degree adverb
RP	prep. adverb, particle
RR	general adverb
RRR	comparative general adverb
RRT	superlative general adverb
TO	infinitive marker
VBI	be, infinitive

VBZ	is
VM	modal auxiliary
VV0	base form of lexical verb
VVD	past tense of lexical verb
VVG	-ing participle of lexical verb
VVI	infinitive
VVN	past participle of lexical verb
VVZ	-s form of lexical verb
ZZ1	singular letter of the alphabet

Figure 29: **Tags used in the tagged extract in Figures 27 and 28**

The refined tagging **process** of *CLAWS* contributes to the **performance of the tagger**. The process and performance of *CLAWS* has been thoroughly considered by Leech, Gearside and Bryant (1994), Gearside (1996), Gearside, Smith (1997), who assert that this tagger has been constantly developed since the early 1980s (up to the version *CLAWS8*). *CLAWS* is a hybrid tagger that employs a mixture of probabilistic and non-probabilistic techniques by assigning tags during four main processes: tokenization, initial assignment of tags, disambiguation and idiom tagging.

In order to do the correct part-of-speech assignment during the tagging process, *CLAWS* has available a large database of words and their associated *POS*, as well as a set of heuristic algorithms which are used to predict the part-of-speech of words not included in the database. The heuristic algorithms are constantly updated and they include probabilistic models of adjacency tag checking (for example, if a particular word has an adjective to its left and a verb to its right, how far this word might have a chance to be a noun) and suffix lists detailing the *POS* of various words ending in particular suffixes. Overall, *CLAWS* boasts high (96-97%) accuracy for most written language texts. Rooy and Schaffer (2003) have evaluated three taggers (*Brill, CLAWS7* and *TOSCA-ICLE*) before starting with *POS* tagging on their learner English corpus in order to investigate which of these three taggers give the best results when tagging non-native English language. They have found that *CLAWS7*, according to the overall accuracy, showed the best (96.26% accuracy). However, they have also drawn attention to the fact that, apart from overall accuracy, it is essential to identify and, if necessary, post-edit the tags with lowest precision (according to these researchers for *CLAWS* the lowest precision tags are *RGR, RRR, DDQ*) and the tags that contribute most significantly to the overall error rate (*NN1, JJ, VVO, ND1*).

After the exploration and the evaluation of *CLAWS*, all student-composed texts included in the corpus were subjected to the four, previously considered, tagging processes in order to assign *C7* tags: tokenisation, initial assignment of tags,

disambiguation of tags and idiom tagging. However, before these four processes, all *ASCII* files of student-composed texts were run through *CLAWS* pre-edit programme, which made them machine readable. During tokenization corpus texts were divided into individual word tokens. Then, during tag assignment, each word was assigned one or more tags. The next step, idiom tagging, matched words and tags against a 'template' and depending on a match, the tags were disambiguated or corrected. When student-composed texts had undergone all four stages each word had one or more tags associated with it (see *Figure 27*). Each tag is associated with a probability represented as a percentage. An example of the tagged word *writing* from *Figure 27* is as follows:

writing VVG71%NN128%JJ0%

This example shows that VVG (-ing participle of the verb *write*) is considered by *CLAWS* to be the most likely tag in this particular case. The reference number, as is presented in *Figure 27*, at the start of each line shows which line of the input file a word comes from. Sentence breaks are identified by lines of hyphens. The two digit number to the left of the *POS* tags is a decision code produced by *CLAWS7* to aid manual post-editing. Each *POS* tag on an ambiguous word is followed by a slash and a likelihood value, expressed as a percentage. The final step, therefore, was the manual post-editing of each output file in order to disambiguate the tags according to *Wordclass Tagging quidelines*. For example, *Figure 28* contains the tagged text that has been disambiguated.

Thus, *POS* annotated texts of the corpus had been prepared for the exploration of the linguistic characteristics according to the chosen linguistic features with the help of the text retrieval programme *WordSmith Tools*. The choice of dimensions, linguistic features and their extraction process with the text retrieval programme *WordSmith Tools* as well as the variation of the linguistic features in the student-composed texts are considered in the next chapter.

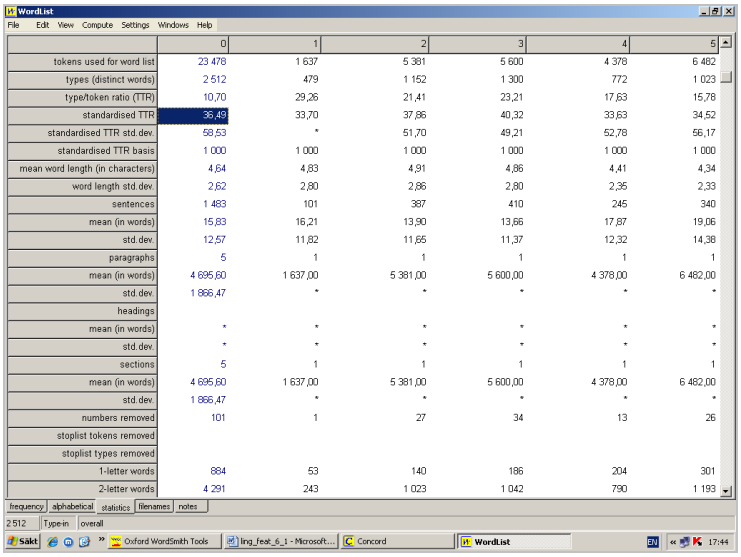
2.1.6 Retrieval of Linguistic Features

The extraction of the above-presented examples of 38 linguistic features and their frequency (see chapter 4.7 *Descriptive Statistics of Linguistic Features*) from the corpus of student-composed texts was performed with the help of the text retrieval programme *WordSmith* (Scott: 2004). This programme has been chosen for two reasons. Firstly, it

has been trained to process both untagged and tagged texts including *CLAWS7 POS* tagged texts. Secondly, *Word*

Smith has been implemented as a research tool in the investigation of the linguistic characteristics of native student-composed texts by such scholars as Tribble (2002), Scott and Tribble (2006), Xiao and McEnery (2005) and their research methodology has been illuminated by McEnery et al (2006). Within the framework of the present study two functions of *WordSmith* have been used: *Wordlist* and *Concord*.

Wordlist was implemented for the identification of average word length and standardized type/token ratio (*TTR*) of the texts arrangedaccording to task type. Firstly, *WordSmith Settings* and *Adjust Settings* were attended and *Tags to ignore* <*> were activated in order to ignore the metadata included in the angle brackets in the header of each file, then frequencies were normalized to 1,000 tokens. Next, *Wordlist* was started and all raw corpus files belonging to a particular register group were uploaded and processed. For example, Figure 30 shows the basic *Wordlist* statistics output of the files containing post-essays. This output includes average word lengths and standardized *TTR* that were manually entered in the *SPSS* spreadsheet to compute the mean, minimum/maximum frequency, range and standard deviation (consult chapter 4.7 for descriptive statistics).



	0	1	2	3	4	5
tokens used for word list	23 478	1 637	5 381	5 600	4 378	6 482
types (distinct words)	2 512	479	1 152	1 300	772	1 023
type/token ratio (TTR)	10,70	29,26	21,41	23,21	17,63	15,78
standardised TTR	36,49	33,70	37,86	40,32	33,63	34,52
standardised TTR std.dev	58,53	*	51,70	49,21	52,78	56,17
standardised TTR basis	1 000	1 000	1 000	1 000	1 000	1 000
mean word length (in characters)	4,64	4,83	4,91	4,86	4,41	4,34
word length std.dev	2,62	2,80	2,86	2,80	2,35	2,33
sentences	1 483	101	387	410	245	340
mean (in words)	15,83	16,21	13,90	13,66	17,87	19,06
std.dev	12,57	11,82	11,65	11,37	12,32	14,38
paragraphs	5	1	1	1	1	1
mean (in words)	4 695,60	1 637,00	5 381,00	5 600,00	4 378,00	6 482,00
std.dev	1 866,47	*	*	*	*	*
headings						
mean (in words)	*	*	*	*	*	*
std.dev	*	*	*	*	*	*
sections	5	1	1	1	1	1
mean (in words)	4 695,60	1 637,00	5 381,00	5 600,00	4 378,00	6 482,00
std.dev	1 866,47	*	*	*	*	*
numbers removed	101	1	27	34	13	26
stoplist tokens removed						
stoplist types removed						
1-letter words	884	53	140	186	204	301
2-letter words	4 291	243	1 023	1 042	790	1 193

Figure 30: ***Wordlis*t output**

The tool *Concord* was used to extract the frequencies of 36 linguistic features (apart from average word length and *TTR*), enumerated in chapter 4.5, from the *CLAWS7 POS* tagged texts of the corpus. The search patterns were based on McEnery et al (2006).

The linguistic features have been extracted by starting *Concord* and uploading *POS* tagged files. Then a query in a form of a particular search pattern, for example, the pattern for the identification of the frequency of analytic negations *XX, was entered in the search box (see Figure 31), then, as seen in Figure 32 the concordances of analytic negations were displayed. All lines were viewed and manually checked. Finally, the number of hits in each file was identified by pressing on *plot* (*Figure 33*). The obtained output also shows visually the dispersion of a particular linguistic feature across the files. In the case of more sophisticated search patterns, requiring the context tags, only the files containing the texts of a particular text group were uploaded at one attempt and the displayed concordances were manually validated.

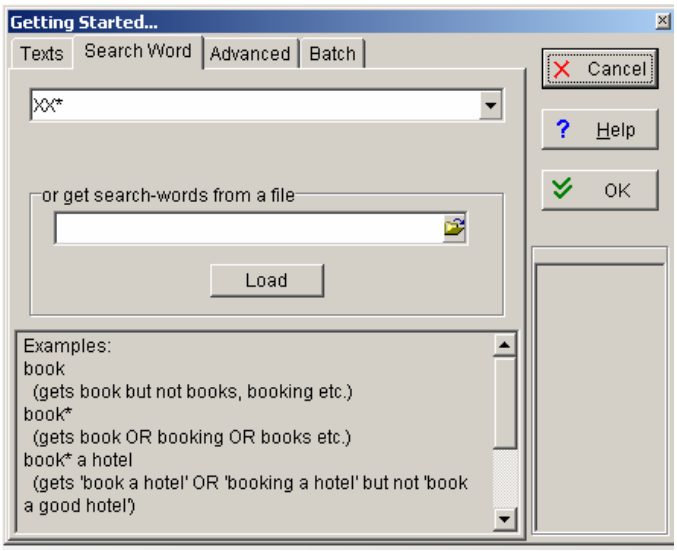


Figure 31: **The search pattern**

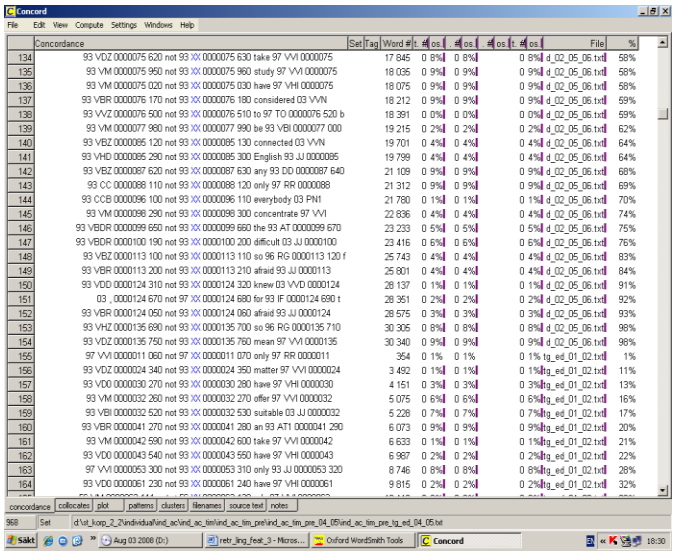


Figure 32: **The concordance of analytic negations**

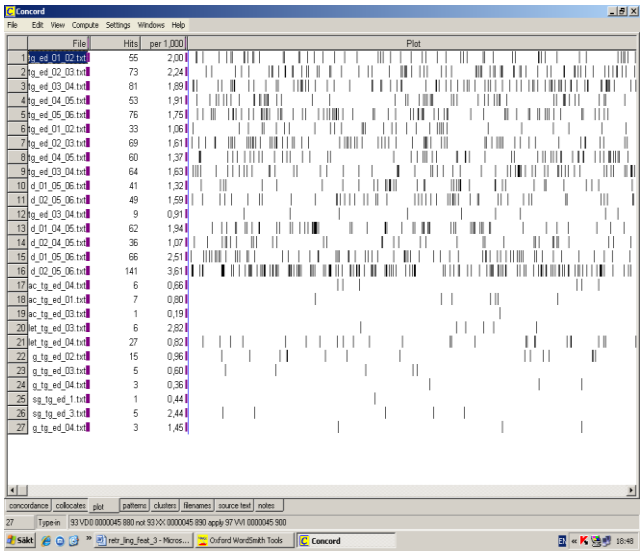


Figure 33: **The hits of analytic negations in files**

The obtained frequency counts were entered in the *SPSS* spreadsheet to compute the mean, minimum/maximum frequency, range and standard deviation (see chapter 4.7 for descriptive statistics) for the texts of a particular register groups and for *NNS* corpus.

2.1.7 Descriptive Statistics of Linguistic Features

NNSE corpus contains complete electronic texts of diverse magnitude. In order to obtain an accurate assessment of frequency distribution of linguistic features across the corpus and compare the obtained results with the previous research of linguists who have

explored linguistic variation the below-presented descriptive statistics for each linguistic feature was calculated.

1 **Normalized frequency:** all raw frequency counts were normalized to a text length of 1000 words. Normalized frequency f_i was formulated as

$$f_i = \frac{F_i}{N} * 1000$$

In the formula, F_i is raw frequency of a particular linguistic feature in a text, but N stands for the number of the words in the text.

2 **Mean frequency** of each linguistic feature has been obtained as the sum of normalized frequencies divided by a number of the texts *n*.

$$\overline{f} = \frac{\sum_{i=1}^{n} f_i}{n}$$

3 **Standard deviation** *s* is a measure of the dispersion of linguistic features within the compiled corpus. It is defined as the root-mean-square (*RMS*) deviation of the frequencies from their mean. Thus, standard deviations of the linguistic features were calculated as the square root of the sum of squared differences of normalized frequencies f_i and their arithmetical mean $\overline{f}$ divided by the number of the texts *n – 1*.

$$s = \sqrt{\frac{\sum_{i=1}^{n} (f_i - \overline{f})^2}{n-1}}$$

4 Difference or **range** *R* between the maximum and minimum frequency of a particular linguistic feature was identified in the following way:

$$R = f_{\max} - f_{\min}$$

The above-presented descriptive statistics of *NNS* corpus is shown in *Table 11*. The descriptive statistics for each student-composed text group included in the corpus is presented in *Annex V*. These calculations were performed with the help of *Microsoft Office Excel* and *SPSS-15.0*.

Table 11: **Descriptive statistics of linguistic features for the whole *NNSE* corpus**

Dimensions	Features	Mean	Std dev.	Min.	Max.	Range
A Involved versus informational production	+ (1) Private verbs	15.75	5.39	9.62	26.95	17.31
	+ (2) Contracted forms	3.85	4.07	0	15.27	15.27
	+ (3) Present tense v.	76.20	19.79	40.15	125.95	85.80
	+ (4) Second person pr.	15.34	13.82	0	46.45	46.45
	+ (5) Do as pro-verb	1.41	1.71	0	7.60	7.60
	+ (6) Analytic negation	9.32	5.02	3.24	21.75	18.52
	+ (7) Demonstrative pr.	4.29	3.30	0.61	11.42	10.81
	+ (8) General emphatics	6.37	3.82	0.51	14.28	13.77
	+ (9) First person pr.	32.97	23.58	2.21	79.85	77.64
	+ (10) Pronoun *it*	12.38	5.20	3.24	21.46	18.22
	+ (11) *Be* as main verb	25.73	20.40	3.89	87.45	83.56
	+ (12) Causative sub.	2.26	1.95	0	7.14	7.14
	+ (13) Discourse markers	0.77	1.68	0	7.14	7.14
	+ (14) Indefinite pr.	4.97	2.97	0	11.41	11.41
	+ (15) Hedges	1.30	1.13	0	3.82	3.82
	+ (16) Amplifiers	4.59	2.58	0	10.99	10.99
	+ (17) Sentence relatives	1.33	1.04	0	3.09	3.09
	+ (18) WH questions	3.75	7.91	0	34.35	34.35
	+ (19) Possibility modals	8.66	3.12	2.16	15.27	13.11
	+ (20) Non-phrasal coord.	11.29	3.99	6.14	20.46	14.31
	+ (21) WH clauses	1.56	1.19	0	4.28	4.28
	+ (22) Final prepositions	1.15	3.12	0	14.29	14.29
	- (23) Other nouns	199.13	31.32	151.95	290.08	138.13
	- (24) Word length	4.81	0.32	4.30	5.44	1.14
	- (25) Prepositions	86.55	21.40	51.40	131.73	80.33
	- (26) Type/token ratio	35.03	3.17	30.14	42.50	12.36
B Explicit versus situation dependent reference	+ (27) WH relative clauses	2.11	1.26	0	4.54	4.54
	+ (28) Pied piping constr.	0.87	0.76	0	3.25	3.25
	+ (29) Phrasal coordin.	17.94	9.81	3.57	41.89	38.32
	+ (30) Nominalizations	43.47	24.66	7.60	102.71	95.11
	- (31) Time adverbs	3.52	2.17	0	7.92	7.92
	- (32) Place adverbs	2.74	1.79	0.52	7.63	7.11
	- (33) Other adverbs	90.55	27.97	34.52	124.04	89.52
C Abstract versus nonabstract information	+ (34) Conjuncts	5.02	3.14	0	10.76	10.76
	+ (35) Agentless passives	9.85	5.40	0	20.50	20.50
	+ (36) BY-passives	0,78	0.87	0	3.80	3.80
	+ (37) Past part. WHIZ d.	2.48	2.56	0	9.71	9.71
	+ (38) Other adv. sub.	0.66	0.60	0	2.35	2.35

According to *Table 11*, some linguistic features distribute considerably evenly across the corpus, for example, *sentence relatives, hedges, by-passives* and *other adverbial subordinators*. However, there are linguistic features that show appreciable difference in dispersion, for example, first person pronouns, other nouns, prepositions, nominalizations and other adverbs. The dispersion variation of the explored linguistic features in the corpus of student-composed texts prove that the use of them can also show significant differences across the electronic texts arranged in groups according to task type. Therefore, within the framework of the present study a more detailed comparison of the texts arranged into groups has been performed by computing dimension scores *(Dimension A, Dimensions B, Dimension C)* for each text and text group. The statistics in *Table 11* has been used to compute the said dimension scores.

SUMMARY

The explicated research methodology and procedure aims at the implementation of technology-enhanced collaborative and individual writing process model in the English language studies, the compilation of student-composed texts in a corpus and the detection of the frequency of linguistic features in students' electronic texts. The implementation of the model has envisaged the exploration of the project *IDEELS* e-environment and the design of e-course *English Academic Writing III*, which have been used by students during their collaborative and individual writing of electronic texts. Student-composed texts have been compiled in electronically transferrable, structurally and part-of-speech annotated corpus in which the frequency of the selected linguistic features has been detected. The obtained frequency counts and their descriptive statistics across the whole corpus as well as collaboratively and individually developed texts enable the analysis of student-composed electronic texts as discourse manifestations.

2.2 Analysis of Student-composed Electronic Texts as Discourse Manifestations

The chapter is devoted to the *MDA* of the linguistic features in the corpus of student-composed electronic texts. The purpose of this analysis is to reveal the relations among variation of lexico-grammatical features in -s' corpus of electronic texts (Figure 22, see Chapter 2.1.2*Corpus of Student-composed Texts*) and compare the obtained results with the previous research in linguistics, namely, Biber's (1988) *MDA* of the texts of General English corpus of authentic texts. Further in the chapter his investigation is referred to as Biber's study.

The linguistic variation have been revealed by the descriptive statistics calculated for the whole corpus (Table 11, see Chapter 2.1.6 *Descriptive Statistics of Linguistic Features*). These results enabled the calculation of *Dimension A, Dimension B* and *Dimension C* scores for each text group. In order to identify the differences among student-composed electronic texts grouped according to task types and compare them with Biber's study, the dimension score of each text group has been obtained by adding the scores of all linguistic features with positive weights and then subtracting the scores of all features with negative weights. For example, the score of Dimension *B* for pre-essays (see Table 16) equals the sum of scores of all features of this text group labelled with the positive sign (+) minus the total of factor scores of the features that appear under the negative sign (-). Thus, for pre-essays the four features with positive weights on *Dimension B* are +0.18, -0.37, +0.83, -0.23, whereas three features with negative weights are +0.64, -0.54, +0.86. *Dimension B* score for this text group accordingly is -0.55. The score of a linguistic feature in a text group is the sum of the scores of that feature in each text divided by the number of the texts nevertheless some texts do not contain such linguistic feature. The score *k* for each linguistic feature in each text has been calculated as follows:

$$\kappa = \frac{f_i - \mu}{\sigma}$$

In the above-presented formula, f_i is the normalized frequency of a particular linguistic feature in a text, μ is the mean frequency of this feature in the corpus, but σ stands for standard deviation of that linguistic feature in the corpus. The statistical significance of varitions of dimension scores across the text groups has been identified by

the implementation of the Pearson's correlation the results of which show that the obtained p values (p = 0.01) are smaller than 0.05.

The frequency of linguistic features across the text groups and their functional analysis is presented according to each dimension. The titles of the text groups from the corpus of student-composed texts have been abbreviated and labelled in the tables and figures of the present paper as follows:

1 collaborative statements	c. statements
2 collaborative e-letters	c. e-letters
3 collaborative messages of simulation teleconferences	c. sim. messages
4 collaborative messages of post-simulation teleconferences	c. post-sim. messages
5 individual messages of asynchronous discussions	ind. messages
6 academic pre-essays	pre-essays
7 academic post-essays	post-essays

2.2.1 Dimension A: involved versus information production

Involved versus informational production is *Dimension 1* or *A* as labelled by Biber and Finegan (1989) and presented in their study. The linguistic features clustered within this dimension distinguish between the discourse that possesses high informational density and exact content versus affective and interactional discourse (see Biber 1988, 104-108). This dimension comprises 28 linguistic features 26 of which have been chosen for the present investigation. The first 22 linguistic features possess positive loadings, but the final 4 possess negative loadings. A more frequent use of all linguistic features with positive weight on *Dimension A* reveal interactional, non-informational focus of discourse, whereas more frequent features with negative weight mark transactional and informational characteristics of discourse. Biber (1988) has proved that more interactional and less formal texts display higher scores, whereas more transactional and informative texts present lower scores in *Dimension A.* Table 12 and Table 13 display dimension scores of linguistic features in collaboratively and individually written texts by students and the total dimension scores for each text group.

Table 12: **Dimension A (Involved versus informational production) scores of linguistic features of collaboratively written texts**

Linguistic features	Statements	E-letters	Sim. messages	Post-sim. messages
+ (1) Private verbs	-8.85	-0.45	-0.62	+3.96
+ (2) Contracted forms	-0.95	-0.57	-0.5	+1.8

+ (3) Present tense v.	-1.58	-1.46	+0.1	+1.63
+ (4) Second person pr.	-1.11	+0.9	-0.08	+1.78
+ (5) Do as a pro-verb	-0.7	+0.08	-0.37	+2.1
+ (6) Analytic negation	-1.17	+0.24	-0.97	+0.36
+ (7) Demonstrative pr.	+1.06	-0.24	-0.55	+0.91
+ (8) General emphatics	-0.98	-1.24	-1.15	+1.24
+ (9) First person pr.	-1.05	-0.08	+0.46	+1.5
+ (10) Pronoun *it*	-1.51	-0.97	-0.74	-0.47
+ (11) *Be* as a main verb	-1.05	-0.49	-0.04	+2.52
+ (12) Causative sub.	-1.05	-1.12	-0.96	+0.71
+ (13) Discourse markers	-0.46	-0.46	0	+2.47
+ (14) Indefinite pr.	-1.48	-1.04	-0.99	+1.26
+ (15) Hedges	-1.15	+0.2	-0.59	+2.15
+ (16) Amplifiers	-1.78	-0.59	-0.27	-0.3
+ (17) Sentence relatives	-0.86	+0.53	-0.92	-1.28
+ (18) WH questions	-0.47	-0.46	+0.04	+2.53
+ (19) Possibility modals	-1.65	-0.91	-0.04	+0.84
+ (20) Non-phrasal coord.	-0.58	+0.38	-0.92	-0.04
+ (21) WH clauses	-0.83	-0.02	-0.23	-1.31
+ (22) Final prepositions	-0.37	-0.37	-0.32	+2.38
- (23) Other nouns	+1.15	+0.43	-0.06	+1.12
- (24) Word length	+1.80	+0.64	+0.54	-1.1
- (25) Prepositions	+1.34	+0.9	+0.51	-0.64
- (26) Type/token ratio	-1.16	+0.27	+1.33	-
Score of *Dimension A*	**-31.70**	**-8.14**	**-11.98**	**+27.36**

Table 13: ***Dimension A*** **(Involved versus Informational Production) scores of linguistic features of individually written texts**

Linguistic features	Pre-essays	Post-essays	Messages
+ (1) Private verbs	+0.90	-0.69	+0.68
+ (2) Contracted forms	+0.27	-0.81	+0.54
+ (3) Present tense v.	+0.54	-0.1	+0.01
+ (4) Second person pr.	+0.04	-0.78	+0.02
+ (5) Do as a pro-verb	-0,23	-0.77	+0.35
+ (6) Analytic negation	+0,68	-0.26	+0.54
+ (7) Demonstrative pr.	+1.60	-0.48	-0.61
+ (8) General emphatics	+0.72	+0.77	-0.44
+ (9) First person pr.	+0.56	-1.16	+0.08
+ (10) Pronoun *it*	+1.12	-0.14	+1.04
+ (11) *Be* as a main verb	-0.03	-0.15	-0.48
+ (12) Causative sub.	+0.85	+0.12	+0.26
+ (13) Discourse markers	-0.39	-0.36	-0.03
+ (14) Indefinite pr.	+0.82	+0.38	-0.06
+ (15) Hedges	-0.44	-0.27	+0.27

+ (16) Amplifiers	+0.37	+0.05	+0.97
+ (17) Sentence relatives	+1.43	+0.30	-0.08
+ (18) WH questions	-0.39	-0.43	+0.33
+ (19) Possibility modals	+0.49	+0.36	+0.06
+ (20) Non-phrasal coord.	+0.86	+0.15	-0.22
+ (21) WH clauses	+0.67	-0.38	-2.1
+ (22) Final prepositions	-0.2	-0.34	-0.32
- (23) Other nouns	-1.01	+0.09	-0.57
- (24) Word length	-0.78	+0.19	-0.22
- (25) Prepositions	-1.38	+0.64	-0.35
- (26) Type/token ratio	-0,63	-0.10	+0.27
Score of *Dimension A*	**+14.06**	**-4.17**	**+1.68**

In addition, Figure 34 shows the variation of dimension scores of student-composed texts in comparison with the dimension scores obtained by Biber in his study. Collaborative messages of post-simulation conferences (+27.36) hold the highest position in this dimension and, in comparison with Biber's (1988) findings, they are more involved than personal letters; however, less involved than face-to-face conversation. The next text group showing a comparatively high score in *Dimension A* are pre-essays (+14.06). However, these texts, according to their communicative purpose, should have been considerably less involved, more transactional and correlate with the register score of academic prose in relation to Biber's (1988) findings. Individual messages (+1.68) of asynchronous discussions, rank lower than the personal letters in Biber's study explored texts in this dimension. These results confirm that students' individual messages are more transactional than personal letters and these messages correlate with professional letters in Biber's investigation (ibid.). Post-essays, which students had written after their participation in collaborative and individual virtual writing activities, display considerably lower dimension score than pre-essays, which means that in this dimension they are more transactional than pre-essays and are closer to academic prose in Biber's study. Collaboratively written e-letters (-8.14) in this dimension are even more transactional than professional letters in Biber's (1988) study and collaborative messages of simulation conferences (-11.98) they are the closest to academic prose in Biber's study. Although collaborative messages of simulation conferences have been produced during virtual conferences, these texts are informational. Collaborative statements (-31.70) are the most information and thus transactional texts developed by students.

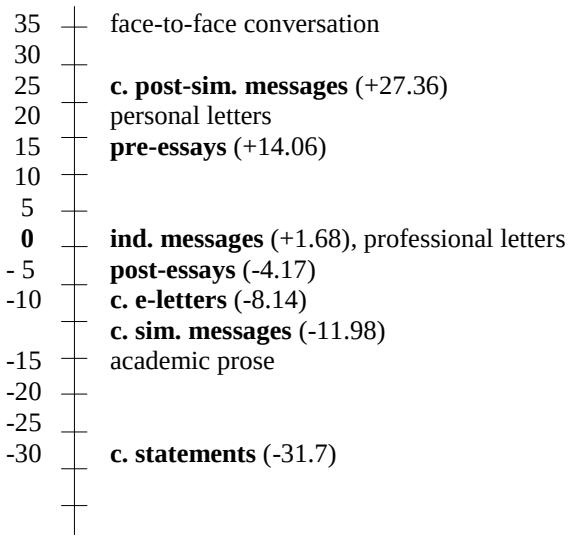


Figure 34: **Dimension A Involved versus Informational production for student-composed texts (in bold) grouped according to task types in comparison with the texts grouped in registers by Biber (1988)**

In order to reveal the linguistic features that account for the above-presented dimension scores (Table 12, Table 13, Figure 34) as well as identify their possible overuse or underuse in particular electronic text groups, the frequency of the selected linguistic features with positive and negative loadings are analysed in detail. This analysis is related to the situational components in which the texts have been composed by the students. The frequency of linguistic features in student-composed electronic texts has been compared with the previous linguistic research, namely, Biber's (1988) findings in relation to the texts arranged in four register groups face-to-face conversation, personal letters, professional letters, academic prose.

Private verbs

Private verbs, also labelled by Biber (1988) as cognition verbs or by Biber et al. (2004) mental verbs that describe mental activities: express intellectual states (e.g. *believe, fear*) or non-observable intellectual acts (e.g. *determine, discover*). Private verbs are used for the overt expression of attitudes, thoughts and emotions, and therefore are predominantly used in more interactional discourse. The examples of private verbs from corpus are presented in *Examples 1*. The overview of the examples of all linguistic features explored in the corpus of *NNSE* is included in *Appendix II*.

Examples 1

I **hope** I have shown you the main points. (pre-essays)
We **believe** we can turn them into benefit (post-sim. messages)
Bardland **thinks** that each person must have an access to all data concerning himself. (sim. messages)

The messages of post-simulation conferences (37.13), pre-essays (20.70) and individual messages (19.40) display the greatest frequency of private verbs (Figure 35). According to their frequency, pre-essays and individual messages are the closest to professional letters, whereas the post-simulation conference messages are the closest to such interactional discourse as face-to-face conversation (35.40) in Biber's study. The above-presented frequency in post-simulation conferences and individual messages corresponds the purpose of these texts: free sharing of personal cognition and in the case of the post-simulation conferences even sharing of emotions. However, the frequency of private verbs in pre-essays is too high to fully correlate with the communicative purpose of academic essays. In contrast, the number of private verbs in post-essays (12.20) is close to their frequency in academic prose (12.50) in Biber's study. Surprisingly, simulation messages (12.39) contain nearly the same frequency of private verbs as post-essays and Biber's investigated academic prose, which means that students' have recognized the peculiarities of the communicative purpose of these texts, although they have been written during synchronous virtual communication using chat messaging system. As it is seen in Figure 35, the lowest frequency of private verbs is found in statements (5.76), which are the most informational texts according to their communicative purpose. The frequency of private verbs in statements is even lower than in academic prose investigated by Biber (1988). These findings show that students have relevantly correlated the use of private verbs in their electronic texts according to the communicative purpose of these texts apart from one electronic text group – pre-essays, which were written before the students had participated in writing and analysis of electronic texts.

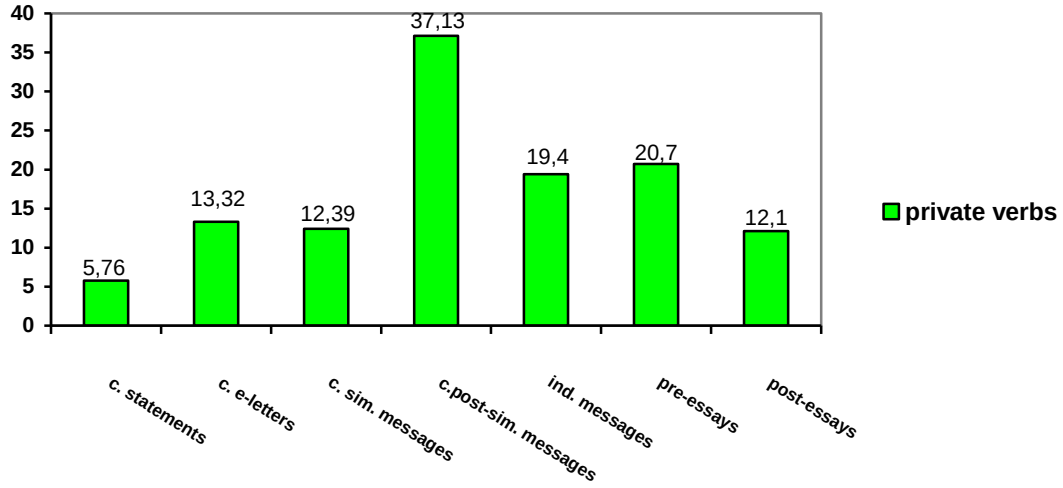


Figure 35: **Number of private verbs (per 1000 words) in the corpus of student-composed electronic texts**

First and second person pronouns

First-person pronouns (*I. my. our. myself. ourselves. mine. ours*). as Biber et al. (2004) point out. are markers of involvement in discourse. because they refer directly to the addressor and the addressee thus indicating an interpersonal focus and an involved style. Second-person pronouns (y*ou. your. yourself. yourselves. yours*) also require a particular addressee and indicate involvement with that addressee in discourse. All these pronouns referring directly to the addressor and addressee are more frequently used in more involved discourse. Examples of pronouns from student-composed texts are included in *Examples 2.*

Examples 2

> **We** agree with Northland's amendment. (sim. konferenču ziņojumi)
> **We** consider the integration one of the most essential issues concerning immigration. (sim. konf. Ziņojumi)
> **Your** idea of dividing the responsibilities of the agency is really very useful as it could make the data operation process more accurate and elaborate. (e-vēstules)
> **You** did a great job! (pēcsim. konf. ziņojumi)
> Of course **we** do! (pēcsim. konf. ziņojumi)
> But we are **FOR SURE** more informed about this subject now. (pēc-sim. konf. ziņojumi)

The highest frequency of first and second-person pronouns (see Figure 36 where the mentioned pronouns are denoted PRO 1 and PRO 2) is in the collaborative post-simulation messages (first-person pronouns 68.23. second-person pronouns 39.92). the purpose of which was to share the opinions and insights among the simulation

participants – students of tertiary institutions. Their frequency is even higher than in face-to-face conversation (first-person pronouns 57.90. second person pronouns 30.80) in Biber's study. The next, according to their frequency, rank collaborative simulation messages (first-person pronouns 31.26, second-person pronouns 27.76) and pre-essays (first-person pronouns 46.35, second-person pronouns 5.94), which, according to the frequency of the first-person pronouns, can be compared with professional letters (40.90) in Biber's study. Collaborative simulation messages (first-person pronouns 43.82, second person pronouns 4.29) that have been exchanged during teleconferences, namely, synchronous online communication. are considerably less involved than post-simulation messages, although both simulation and post-simulation teleconferences were enabled by the implementation of the same online communication tools. However, the frequency of first-person (48.35) and second-person (5.94) pronouns is irrelevantly high as pre-essays were supposed to be informational, uninvolved texts. The frequency of first-person pronouns in e-letters (31.16) and individual messages (34.85) is lower than in Biber's study in professional letters (first-person pronouns 40.90. second person-pronouns 15.20), whereas the frequency of second-person pronouns is considerably higher in e-letters (27.76), which indicates a greater involvement with addressees during the exchange of these letters. Collaborative statements and post-essays display the lowest frequency of the mentioned pronouns and with respect to them correlate with academic prose in Biber's study. However, the frequency of first-person pronouns in statements (8.13) is a bit higher than in Biber's academic prose (5.70), which can be explained by the repeated use of the plural first-person pronoun *we* denoting in the statements the group of the representatives of the simulated country belonging to Eutropian Federation. Post-essays exhibit higher frequency of second-person pronouns (4.57) than academic prose (0.20) in Biber's study. These results imply that students have occasionally used in their post-essays direct addressing in a form of the personal pronoun *you*, which, however, is considerably less frequent in post-essays than pre-essays. The findings of first and second-person pronouns in student-compposed electronic texts show that they have correlated the use of these pronouns with the communicative purpose of electronic texts. First and second-person pronouns are substantially overused only in students' pre-essays, which were written before their participation in collaborative and individual writing activities. First and second-person pronouns are more relevantly used in post-essays than pre-essays although the students tend slightly overuse Second-person pronouns in them. It shows that students

might experience difficulties in the recognition of the virtual academic essays to be the texts that are an uninvolved with the addressee.

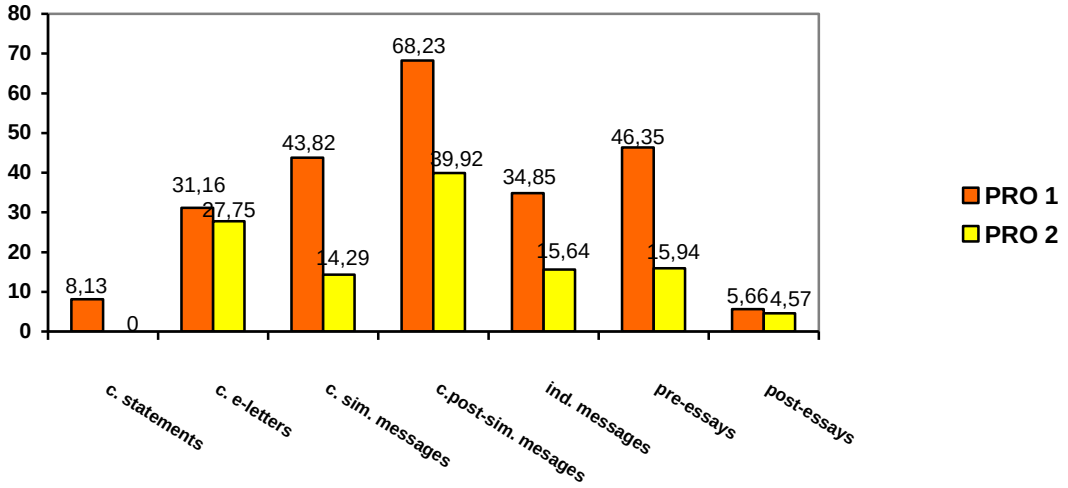


Figure 36: **Number of first-person and second-person pronouns (per 1000 words) in the corpus of student-composed electronic texts**

Pronoun it, demonstrative and indefinite pronouns

Pronoun *it* substitutes for a fuller noun phrase and Biber considers (1988) that frequent use of pronoun *it* marks a relatively inexplicit content that is connected with these time constraints as well as non-informational focus of a text produced in a more interactional communicative context. Demonstrative pronouns (*this. that. these. those*) not followed by a noun) can refer to previously denoted referent within a text or unspecified nominal referents outside a text, which can be a specific nominal entity or an abstract concept (e.g. *this shows*…). Halliday and Hasan (1976) argue that demonstrative pronouns are important devices for achieving discourse cohesion. Indefinite pronouns, like pronoun *it* and the demonstrative pronouns, are markers of generalized pronominal reference. Examples of pronoun *it*, demonstrative and indefinite pronouns are given in *Examples 3*.

Examples 3

We mean we have to arrange **it** in logical order. (sim. messages)
Yes. **it** would be a brilliant idea. (sim. messages)
That's terrible! (pre-essays)
By **this** I mean formal and informal writing. (pre-essays)
There are specialists who take care of **that**. (sim.messages)
This is a part of out Recommendation Paper. (sim. messages)
Why can't we add **this**? (sim. messages)
We noticed **that**! (post sim. messages)
Everything depends on a person. (post-essays)

Everyone has a right to know what information is available about him/her and where it is used. (sim. messages)
Nobody can say that one method is better than the other. (post-essays)
Good morning **everybody**! (sim. messages)
We had technical difficulties. but it seems that **everything** is ok now. (e-letters)

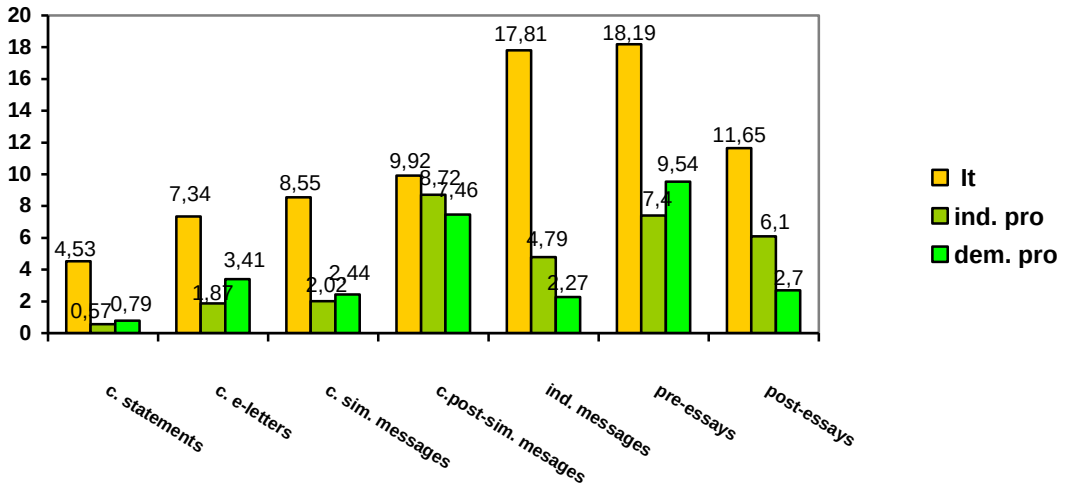


Figure 37: **Number of pronoun *it*. demonstrative and indefinite pronouns (per 1000 words) in the corpus of student-composed electronic texts**

Pronoun *It*. indefinite and demonstrative pronouns (see Figure 37. where the said pronouns are denoted in the following way: It, ind. pro., dem. pro) occur with the highest frequency in pre-essays, which shows non-informational focus of these essays. The frequency of the mentioned pronouns in pre-essays (pronoun *It* 18.19. indefinite pronouns 7.40. demonstrative pronouns 9.54) is even considerably higher than in post-simulation messages (pronoun *It* 9.92. indefinite pronouns 8.72. demonstrative pronouns 7.46). simulation messages (pronoun *It* 8.55. indefinite pronouns 2.02. demonstrative pronouns 2.44). and individual messages (pronoun *It* 17.81. indefinite pronouns 4.79. demonstrative pronouns 2.27) that, according to their communicative purpose, are more involved texts than pre-essays. In pre-essays, however, the frequency of pronoun *it* (18.19) and demonstrative pronouns (9.54) are even close to their use in face-to-face conversation (pronoun *It* 20.00. demonstrative pronouns 13.10) in Biber's study. As to collaborative e-letters, the use of the mentioned pronouns (pronoun *It* 7.34. indefinite pronouns 1.87. demonstrative pronouns 3.41) tends to correlate with their use in professional letters (pronoun *It* 7.10. indefinite pronouns 1.10. demonstrative pronouns 2.40) according to Biber's study. The obtained results show that students have carefully explored the communicative context of simulation discourse, and therefore have produced considerably informative texts according to the use of the said pronouns during

simulations even if the texts of this discourse were produced by the application synchronous and asynchronous online communication tools. The obtained results reveal that students have considerably reduced their use of the pronouns under discussion in post-essays, which, however, tend to be a bit overused in these texts.

Contracted forms and final prepositions

Contracted forms or contractions, explored in the present study (*n't. 'll. 'd. 're. 've. 's. 'm*), are verb contractions with the verbs *be* and *have*, modal verbs, *will* and *would* as well as negative or not-contractions when *not* is reduced and attached to a preceding verb. Biber (1988) defines contracted forms as reduced surface structured form and emphasises that they usually are not preferred in edited formal writing. These contractions substitute for a fuller verb phrase and are a characteristic feature of less formal, interactional texts. The use of contracted forms results in a more generalized, uncertain content. For example, *its* can be used to express the following: *it is, it has,* or *it*-possessive. Final prepositions, as pointed out by Biber (ibid.) are not followed by any compliments and their use results in a more generalized, uncertain content denoting a surface disruption in form-meaning correspondence. See *Examples 4* that contain sentences from the corpus of student-composed texts. The distribution of contracted forms and final prepositions in the investigated corpus is shown in Figure 38 (where the said linguistic features are denoted in the following way: contr. forms. fin. prep).

Examples 4

I **can't** do everything I want. (pre-essays)
That's what we meant. (sim. messages)
Let's keep in touch. (sim. messages)
All the topics were quite interesting because it is actually the everyday problems that affect all the countries. **aren't** they! (post-sim)
That is what I am thinking **about**. (pre-essays)
People who enter the country illegally should be returned to the country they come **from**.(sim. messages)
Now I know were everybody is **from.** (post sim
messages)

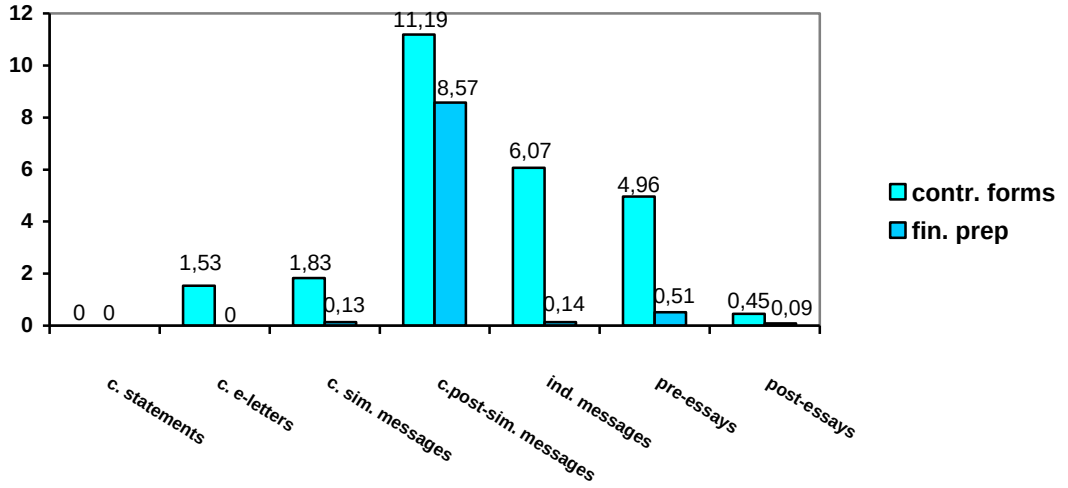


Figure 38: **Number of contracted forms and final prepositions (per 1000 words) in the corpus of student-composed electronic texts**

The number of contracted forms in two text groups (statements and post-essays) is similar with their frequency in academic prose (0.10) in Biber's study. Surprisingly that even simulation messages (1.83) written by students during virtual conferences with the respect to the frequency of contracted forms are closer to such informational texts investigated by Biber as academic prose than professional letters (4.70). These results confirm that the serious agenda of simulation conferences has encouraged the students to refrain from frequent use of contracted forms. Moreover, the conference messages contain the extracts from the most informational texts – statements. A more frequent use of contracted forms is found in individual messages and pre-essays. The frequency of them is close to their number in professional letters in Biber's study. As to final prepositions, all text groups, apart from simulation messages and post-simulation messages, contain insignificant number of them. Thus, post-simulation messages show marked frequency of final prepositions (8.57) as well as contracted forms (11.19), which confirms that post-simulation messages are the most interactional of student-composed texts in the corpus. Moreover, the frequency of these linguistic features is considerably reduced in the post-essays in comparison with the pre-essays. These results indicate that students have varied their use of contracted forms and final prepositions according to the purpose of the electronic texts.

Wh clauses, sentence relatives, causative subordination and non-phrasal coordination

The subordination features within *Dimension A. Wh* clauses, sentence relatives and causative subordination. as noted by Biber (1988), are associated with a relatively loose presentation of information and marks affective functions that relate to the elaboration of personal attitudes or feelings characteristic for a more transactional discourse. Biber (ibid.) points out that non-phrasal coordination can make various logical relations between clauses. Chafe (1985) considers that independent coordination indicates fragmentation of discourse, which is caused by the simple chaining of idea units. Non-phrasal coordination also is associated with fragmented presentation of information that results in a low informational density, so it is predominantly used in interactional and less formal discourse. The examples of all four linguistic features from student-composed electronic discourse are presented in *Examples 5*.

Examples 5

I didn't know **what to do with my independence**. (pre-essays)
I remember **what it looked like**. (pre-essays)
Would you please specify **what you mean by this question**? (sim. messages)
Tell us **what you think**. (e-letters)
The last point, **which is also important**. is the safety. (sim. messages)
This is a wrong point of view **because** we have never neglected the importance of human rights. (sim. messages)
We were 20 people and it was easier to get things done **because** we could share the tasks. (post-sim. messages)
We write letters to each other. **and we can not imagine our life without a mobile phone.** (pre-essays)
We are from Latvia **and we think that France is from Norway.** (post -sim. messages)
It was great **and we gained new experience!** (post-sim. messages)
Thank you for your greeting message. **and we are also looking forward to a successful negation process**. (e-letters)

The frequency of the use of *Wh* clauses and sentence relatives in student-composed electronic texts reveal curious results (Figure 39): post-simulation messages do not exhibit *Wh* clauses and sentence relatives although, according to all other linguistic features investigated in *Dimension A.* post-simulation messages are the most interactional texts. The absence of these subordination features can be explained by the fact that post-simulation messages predominantly contain simple sentences. The mentioned subordination features are the most frequent in e-letters (*Wh* clause 1.58. sentence relatives 1.87), individual messages (*Wh* clause 2.93. sentence relatives 1.25) and pre-essays (*Wh* clause 2.34. sentence relatives 2.02). They are even more frequent in these

texts than in face-to-face conversation (*Wh* clause 1.30. sentence relatives 0.70) found by Biber in his study. Another finding refers to the statements that being the most transactional texts according to all other linguistic features in *Dimenasion A,* contain *Wh* clauses (1.57) and sentences relatives (0.43) the number of which, however, is insignificant. Although in post-essays *Wh* clauses and sentence relatives are less frequent than in pre-essays, they are considerably overused in post-essays in comparison with their frequency in academic prose (*Wh* clauses 0.30. sentence relatives 0.0) in Biber's study.

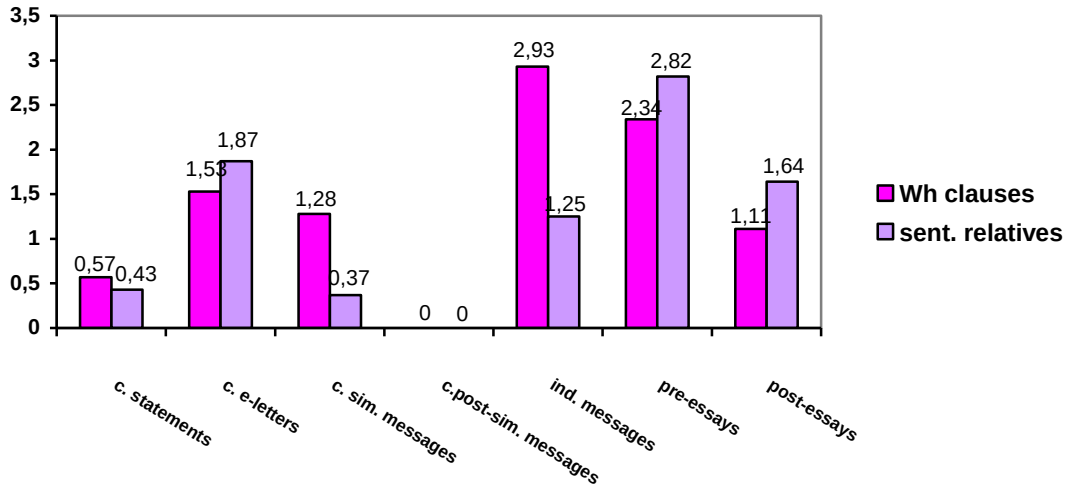


Figure 39: **Number of *Wh* clauses and sentence relatives (per 1000 words) in the corpus of student-composed electronic texts**

These findings show that students tend to experience the difficulties of the correlation of functions of the said two linguistic features with the communicative purpose of the transactional texts.

Causative subordination includes *because,* since only this causative adverbial functions unambiguously. Other forms, for example, *as, for, since,* can perform a range of other functions alongside with the causative function. Biber (1988) emphasizes that researchers have found a comparatively greater frequency of this causative adverbial in more interactional discourse. as it marks affective functions relating to the elaboration of personal attitudes or feelings. As it is seen in Figure 40 (where the linguistic features are denoted as follows: causative sub. non-phrasal coordination). the variation of causative subordination in student-composed electronic texts reveal that this linguistic feature is the most infrequent in statements (0.21), e-letters (0.08) and simulation messages (0.39). As a result, these texts according to causal subordination correlate with academic prose (0.30) in Biber's study. The most frequently causative subordination is used in students' pre-

essays (3.90), post-simulation messages (3.65) and individual messages (2.76). These groups of texts in relation to the frequency of causative subordination correlate with their frequency in face-to-face conversation (3.54) in Biber's (1988) investigation. Post-essays display considerable less frequent use of causative subordination (2.50) than pre-essays (3.95), but according to the communicative purpose of these texts they would require even lower frequency of these linguistic features.

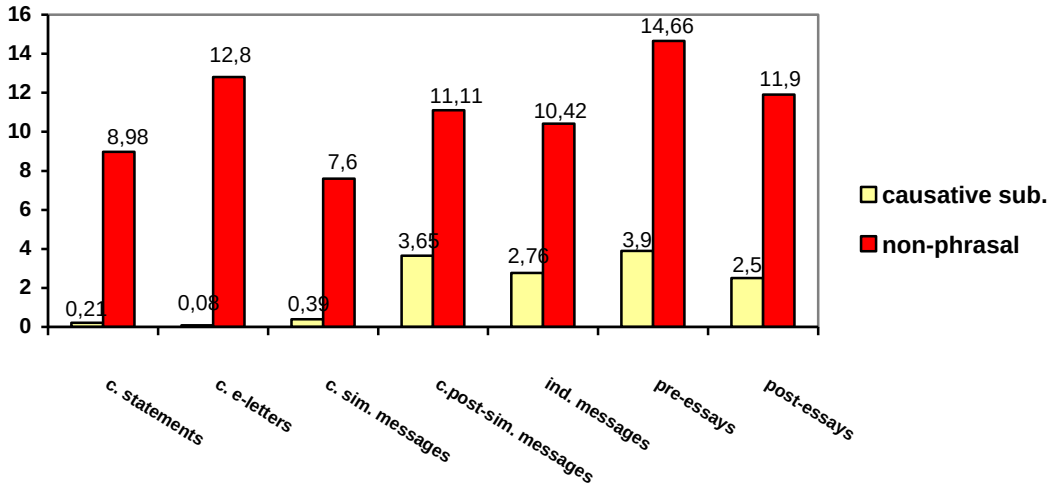


Figure 40: **Number of causative subordination *because* and non-phrasal coordination (per 1000 words) in the corpus of student-composed electronic texts**

Frequency variation of non-phrasal coordination correlates with causative subordination. This coordination is predominantly used in interactional and less formal texts. in which it links clauses in a loose, logically unspecified way, instead of integrating the information through the use of prepositional phrases, restrictive relative clauses. adjectives and other linguistic features used in more transactional written discourse. Non-phrasal coordination (see *Figure 40*) is considerably frequently used by students in their electronic texts: it is the most frequent in pre-essays (14.66), but the least frequent in simulation messages (7.60), statements (8.98). Non-phrasal coordination tends to be overused by students in all text groups. This reveals that the correlation of this linguistic feature with the communicative purpose of the various electronic texts can cause difficulties to *NNSE*.

Do as a proverb and analytic negation

Do as a proverb reduces the informational density of a text and Biber (1988) emphasises that it can substitute for the whole clause. *Do* results in a more generalized, uncertain

context, as it stands for an unspecified verbal referent. It is more frequently used in interactional discourse in which the addresser and the addressee can rely on the shared information and they have a possibility for immediate clarification to identify the implied meaning. Like *Do* as a proverb, analytic negation (*not*) is also associated with fragmented presentation of information, resulting in a low informational density. Analytic negation *not*, being less formal than synthetic negation, is more frequently used in less formal discourse (see *Examples 6*).

Examples 6

It will **do**. (pre-essays)
I'll **do** it. (pre-essays)
You **did** a great job! (post. sim messages)
If one writes a letter to a friend. he should **not** use formal expressions. (pre-essays)
We **do not** consider it to be a division. (sim. messages)
The delegation hopes that this situation **will not** affect the overall process of gathering votes.(e-letters)

The distribution of *Do* as a proverb (Figure 41) in students' corpus reveals that it is considerably infrequent in all groups of texts and its use correlates with the respective texts in Biber's study (1988). For example, in the most transactional texts, found in the corpus, post-simulation messages (4.99), *Do* as a proverb correlates with its use in personal letters (4.30) in Biber's study. The frequency of the mentioned linguistic feature in student-composed individual messages (2.0) correlates with their frequency in professional letters (2.6), as explored by Biber.

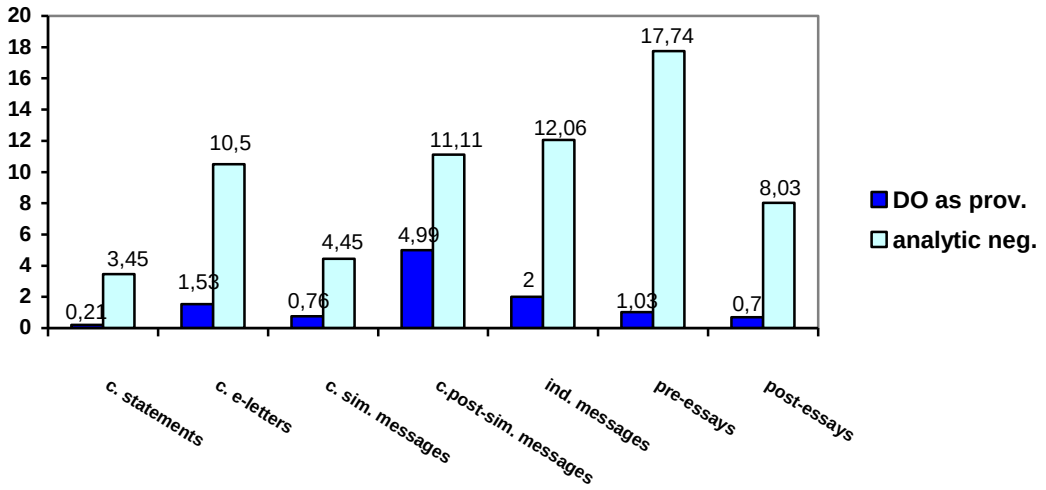


Figure 41: **Number of *Do* as a proverb and analytic negations (per 1000 words) in the corpus of student-composed electronic texts**

The use of *not*, however, shows considerably greater variation than the use of *Do* as a proverb (*Figure 41*). The lowest frequency of this negation is found in statements (3.45) and simulation messages (4.45), which in this respect correlate with their frequency in academic prose (4.30) in Biber's study. Analytic negation is considerably frequently used in such interactional texts written by students as e-letters (10.50), post-simulation messages (11.11), individual messages (12.06) and also expectedly transactional texts pre-essays (17.74) and post-essays (8.03). The number of analytic negation in pre-essays is unexpectedly high, so that it even considerably exceeds their number in face-to-face (18.50) conversation presented in Biber's (1988) study. Although the frequency of analytic negations in students' post-essays is twice lower than in pre-essays, the overall results confirm that students tended to overuse analytic negations in transactional texts.

General emphatics, hedges and amplifiers

These linguistic features indicate in discourse the degree of certainty and involvement with the topic. Thus, Biber (1988) points out that emphatics (*for sure, a lot, such a, real, just, really, most, more, so, do+verb*) denote presence or absence of certainty. whereas amplifiers (*absolutely, altogether, completely, enormously, entirely, extremely, fully, greatly, highly, intensely, perfectly, strongly, thoroughly, totally, utterly, very)* show the degree of certainty. Amplifiers mark heightened feeling, so they are a characteristic feature of less formal, interactional discourse. Chafe (1985) claims that amplifiers (or intensifiers) show the reliability of proposition. As to emphatics, Chafe (ibid.) considers that this linguistic feature indicates involvement with the topic. Hedges, chosen for the present study (*about* [not as a preposition] *something like. more or less. almost. maybe. sort of. kind of*) like emphatics and amplifiers function as informal. less specific markers of probability or uncertainty. Hedges are used to signal uncertainty or lack of precision in the presentation of information (see *Examples 7*).

Examples 7

Sometimes we can spend many hours **just** sitting and talking about nothing with our friends. (pre-essay)
It takes hours and hours to find something **really** useful. (pre-essays)
It was a pleasure to have **such a** nice discussion. (sim. messages)
We **completely** agree with the DAP policy vision of Northland. (sim. messages)
We **fully** support this point. (sim. messages)
Our delegation wants to thank you for a **very** productive and fruitful teleconference. (sim. messages)

Your idea of dividing the responsibilities of the agency is really **very** useful as it could make data operation process more accurate and elaborate.(e-letters)
This institution is based on the platform of the Governmental Intelligence Service and is a **kind of** independent organization with all the actions regarded as a part of State Restricted Information Service. (sim. messages)

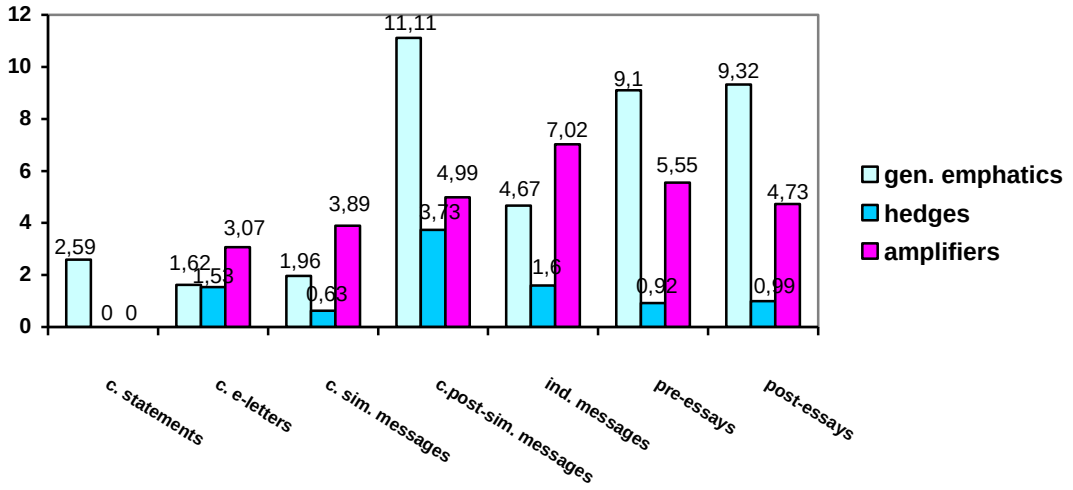


Figure 42: **Number of general emphatics, hedges and amplifiers (per 1000 words) in the corpus of student-composed electronic texts**

All three linguistic features occur with low frequency (*Figure 42*) in statements, namely, the frequency of general emphatics is 2.59, but amplifiers and hedges are not found in this text group. Thus, statements with respect to these linguistic features overlap with academic prose in Biber's study in which their frequency is also low: emphatics 3.60. hedges 0.20 and amplifiers 1.40. E-letters and even simulation messages also show considerably low frequency of these linguistic features. The frequency of general emphatics in e-letters (1.62) and in simulation messages (0.63) is even lower than in statements. The use of hedges is a bit more frequent in e-letters (1.53) and in simulation messages (0.63). As to amplifiers in students' e-letters (3.07) and simulation messages (3.89), they are more frequent than in personal letters (2.20) in Biber's study, but less frequent than in face-to-face conversation (6.0) in the said study. The use of emphatics, hedges and amplifiers ranks the highest in post-simulation messages (general emphatics 11.11, hedges 8.73, and amplifiers 4.99). The frequency of hedges in post-simulation messages exceeds their frequency in face-to-face (2.10) conversation in Biber's study. while the use of emphatics tend to correlate with face-to-face conversation (12.20) and the frequency of amplifiers is a bit lower (6.0) than in the said study. Although students have successfully varied the use of these three linguistic features in all the above analysed text

groups. the variation of general emphatics and amplifiers are less marked in pre-essays and post-essays than expected. According to the frequency of general emphatics (9.32), post-essays are close to professional letters in Biber's study (7.80), but according to the frequency of amplifiers (4.73) these essays are between personal letters (2.20) and face-to-face conversation (6.00) in Biber's study. The frequency of hedges, however, is insignificant in pre-essays (0.92) and in post-essays (0.99) and is close to their frequency in academic prose (0.20) in Biber's study. The obtained results confirm that general emphatics and amplifiers have been used relevantly in the cases when the communicative purpose of the texts has been analysed in detail by students during their participation in the simulations of the project *IDEELS*. However, the frequency variation of the said linguistic features in pre-essays and post-essays reveal that the correlation of the use of these two linguistic features in transactional electronic texts tends to cause difficulties to the students.

Prepositions

Prepositional phrases are used in order to integrate high amounts of information into discourse and thus they are used in more formal texts. Thus, this linguistic feature is more frequent in informational texts. Biber (1988) has found that prepositions tend to co-occur with nominalizations and passives in academic discourse (the frequency of these linguistic features are considered in the subchapter *Summary of the Overall MD Relations among the Electronic Texts*), professional letters and other informational texts of more formal discourse (see *Examples 8*).

Examples 8

They need to think **about** the better variant. (post-essays)
The data **on** a person or a group should be taken **from** the Database. (e-letters)
Being a part **of** the Eutropian Federation, we expect to get support **from** other Eutropian countries. (statements)

The occurrence of prepositions (*Figure 43*), as expected, is the highest in the most transactional of student-composed texts according to their communicative purpose - statements (115.04) and it is a bit lower in e-letters (105.81) and simulation messages (97.47). but the lowest in post-simulation messages (72.94). As to the variation of prepositions in pre-essays and post-essays, the frequency of prepositions in pre-essays is

even lower (57.09) than in post-simulation messages. Prepositions are considerably more frequent in post-simulation essays (100.4). However, according to the overall frequency variation of prepositions in the student-composed texts, their frequency is lower than in respective texts in Biber's (1988) study. For example, academic prose contains 139.50 prepositions in Biber's study, whereas the most transactional texts in the corpus – statements – contain only 115.04 prepositions. These findings imply that students require practice in the use of prepositional phrases in more transactional texts in order to use them for the integration of the information in this discourse.

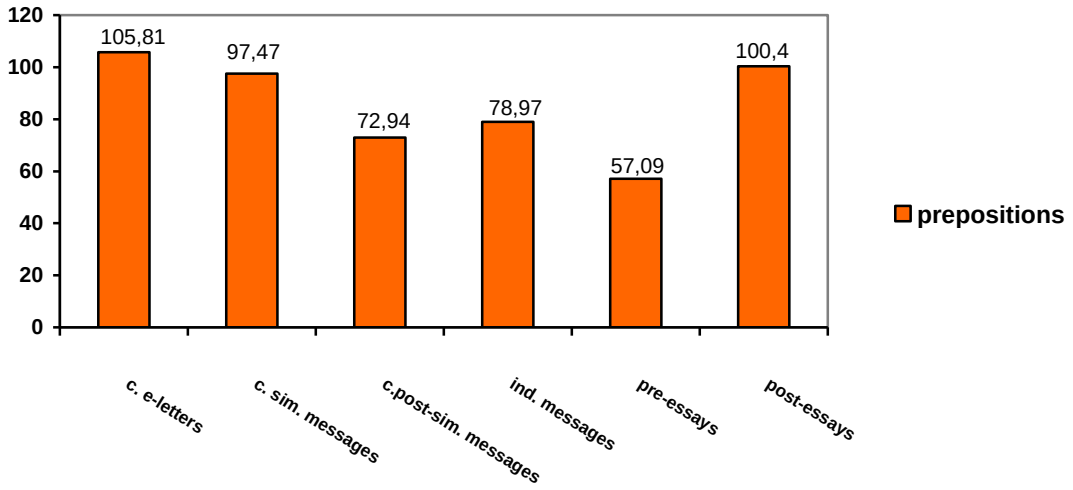


Figure 43: **Number of prepositions (per 1000 words) in the corpus of student-composed electronic texts**

2.2.2 Dimension B: explicit versus situation dependent refence in discourse

Explicit versus situation dependent reference in discourse is *Dimension 3* or *B.* as labelled by Biber and Finegan (1989). This dimension comprises the linguistic features that distinguish between highly explicit, context-independent versus non-specific, situation-dependent reference. This dimension includes 7 linguistic features of which 4 are with positive, but 3 with negative weight. According to Biber's (1988) findings more transactional and informative texts present higher scores, whereas less formal texts display lower scores in *Dimension B.*

A more frequent use of all linguistic features with positive weight in this dimension denote explicit focus of discourse, whereas more frequent features with negative weight mark situation dependent characteristics of discourse. Biber (1988) has

proved that more transactional and informative texts present higher scores in *Dimension B.* but interactional and less formal texts display lower scores in this dimension.

Table 14 and *Table 15* display dimension scores of linguistic features in collaboratively and individually student-composed texts and the total dimension scores for each text group. Summary of the overall MD relations among students' electronic texts is presented in detail in the subchapter *Summary of the Overall MD Relations among the Electronic Texts.*

Table 14: **Dimension B (explicit versus situation dependent reference) scores of linguistic features of collaboratively written texts**

Linguistic features	Statements	E-letters	Simulation messages	Post-simulation messages
+ (27) WH relative clauses	+1.35	+0.36	+1.17	-1.67
+ (28) Pied piping constr.	+0.47	+2.78	+0.05	-1.14
+ (29) Phrasal coordin.	+0.67	-0.47	-0.58	-1.44
+ (30) Nominalizations	+2.01	+0.55	+0.78	-1.31
- (31) Time adverbs	-1.52	-0.91	+0.61	+1.23
- (32) Place adverbs	-0.53	-0.24	-0.11	+1.97
- (33) Other adverbs	-1.66	-1.37	-0.93	+0.9
Score of *Dimension B*	**+8.21**	**+5.74**	**+1.85**	**-9.66**

Table 15: **Dimension B (explicit versus situation dependent reference) scores of linguistic features of individually written texts**

Linguistic features	Pre-essays	Post-essays	Messages
+ (27) WH relative clauses	+0.18	-0.21	-0.64
+ (28) Pied piping constr.	-0.37	-0.3	-0.02
+ (29) Phrasal coordin.	+0.83	+0.63	-0.6
+ (30) Nominalizations	-0.23	-0.44	-0.44
- (31) Time adverbs	+0.64	-0.17	-0.28
- (32) Place adverbs	-0.54	-0.56	+0.39
- (33) Other adverbs	+0.86	+0.11	+0.56
Score of *Dimension B*	**-0.55**	**+0.30**	**-2.37**

Figure 44 shows the variation of dimension scores of student-composed texts in comparison with the dimension scores obtained by Biber in his study. Collaborative statements (+8.21) hold the highest position in this dimension and they are more explicit and transactional than professional letters and academic prose in comparison with Biber's study. The next rank collaborative e-letters (+5.74) that exhibit even lover score than professional letters in Biber's study, which means that student-composed e-letters are

more explicit texts when the said professional letters. Although simulation messages (+1.85) written during virtual conferences have lower score than e-letters, they are considerably more explicit than individual messages (-0.55). The most situation dependent texts are messages of post-simulation conferences (-9.66), which rank considerably lower than face-to-face conversation in Biber's study. According to dimension scores, post-essays (+0.30) are more explicit than pre-essays (-0.55), although in this dimension this difference is less significant than in *Dimension A.*

Dimension B

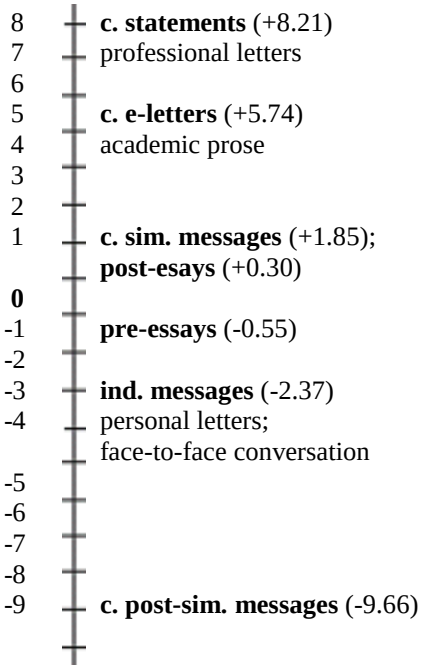


Figure 44: **Dimension B *Explicit versus Situation Dependent Refence in Discourse* for *NNSE* texts (in bold) grouped according to task types in comparison with the texts grouped in registers by Biber (1988)**

The selected linguistic features that account for the above-presented dimension scores (*Table 12*, *Table 13*, *Figure 44*) are analysed in detail. The frequency of linguistic features in student-composed electronic texts has been compared with the previous linguistic research, namely, Biber's (1988) findings regarding the four groups of texts denoted in *Figure 34*: face-to-face conversation, personal letters, professional letters, academic prose.

Adverbs of time and place

Time adverbials (e.g. *earlier. soon. now. then. recently*), as emphasised by Biber (1988) are used for text-internal referents and for reference to times outside of the text. Thus, they often serve as deictics that can be understood by reference to an external temporal situation. Place adverbs (e.g. *above. behind. there. backward. forward*) are used to denote temporal reference within a particular text or for reference to places outside the text. These adverbs, like time adverbs, can serve as deictics that is understood by reference to an external physical situation. These exmples of the mentioned adverbs found in the corpus of *NNSE* are presented in *Examples 9*

Examples 9

We are glad to start our negations **today**. (sim. messages)
We will be **there.** (sim. messages)
I am **here**. (post sim. messages)

Time and place adverbs have been used by *NNSE* in all text groups; however. their frequency variation are less marked than the frequency of. for example. such linguistic features as the first and second person pronouns. The most frequently these adverbs (see Figure 45) are used in the most situation dependent interactional texts – post-simulation messages (time adverbials 6.19 and place adverbials 6.27). This frequency correlates with their use in personal letters (place adverbs 2.0. time adverbs 8.30) in Biber's (1988) study, although time adverbs are a bit less frequent in this study.

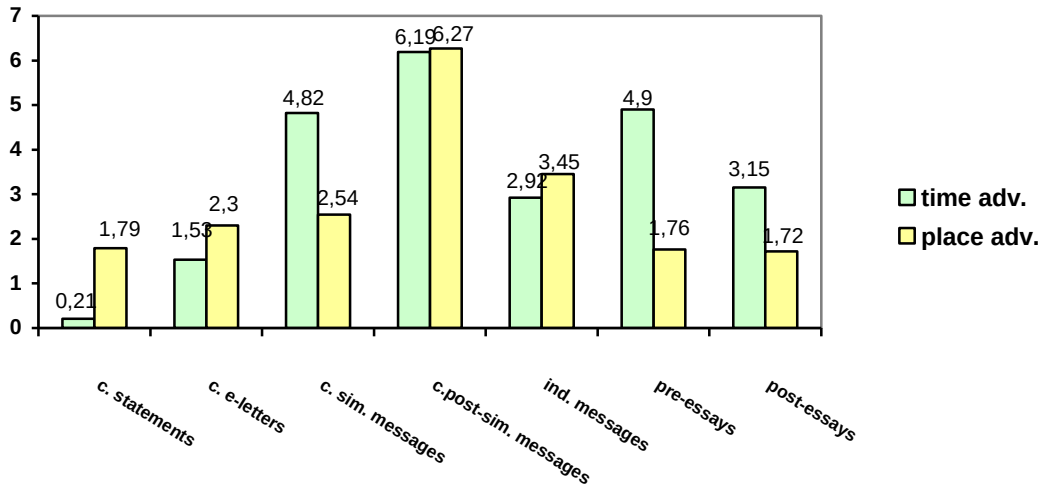


Figure 45: **Number of time adverbs and place adverbs (per 1000 words) in the corpus of student-composed electronic texts**

Time adverbs (6.19) and place adverbs (6.27) are the most frequent in post-simulation messages as well as considerably frequent in simulation messages (time adverbs 4.82. place adverbs 2.54). Their frequency in these electronic texts is caused by the purpose of these messages: during the virtual conferences. it was important to provide time and place reference to the events and the time of occurrence of these events in the previous or future conferences to ensure smooth flow of the virtual simulations.

Statements, in contrast, contain a considerably lower frequency of time and place adverbs, which is even lower than in academic prose (time adverbs 2.80. place adverbs 2.40.) in Biber's study (1988). Similarly in e-letters the frequency of these linguistic features (time adverbs 1.53. place adverbs 2.30) also correlates with their frequency in academic prose in Biber's (ibid.) study. but these linguistic features are a bit more frequent in individual messages (time adverbs 2.92. place adverbs 3.45). In pre-essays, time adverbs (4.90) are overused by students as their frequency correlates with the use in face-to-face conversation (5.10) in Biber's (1988) study. The students have reduced the use of time adverbs denoting the reference to times outside the text in their post-essays. Thus, students have correlated their use of time and place adverbs with the communicative purpose of their electronic texts.

2.2.3 Dimension C: abstract versus non-abstract information

Abstract versus Non-abstract Information is dimension *5* or *C* as labelled by Biber and Finegan (1989). The linguistic features belonging to this dimension distinguish informational discourse that is abstract, more technical and formal versus more interactional discourse. This dimension includes six linguistic features and five of them have been explored within the present study. All these features are with positive weight. Higher scores in *Dimension C*, according to Biber's (1988) findings, have been displayed by more informational and formal discourse.

Table 16 and *Table 17* show *Dimension C* scores of linguistic features in collaboratively and individually written texts and the dimension scores for each text group. *Figure 46* exhibits variation of dimension scores of student-composed texts in comparison with the scores obtained by Biber (1988) in his investigation of the texts grouped according to registers.

Table 16: **Dimension C (abstract versus non-abstract information) scores of linguistic features of collaboratively written texts**

Linguistic features	Statements	E-letters	Sim. messages	Post-simulation messages
+ (34) Conjuncts	-0.17	+0.25	-0.7	-1.60
+ (35) Agentless passives	+0.42	+0.83	+0.88	-1.82
+ (36) BY-passives	+0.51	-0.9	+0.87	-0.9
+ (37) Past part. WHIZ d.	+1.81	+1.2	+0.44	-0.97
+ (38) Other adv. Sub.	-1.1	-1.1	+0.18	-1.1
Score of *Dimension C*	**+1.47**	**+0.28**	**+1.67**	**-6.39**

Table 17: **Dimension C (abstract versus non-abstract information) scores of linguistic features of individually written texts**

Linguistic features	Pre-essays	Post-essays	Messages
+ (34) Conjuncts	-0.01	+1.41	0
+ (35) Agentless passives	-0.37	-0.04	+0.4
+ (36) BY-passives	-0.53	-0.13	-0.14
+ (37) Past part. WHIZ d.	-0.69	-0.34	-0.14
+ (38) Other adv. Sub.	+0.43	+1.25	-0.01
Score of *Dimension C*	**-1.17**	**+2.15**	**+0.11**

Dimension C

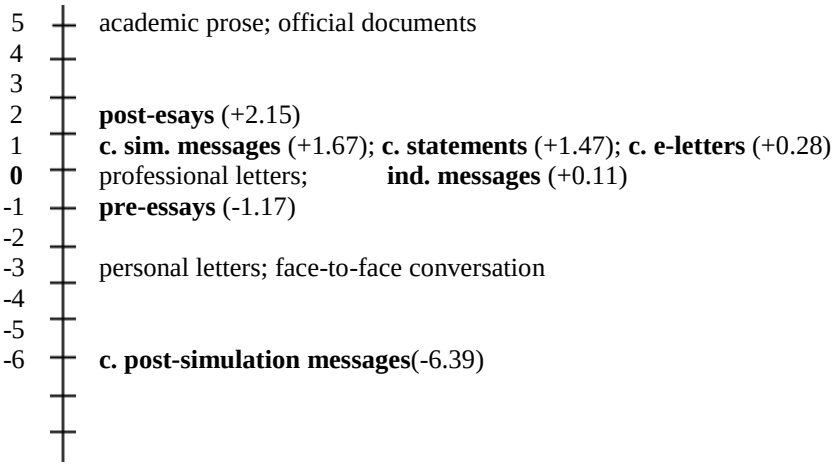


Figure 46: **Dimension C *Abstract versus non-abstract information* for *NNSE* texts (in bold) grouped according to task types in comparison with the texts grouped in registers by Biber (1988)**

In this dimension, student-composed post-essays (+2.15) rank the highest and their dimension score that is the closest to academic prose in Biber's study. Surprisingly that simulation messages (+1.67), which rank next after post-essays, are even more abstract and thus more transactional texts according to this dimension than statements (+1.47). According to the dimension score e-letters (+0.28) are close to Biber's professional letters. However, the messages of post-simulation conferences display the lowest dimension score (-6.39) and are even considerably more non-abstract than face-to-face conversation in Biber's study.

The frequency of three linguistic features in the corpus of *NNSE* have been analysed in detail in order to reveal the way they account for the presented dimension scores of the students' electronic texts.

Conjuncts

Biber (1988) points out that conjuncts mark more complex logical relations among clauses that characterize more abstract and formal discourse. The following conjuncts have been explored in the present study:

Alternatively, altogether, consequently, conversely, eg., e.g., else, furthermore, hence, however, i.e,. instead, likewise, moreover, namely, nevertheless, nonetheless, notwithstanding, otherwise, rather, similarly, that is, therefore, thus, viz, in (comparison, contrast, particular, addition, conclusion, consequence, sum, summary, any event, any case, other words), for example (instance), by contrast (comparison), as a result, (consequence), on the contrary, (other hand)

The students have refrained from the use of conjuncts in post-simulation messages (see Figure 47) that, according to the communicative purpose and the obtained scores of all three dimensions, are the most interactional texts. However, all the other texts display considerable frequency of conjuncts. In simulation messages (2.79) students have used conjuncts nearly as frequently as they have been used in academic texts (3.0) in Biber's study. The conjuncts have been used considerably frequently in other electronic texts written by students: statements 4.45, e-letters 5.80, individual messages 5.03, pre-essays 4.96, post-essays 9.43 (see *Examples 10*). These results show that students tend to overuse this linguistic feature in transactional texts, such as post-essays. The overuse of conjuncts in post-essays can be explained by students' use of them for linking of the ideas across the sentences of a text.

Examples 10

To sum up, I can say that both ways are very important. (pre-essays)

As a result, we discovered that subjects at the University are more interesting. (pre-essays)
However, the second type of learning is not so widespread. (post-essays)
Moreover, vocabulary of the spoken and written communication differs as to nominal style and lexical density. (post-essays)

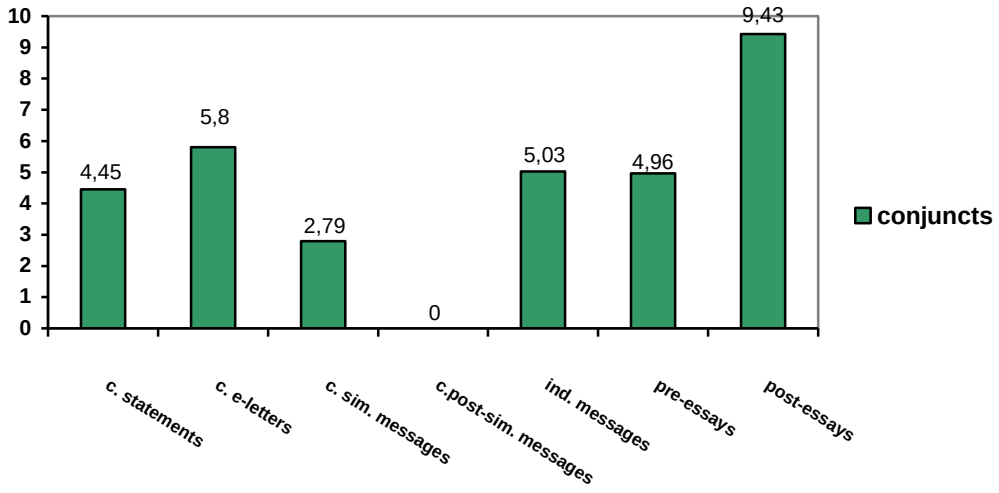


Figure 47: **Number of conjuncts (per 1000 words) in the corpus of student-composed electronic texts**

Agentless passives and BY passives

Agentless passives (the passive without agents) function as markers of the decontextualized and detached discourse. Biber emphasizes (1988) that these passives are used in the case of more abstract presentation of the ideas, where the same agent is refers to several clauses and the specific agent of a clause is not important to the discourse purpose. *By*-passive constructions are applied to present abstract, technical information in discourse in formal way. *Examples 11* contain the passive constructions used by students in their electronic texts.

Examples 11

Formal communication **is** basically **used** at work. (pre-essays)
At university more detailed writing practice **is provided**. (post-essays)
The commission **will be divided** into several specific departments. (e-letters)
These stylistic differences **are studied by linguists**. (pre-essays)
Both learning institutions **are** predominantly **attended by young people**. (post-essays)
The celebration was organized **by the non-governmental organizations**. (e-letters)

The variation of agentless and *BY* passives is considerably marked across student-composed discourse. Agentless passives (*Figure 48*) are the most frequently used in four

groups of student-composed electronic texts: individual messages (12.00) statements (12.15), e-letters (13.34) and simulation messages (14.59). Their frequency, especially in e-letters and simulation messages, is considerably close to academic prose (17.0) explored by Biber (1988). It is surprising that students' statements, which according to other linguistic features tend to be the most transactional texts, exhibit lower frequency than simulation messages. The frequency of agentless passives in pre-essays (7.83) and post-essays (9.60) reveals that their use in these texts is the closest to professional letters (7.30) in Biber's study. In post-essays this linguistic feature is noticeably more frequent than in pre-essays. It means that post-essays contain more abstract ideas that correlate with the communicative purpose of these texts. In post-simulation messages, which are the most interactional text group according to the dimension scores, agentless passives are not found. However. this text group contains a comparatively noticeable frequency of BY passives (1.27). Similar frequency of BY passives is found in simulation messages (1.53) and statements (1.22) that rank them close to the use of this passive construction in academic prose (2.0) in Biber's study. Student-composed individual messages (0.65), pre-essays (0.32) and post-essays (0.67) contain even less significant frequency of *BY* passives and with this the said text groups are the closest to professional letters (0.60) in Biber's study.

The obtained results confirm that the students have varied the use of the passive constructions according to the purpose of the electronic text groups. but the overall frequency results also reveal that both passive constructions are used less frequently across all text groups in comparison with the respective texts explored by Biber.

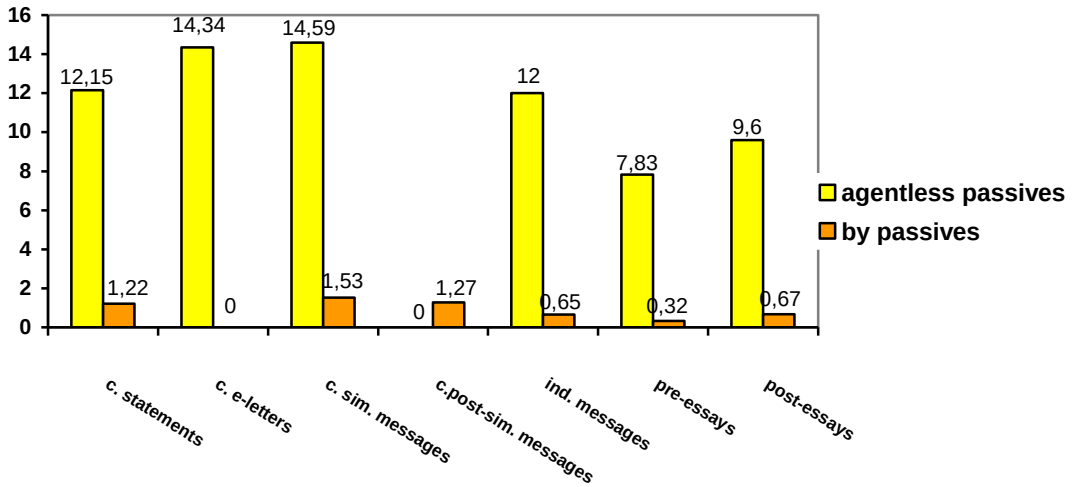


Figure 48: **Number of agentless and BY passives (per 1000 words) in the corpus of student-composed electronic texts**

The comparison of the dimension scores and the frequency variation of linguistic features in the student-composed texts with the dimension scores and the frequency of linguistic features per 1000 words in Biber' study of native speaker's texts confirms that the students have varied the linguistic features according to the communicative purposes of the electronic texts. However, this comparative analysis also reveals the linguistic features that students' tend to overuse or underuse in particular text groups in comparison with the native speakers' authentic texts. Therefore, apart from the above comparison of students' electronic texts with the native speakers' texts, it is important to summarise the overall MD variation among all groups of student-composed texts included in corpus and identify the statistical significance of the differences of dimension scores displayed by these texts.

The overall relations among the electronic texts grouped according to the task type that are included in the corpus are presented in the next chapter.

2.2.4 Relations among Student-composed Electronic Texts

This chapter contains the description of relations among all student-composed electronic texts and their variation: collaboratively and individually written texts.

2.2.4.1 Collaborative Texts

Collaborative texts have been written by students during their participation in the project *IDEELS* simulations. The total scores of these texts in each of the three dimensions (*Dimension A. Dimension B. Dimension C*) are exhibited in *Table 18,* whereas *Figure 49* shows the variation of the texts according to these dimensions. The implemented Chi-test proves that the difference between the mentioned text groups is statistically significant. The examples of the collaboratively developed texts are found in *Example 13. 14. 15. 16.* These examples show the way students have varied the use of linguistic features, depending on the purpose of these texts.

Table 18: **Dimensions scores of collaboratively developed texts**

Dimension / Text groups	A	B	C	Sig. level
Statements	-31.70	+8.21	+1.47	0.01
Letters	-8.14	+5.74	+0.28	
Simulation messages	-11.98	+1.85	+1.67	
Post-simulation messages	+27.36	-9.66	-6.39	

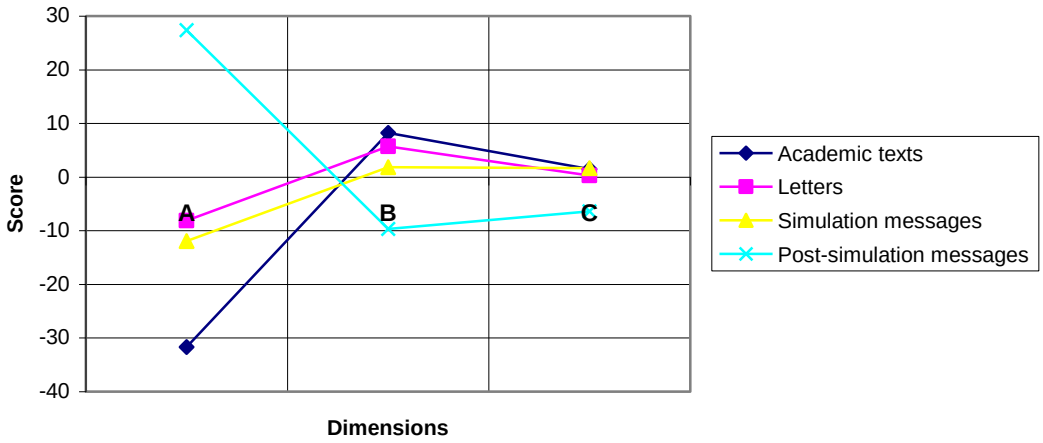


Figure 49: **Comparison of collaboratively developed texts**

As it is seen in *Table 18* in *Dimension A*, informational versus involved production, the mentioned four groups of texts show the most marked differences. Thus. the greatest scores are displayed by post-simulation messages (*Example 16*), which the students exchanged after the simulation session and which aimed at the discussion of the insights students had gained during the simulation. According to the scores, these messages are highly interactional texts that resemble conversation. Collaborative statements (*Example 13*), e-letters (*Example 14*) and simulation messages (*Example 15*) display considerably lower sores. Surprisingly, simulation messages, which were written during synchronous online communication, show even lower scores than letters, whereas academic texts, as it was expected, display the lowest scores and are the most transactional texts.

Example 12: **Statements: *Policy Statement***

Table of contents

1 Executive summary
2 Goals
3 Data Protection in *Bardland*
4 Environmental Information Regulations in *Bardland*

5 Freedom of Information in *Bardland*
6 Bibliography

1Executive summary

By means of this policy statement we will try to find solutions to the existing problems connected with Data Access and Protection (*DAP*) in *Bardland*. Being a part of the *Eutropian Federation* we expect to get support from other *Eutropian* countries. and to come to common solutions. The solution of problems connected with data access and protection, undoubtedly, are of vital importance. The overall goals of Bardland are to develop and to improve the system of *DAP* in *Eutropian Federation*.

3 Freedom of Information in Bardland

The Freedom of Information Act (FoIA) was issued to give people the right to ask for information from public authorities, which do not fall under data protection or environmental information regulations as listed above. FoIA is intended to promote a culture of openness and accountability among public sector bodies and to facilitate better public understanding of how public authorities carry out their duties. why they make the decisions they do. and how they spend public money. The FoIA requires government agencies to make public their data access and protection.

Example 14: **E-letter**

Subject: RE: GOV: Agencies of the Government.

To the Delegation of Northland

We got acquainted with the formulation of the functions of the Special Agency proposed by your Delegation. and we would like to express our general agreement with all the ideas. Your idea of dividing the responsibilities of the agency is really very useful as it could make the Data operation process more accurate and elaborate.
However. we would like to state our strong conviction that the separation of the functions of the Agency, according to the storage of different kinds of data, could cause a kind of overlapping responsibilities and even could be a reason for raising the bureaucratic tendencies within this sphere.

Concerning all other points our political views are quite similar.

We look forward to further negotiations between our Governments
The Government Delegation from the Republic of Bardland

Example 15: **Examples of simulation teleconference messages**

<Message 89> Our delegation also agrees with the majority of the participants. and we would like to speak over this issue in the afternoon.

<Message 61> This institution is based on the platform of the Governmental Service and is a kind of independent organisation with all the actions regarded as the part of State Restricted Information Service. The access to this Data should be in hands of experts.

<Message 97> There are certain legal requirements to disclose information, for example, in relation to infectious deseases. Doctors are obliged to report incidences of specified infectious diseases to health boards and the Infectious Diseases Surveillance Centre. This information must be disclosed wihout identifying the patient. The confidentiality of personal information such as medical records must be protected by the data protection legislation. beyond doubt.

<Message 99> We would like to make some proposals about Data Access and Protection in financial field. The access to the financial history of the individual is available to: (1) data subject; (2) financial institution; (3) Government (in exceptional cases). By exceptional cases we mean if the data subject is involved in criminal affairs. Then the Government is allowed to access the individual's financial history.

Example 16: **Post-simulation messages**

All the topics were quite interesting bacause they are actually the everyday problems that affect all the countries, aren't they? But we are FOR SURE more informed about them now!! Why do you think so? We think that we also learned a lot of new things. What do you mean? It was great and we gained new experience! Our feelings are the same! Who are you? Where are you from? We're actually not form Spain.

Table 18 and *Figure 49* show that in *Dimension B*, which marks the degree of explicitness in discourse, there are noteworthy differences among the four groups of texts. The highest scores are displayed by statements (*Example 13*). so they are the most explicit among the four text groups. E-letters (*Example 14*) exhibit lower scores than statements, whereas simulation messages (*Example 15*) display lower scores than e-letters. Although both simulation and post-simulation messages display the lowest scores, there is a noteworthy difference in scores between these two text groups (see *Figure 50, Example 15* and *Example 16*) in this dimension. Simulation messages, which aimed at consensus building among the delegations of the simulated *Eutropian Federation* on essential policy issues of this country, are considerably more explicit (see *Example 15*). In contrast, post-simulation messages, which according to dimension scores resemble conversation. are the least explicit (see *Example 16).*

The scores of *Dimension C* (*Table 18, Figure 50*) reveal surprising results. Although statements (*Example 13*), which display considerably high scores in this dimension, prove to be abstract and thus transactional texts, the simulation messages (*Example 15*) tend to display even higher scores than statements. These messages were developed by using the messaging system that functioned like a chat system; however, the content and goal of the virtual negotiations required the choice of the linguistic features used in more transactional texts. Post-simulation messages (*Example 16*), as it has been expected, show considerably lower scores in comparison with academic texts, letters and simulation messages.

According to the dimension scores, shown in *Table 19* and illustrated in *Figure 50*. all four text groups display noticeable differences. However, the greatest variation is along the dimensions *A* and *B*. This variation shows that the most noticeable transfer from interactional to transactional written communication is displayed by the use of linguistic features in post-simulation messages on the one hand and statements, e-letters and

simulation messages on the other hand. Thus, collaboratively written texts display considerable variation.

2.2.4.2. Individual Texts

Individually written texts were developed during writing activities envisages by e-course *English academic writing III* as well as before and after students' participation in collaborative and individual writing activities in a form of pre-essays and post-essays. The scores of individually written pre-essays, post-essays and messages in each dimension: *A. B. C* are presented in *Table 19, Figure 50* shows variation of texts according to these dimensions. The implemented Chi-test proves that the difference between the mentioned text groups is statistically significant. The examples of individually written messages, pre-essays and post-essays are included in *Examples 17. 18. 19.*

Table 19: **Dimensions scores of individually developed texts**

Dimension / Text groups	A	B	C	Sig. level
Pre-essays	+14.06	-0.55	-1.17	0.01
Post-essays	-4.17	+0.30	+2.15	
Messages	+1.68	-2.37	+0.11	

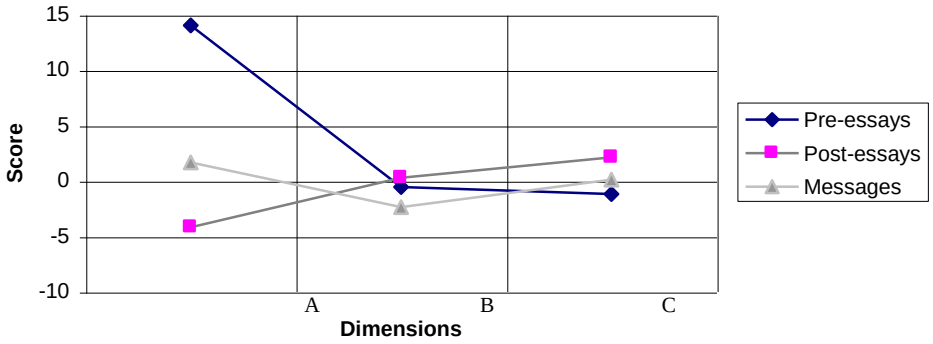


Figure 50: **The comparison of individually developed texts**

The obtained results show that in *Dimension A.* informational versus involved production. all three text groups display noteworthy differences (*Figure 50*). The features with positive loadings have greater scores in pre-essays (*Example 18*) than in post-essays (*Example 19*) and messages (*Example 17*). Similarly the absolute values of the scores of features with negative weights in *Dimension A* are greater in pre-essays than in post-

essays and letters. Moreover, this difference is more marked between pre-essays and post-essays than pre-essays and messages. Since, according to Biber (1988. 1995), higher scores of linguistic features clustered in *Dimension A* are displayed by more involved and interactional texts, pre-essays are the most interactional of all three individually developed texts (see *Example 18* for the use of linguistic features in them). However, post-essays according the scores of linguistic features are considerably more informational and thus transactional than pre-essays (see *Example 19* for the use of linguistic features in them).

Example 17: **Individual message**

> It is possible to use personal pronouns in the essay. But the writer should avoid expressing his personal opinion too straightforwardly by, for example, writing. *I guess*. Instead of this phrase, another sentence should be used, and it should be more formal, and it means that the first person pronoun, most probably, will disappear. However, in my opinion, it is necessary to use personal pronouns in quotations, for example, if the writer wants to make a quotation from a book and the quote contains first person pronoun it won' t be omitted. of course.

Example 18: **Pre-essay**

> It has always been very important to know how formal and informal communication differ. Nowadays it's extremely important.
>
> I can say that we must communicate in a different way with people who are older. I can say that for me informal communication is easier because I know that I can say everything that is on my mind. I don't have to use special words; I don't have to think about what to say. I can say everything what is on the tip of my tongue. But I understand that more formal communication is also necessary.
>
> When do we use formal language? Do we use it only when we speak with teachers and business partners? I guess. But when I'm with my friends, relatives or family I speak absolutely differently. And, of course, this way of communication is much easier. because I don't have to think about the choice of a right word or intonation.
>
> I can say that I use both informal and formal communication. When I communicate with people who are of my age, I use informal language, and when I communiacte with people older than me I use formal language. I don't like formal language because I think it does not make the special connection between two or more people.

Example 19: **Post-essay**

> The question of various kinds of institutions for punishing the criminals has always been a topical issue. Yet, more contradictory are discussions concerning the sentence of under-age law-breakers. As there is no clear evidence of the refinement of behaviour, and lifestyle of people leaving the institutions of punishment, one might doubt whether the system of sentence actually works. A new type of establishments aiming to take a highly disciplined and pedagogically tended care of under-age law-breakers should be introduced in Latvia.
>
> Though there is a distinction between the punishment adapted to grownup criminals and teenage law-breakers. it is not providential enough. Under-age criminality should be controlled by special pedagogical approach. as teenagers are very flexible social group that can easily be influenced to change for better. Although under-age criminals often come from disadvantaged families, thus bringing the vice into society, there is a chance for them to be persuaded of a different way of living. By the help of new establishments, problematic teenagers would become aware of their own abilities to receive from society appreciation by giving positively tended attitude in return.

Alongside *Dimension B* (*Table 14*), explicit versus situation-dependent reference, there is a less noteworthy difference among the three text-groups, pre-essays, post-essays and messages, than the differences displayed by these four of texts in *Dimension A.* The highest score among the three text-groups is displayed by post-essays (*Example 19*). This, according to Biber's (1988) findings, means that post-essays are the most transactional among all three text groups. Pre-essays (*Example 18*) displaying lower total score and messages (*Example 17*) displaying the lowest score of all three text groups are less explicit and thus more interactional than post-essays.

In terms of Dimension *C,* abstract versus non-abstract information, the three text groups show noticeable variation. Biber (1988) has found that transactional texts that contain more abstract information display higher dimension score, whereas more interactional texts show lower dimension score. as they are less abstract. According to the dimension scores shown in *Table 21* and *Figure 51,* the most abstract and thus transactional are post-essays (*Example 19*). Surprisingly, the obtained data shows that pre-essays (*Example 18*), which present considerably lower score than messages, are even less abstract and thus more interactional than messages (*Example 17*).

As can be seen in *Table 21* and *Figure 50,* the most marked variation among the individually developed texts, pre-essays, post-essays and messages lies in Dimension *A.* although considerably noticeable differences are also in Dimensions *B* and *C.* Moreover, according to all three dimension scores pre-essays are the most interactional texts of individually developed electronic discourse.

MD variation in the student-composed electronic discourse proves that the post-essays are considerably more transactional than pre-essays in all three dimensions. The scores of linguistic features in all texts along all three dimensions, the students' individually developed pre-essays are interactional texts. Pre-essays were meant for presumable audience that, according to Barton (2004: 71), has been named 'a general academic audience'. However, the scores of linguistic features in pre-essays have shown that these essays have been written by the students as interactional texts. Therefore the students were offered to write individual post-essays in order to explore the impact of collaborative research and writing activities on their advancement of register-pertaining language use in transactional texts. The post-essays thus aimed at exploration of the possible correlations of linguistic features used by the students' after their participation in the technology-enhanced writing and research activities. According to the above results

student-composed post-essays show marked shift from more interactional to transactional texts.

Variation of dimension scores allows concluding that the students have varied the linguistic features according to the communicative purpose of the texts during their collaborative and individual writing activities. The value of these activities in the activation and enrichment of their writers' competences is confirmed by the variation of dimension scores in pre-essays and post-essays.

SUMMARY

Thus, the obtained results reveal noteworthy register variation among the individually and collaboratively developed texts. The comparison of dimension scores of all individually and collaboratively developed texts along Dimension *A* proves that student-composed electronic texts, as shown in Tables 14 – 16. 18 – 20, display the most significant shift from involved to informational and thus from interactional to transactional written communication. The most involved texts are collaboratively developed post-simulation messages, individually developed pre-essays and messages. The dimension scores of these text groups display positive weights. Less involved and hence more transactional texts are the collaborative statements, simulation messages, letters and individually developed post-essays. These texts display negative weights in their dimension scores.

The variation of the scores of all individually and collaboratively developed texts on Dimension *B* confirms the above-presented tendency, namely, the shift from interactional to transactional texts. However, the variation on *Dimension B* is less marked than the variation on *Dimension A*. Less explicit and therefore the more interactional texts are the collaboratively developed post-simulation messages, individually developed messages and pre-essays, which display negative dimension scores. More explicit and accordingly more transactional on this dimension are collaborative academic texts, letters, simulation messages and individual post-essays, which display positive dimension scores.

According to Dimension *C*. individually and collaboratively developed texts tend to range in a similar way as they range along Dimensions *A* and *B*. On Dimension *C*. the most abstract and therefore interactional texts are collaboratively developed post-simulation messages, individually developed pre-essays, since they show negative dimension scores. More abstract and accordingly more transactional are individually

developed messages, collaboratively developed letters, academic texts, simulation messages and individually written post-essays.

CONCLUSIONS

The theoretical and empirical research of the existing theories on the variation of linguistic elements in written and electronic written discourse has enabled the author to adapt these theories for the involvement of *NNSE* students in the development of interactional and transactional electronic texts to design the corpus of these texts and undertake multi-dimensional analysis of the designed corpus. The conclusions of this study are as follows.

1 Electronic discourse is hybrid. as it manifests itself in a wide range of interactional and transactional texts in use in the virtual environments. The choice of linguistic elements in electronic texts depends on their interactional, transactional and textual functions as well as the technical characteristics of the messaging systems that have been used to develop, maintain and use these texts in communication. Consequently, one of the biggest challenges of the research is the choice of lexical and grammatical elements as well as means of expression corresponding to the facilities and requirements of the medium used in the process of electronic discourse creation.

2 Writing of electronic texts in a foreign language requires special knowledge, competence and language system knowledge, which correlates with the knowledge of interactional and transactional functions of the electronic texts. In the course of the research, technology-enhanced model has been worked out for the development of interactional and transactional electronic texts that vary according to their communicative purpose. The model aims at upgrading students' knowledge and competence required for the development of electronic texts during online simulations within the framework of the project *IDEELS* and electronic writing activities offered to the target audience in the form of a specially designed e-course (*English Academic Writing III)*. The identification of different patterns of language use can contribute to better understanding of relevant language phenomena and the level of computer competence.

3 The analysis of theories on corpus design, in particular, corpus of student-composed texts leads to conclusion that part-of-speech taggers that are designed for native-

language marking of the texts of students who are advanced non-native learners of English. The specially designed structurally and part-of-speech marked corpus of texts arranged in groups according to their communicative purpose.

4 The investigation of theories on quantitative multi-dimensional analysis in the application of linguistic characteristics of spoken and written discourse leads to conclusion that the electronic discourse determines the selection of linguistic elements underlying functional dimensions along which, the variation in the texts of students' electronic discourse occur. To reveal linguistic variation in student-composed electronic texts, the variation of linguistic elements in them has been investigated along the following three dimensions: *Dimension A* (informational versus involved production). *Dimension B* (explicit versus situation dependent reference). *Dimension C* (abstract versus non-abstract information).

5 The results of investigation of the corpus of electronic texts have proved that students have activated the language system knowledge during their technology-enhanced research. Their participation in the online simulations within the framework of the project *IDEELS* and the activities envisaged in the e-course *English Academic Writing III* enabled them to correlate the choice of linguistic elements with the communicative purpose of electronic texts created and maintained in the virtual environments under discussion.

6 The results of quantitative multi-dimensional analysis show that there are statistically significant differences among student-composed electronic texts explored along the three selected dimensions. The implemented statistical method proves that the communicative purpose of each group of electronic texts determines students' variation of linguistic elements in them. The results of Chi-Square Test show that there is a significant statistical difference among the mean values along each dimension in collaboratively and individually written electronic texts. The seven groups of student-composed texts are different along one dimension, whereas more similar along another dimension that proves multi-dimensional variation of linguistic elements. For example, Synchronous simulation messages differ from e-letters and statements along *Dimension A* (Involved/Informational production) and *Dimension B* (Explicit/Situation-dependent

reference). but they are similar to e-letters and statements along *Dimension C* (Abstract/Non-abstract information).

7 The implemented statistical multi-dimensional analysis has proved that the communicative purpose of synchronous online communication is the reason for the significant frequency variation of linguistic elements in students' synchronous simulation conference messages and synchronous post-simulation messages along all three dimensions. This variation demonstrates that the messaging system used for synchronous online communication can host messages requiring marked variation of linguistic elements that depends on the communicative purpose of these messages.

8 The obtained results prove that there is significant statistical difference between mean values of student-composed the pre-essays written before their participation in electronic text creation within the framework of the project *IDEELS* and e-course *English Academic Writing III* and the post-essays created after their participation in the mentioned activities. These results prove that, according to the frequency co-occurrence of linguistic elements, the post-essays are considerably more transactional texts than the pre-essays, and thus they correlate with their communicative purpose.

9 The analysis of the results of the theoretical and empirical research as well as creation of a model help identify the most characteristic medium-specific means of expression and give recommendations for further development of academic electronic writing.

The results that have been achieved during the research prove the hypothesis advanced at the beginning of the doctoral thesis. Research results emphasise key linguistic issues, assisting acquisition of language use in electronic discourse. The focus of the research has been on the application of the main linguistic theories pertaining to the creation of relevant electronic texts. The investigation has shown that one of its biggest challenges in the creation of electronic discourse is the choice of lexical and grammatical elements. as they exist in virtual communicative environment. Since the quality of electronic discourse the students develop remains the main task, the e-course *English academic writing III* has been designed and offered for practical use. The principles of this course design can be used in the development of electronic materials for multi-purpose language studies.

The creation of electronic discourse may be regarded as new type of academic writing culture. The main advantage of electronic discourse is its development potential – electronic discourse appears to be the future of communication; that is conditioned by further development of modern computer technologies. Consequently, apart from linguistic competence, electronic discourse creation requires computer literacy, which includes the knowledge of virtual environments used in communication.

BIBLIOGRAPHY

1. Aristotelis (1959) *Poētika*. Rīga: Latvijas Valsts izdevniecība.
2. Baker, P. (2006). *Using Corpora in Discourse Analysis*. London and New York: Continuum.
3. Ball, C.N. (1994). Automated text analysis: Cautionary Tales. *Literacy and Linguistic Computing*, 9(4): pp. 295-302.
4. Barber, C.L. (1962). Some measurable characteristics of scientific prose. In: *Contributions to English Syntax and Phology.* Stockholm: Almquist & Wiksell, pp. 1-23.
5. Baron, N. S. (2000). *Alphabet to Email: How Written English Evolved and Where its Heading.* New York: Routledge.
6. Baron, N.S. (1998). Letters by phone or speech by other medium: the linguistic of email. *Language and Communication*, 18, pp. 133-170.
7. Bazerman, C. (1994). *Constructing Experience*. Carbondale, IL: Southern Illinois.
8. Beatty, K. (2003). *Teaching and Researching Computer-assisted Language learning.* Longman.
9. Belcher, D. (2004). Trends in teaching English for specific purposes. *Annual Review of Applied Linguistics*, 24, pp. 165-186.
10. Belcher, D. (2006). English for specific purposes: teaching to perceived needs and imagined futures in worlds of work, study, and everyday life. *TESOL Quarterly*, 40 (1), pp. 133-156.
11. Berkenkotter, C., Huckin,T. N. (1995). *Genre Knowledge in Disciplinary Communication – Cognition/Culture/Power.* Hillsdale, NJ: Lawrence Erlbaum Associates.
12. Bhatia, V. K. (1993). *Analysing Genre: Language Use in Professional Settings.* London: Longman.
13. Bhatia, V. K. (2004). *Worlds of Writen Discourse*. London and New York: Continuum.
14. Bhatia, V.K. (1997). The power of politics of genre. *World Englishes*, 17(3), pp. 359-371.
15. Bhatia, V.K. (1999). Integrating Products, Processes, Purposes and Participants in Professional Writing. In: C.N.Candlin and K.Hyland (eds.) *Writing: Texts, Processes and Practices*. London: Longman, pp. 21-39.
16. Bhatia, V.K., Swales, J.M. (1983). *An Approach to the Linguistic Study of Legal Documents.* Fachsprache 5, 3, pp. 98 -108.
17. Biber, D. (1986). Spoken and written textual dimesions in English: resolving the contradictory findings. *Language*, 62, pp. 384-414.
18. Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
19. Biber, D. (1995). *Dimensions of Register Variation.* Cambridge: Cambridge University Press.
20. Biber, D., Conrad, S. (2003). Register variation: a corpus approach. In: D.Schiffrin, D.Tannen, H.E.Hamillton (eds.), *The Handbook of Discourse Analysis* Blackwell Publishing, pp. 175-196.
21. Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., (2002). Speaking and Writing in the University: A Multidimensional Comparison. *TESOL Quarterly*, 36(1) pp. 9-48.

22. Biber, D., Csomay, E., Jones, J. K., Keck, C. (2004). A Corpus linguistic investigation of vocabulary-based discourse units in university registers. In: U.Connor, T.A.Upton (eds.) *Applied Corpus Linguistics: A Multidimensional Perspective*, Amsterdam-New York, NY, pp. 53-72.
23. Biber, D., Finegan E. (1989). Drift and Evolution of English Style: A history of Three Genres. *Language* 65, pp. 487-517.
24. Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (2004). *Grammar of Spoken and Written English.* Longman.
25. Bierman, K. (2005). Review of technology and teaching English language learners. *Language Learning and Technology,* 9(1), pp. 27-28.
26. Biesenbach-Lucas, S., Weasenforth, D. (2001). E-mail and word processing in the ESL classroom: how the medium affects the message. *Language, Learning and Technology,* 5, pp. 133-165.
27. Bjork, L., Brauer, G., Reinecker, L., Jorgensen, P.S. (2003). Teaching academic writing in European higher education: an introduction. In: L.Bjork, G.Brauer, L.Reinecker, P.S.Jorgensen (eds.) *Teaching Academic Writing in European Higher Education.* Dordrecht, London: Kluwer Academic Publishers, pp. 1-18.
28. Bloch, J. (2002). Student/Teacher interaction via email: the social context of internet discourse. *Journal of Second Language Writing*, 11, pp. 117-134.
29. Bolton, K., Nelson, G., Hung, J. (2002). A Corpus-based study of connectors in student writing. *International Journal of Corpus Linguistics* vol. 7, pp. 171-182.
30. Briggs, C.L., Bauman, R. (1992). Genre, intertextuality, and social power. *Journal of Linguistic Anthropology*. 2(2), pp. 131-172.
31. Brown, G & G. Jule (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.
32. Bruce, B.C., Rubin, A. (1993). *Electronic Quills: A Situated Evaluation of Using Computers for Writing in Classroom.* Hillsdale, N.J: Lawrence Erlbaum Associates.
33. Buhler, K. (1999). *Sprachtheorie. Die Darstellungsfunktion der Sprache.* Mit einem Geleitwort von Friedrich Kainz. Ungekurzter Neudr. der Ausg. Jena, Fischer, 1933, 3. Aufl. Stuttgart: Lucius &Lucius.
34. Burnard, L. (1995). The text encoding initiative: an overview. In: Leech.G., G.Myers, J. Thommas (eds.) *Spoken English on Computer*. Harlow, Essex: Longman.
35. Burnard, L., Sperberg-McQueen, C.M. (2002). *TEI Lite: An Introduction to Encoding for Interchange.* [Electronic version] Avaitable at <http://www.tei-c.org/Lite/index.html> accessed 15.01.2007.
36. Bušs, O., Joma, D., Kalnača, A., Lokmane, I., Markus, D., Pūtele, I., Skujiņa, V. (2007). *Valodniecības pamatterminu skaidrojošā vārdnīca*. Rīga: LU Latviešu valodas institūts.
37. Carter, R., Goddard, A., Reah, D., Sanger, K., Bowring., M. (2001). *Working with texts.* London and New York: Routledge.
38. Ceplītis, L., Rozenbergs, J., Valdmanis, J., (!1989). *Latviešu valodas sintakse.* Rīga: Zvaigzne, 147-167 lpp.
39. Chadwick, S., Bruce, N. (1989). The revision process in academic writing: from pen & paper to word processor. *Honkong Papers in Linguistics and Language Teaching*, 12 (April), pp. 1-27.
40. Chafe, W. (1982). Integration and involvement in speaking, writing and oral

literature. In: D. Tannen (ed.) *Spoken and Written Language: Exploring Orality and Literacy.* Norwood NJ: Ablex, pp. 35-54.
41. Chafe, W. (1992/2003). Discourse: Overview. In: W.Bright (ed.) (1992)/W.Frawley (ed.) (2003), *International Encyclopedia of Linguistics.* New York: Oxford University Press.
42. Chafe, Wallace L., & Jane Danielewicz (1986). Properties of Spoken and Written Language. In: Rosalind Horowitz and S.J.Samuels (eds.), *Comprehending Oral and Written Language.* New York: Academic Press, pp. 82-113.
43. Chen, C.F.E. (2006). The development of e-mail literacy: from writing to peers to writing to authority figures. *Language Learning and Technology*, Vol.10, No.2, May 2006, pp. 35-55.
44. Cherny, L. (1999) *Conversation and Community: Chat in a Virtual World.* Standford: CSLI Publications.
45. Chun, D. (1994). Using computer networking to facilitate the acquisition of interactive competence. *System,* 22(1), pp. 17-31.
46. Cigankova, N. (2005) The Influence of Digital Culture on EAP Learner Language Use in WebCT-Based University Courses. Lingvistikas didaktikas problēmas XIV-XV. Daugavpils: Daugavpils Saule, 9-14.
47. *CLAWS* (2007). part-of-speech tagger for English, UCREL, Lancaster University [online] Available from <http://www.comp.lancs.ac.uk/ucrel/claws/> accessed 16.02.2007.
48. Cobb, T. (1998). Breadth and depth of vocabulary acquisition with hands-on concordancing. [Electronic version]. *Computer Assisted Language Learning 12*, 345-360. Retrieved September 20, 2007, from <http://www.er.uqam.ca/nobel/r21270/cv/Breadth.htm>
49. Cobb, T. *Compleat Lexical Tutor*. [online] Available from <http://www.lextutor.ca/> Retrieved September 15, 2006.
50. Coffin, C., Curry, M.J., Goodman, S., Hewings, A., Lillis, T.M., Swann, J. (2003). *Teaching Academic Writing*. London and New York: Routledge.
51. Collot, M., and Belmore, N. (1996) Electronic language: a new variety of English. In: S.C. Herring (ed.), *Computer-mediated Communication: Linguistic, Social, and Cross-cultural Perspectives.* Amsterdam/Philadelphia: John Benjamins, pp. 13-28.
52. Connor, U. (1999). *Contrastive Rhetoric: Cross-cultural Aspects of Second-language Writing.* Cambridge: Cambridge University Press.
53. Connor, U., Mauranen, A. (1999). Linguistic analysis of grant proposals: European Union research grants. *English for Specific Purposes*, 18(1): pp. 47-62.
54. Cook, G. (1989). *Discourse.* Oxford: Oxford University Press.
55. Couture, B. (ed.) (1986). *Functional Approaches to Writing: Research Perspectives*. Norwood, N.J: Ablex.
56. Crookes, G. (1986). Towards a validated analysis of scientific text structure. *Applied Linguistics*, 7: pp. 57-70.
57. Crystal, D. (1997). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
58. Crystal, D. (2001). *Language and the Internet.* Cambridge: Cambridge University Press.
59. Csomay, E. (2002). Variation in academic lectures: interactivity and level of instruction. In: R.Reppen, S.M. Fitzmaurice, D.Biber (eds.) *Using Corpora*

to Explore Linguistic Variation, Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 203-224.
60. Davis, B. H., Brewer, J.P. (1997). *Electronic Discourse: Linguistic Individuals in Virtual Space.* Albany, NY: State University of New York Press.
61. Davis, B., Thiede, R. (2000). Writing into change: style shift in asynchronous electronic discourse. In: M.Warschauer, R.Kern (eds.) *Network-based Language Teaching,* Cambridge: Cambridge University Press, pp. 87-120.
62. Dijk, T. A. (1977). *Text and Context: Explorations in the Semantics and Pragmatics of Discourse.* London & New York: Longman. van
63. Dijk, T.A. (1997). Discourse as interaction in society. In: T.A. van Dijk (ed.), *Discourse as social interaction*. London: Sage pp. 1-37. van
64. Diniz, L. (2005). Comparative review: TextStat 2.5, ANTCONC 3.0, and compleat lxical tutor 4.0. *Language Learning and Technology,* 9 (3) pp. 22-27.
65. Dix, A., Finlay, J., Abowd, G., and Beale, R. (1993). (eds.). *Human-Computer Interaction* , Hemel Hempstead: Prentice Hall, pp. 96-97.
66. Dorning, N. (1998). Lernen mit dem Internet. In Ludwig J. Issing, P.Klimsa (Hrsg.) *Information und Lernen mit Multimedia.* BELIZ: Psychologie VerlagsUnion, pp. 306-336.
67. Edmonson, W. (1981). *Spoken Discourse: A Model for Analysis.* London & New York: Longman.
68. eEurope - An Information Society for all - Progress report for the Special European Council on Employment, Economic reforms and social cohesion towards a Europe based on innovation and knowledge Lisbon, 23 and 24 March 2000 *European Commission* (author) 52000DC0130.
69. Eggins, S. (2004). *An Introduction to Systemic Functional Linguistics*. New York, London: Continuum.
70. Elbow, P. (1998). *Writing with power: techniques for mastering the writing process.* New York and Oxford: Oxford University Press.
71. Ellis, R. (1994). *The Study of Second Language Acquisition.* Oxford: Oxford University Press, pp. 49-82.
72. Emmerson, P. (2004). *Email English*. MacMillan.
73. Enkvist, N.E. (1984). Contrastive Linguistics and Text linguistics. In: J.Fisiak (ed.) *Contrastive Linguistics, Prospects and Problems*. Berlin: Mouton, pp. 45-67.
74. Erikson, T. (1999). Persistent Conversation: An Introduction. *Journal of Computer-Mediated Communication* 4(4). Available at < http://jcmc.indiana.edu/vol4/issue4/ericksonintro.html> Accessed: 25.06.2006.
75. *E-Universitātes Projekts* (2002) Latvijas Universitāte. Teksts pieejams: http://www.lu.lv/e-universitaate/e-studijas/dokumenti/arhiivs/projekta_ietvars_30052002.doc <http://eur-lex.europa.eu/lv/dossier/dossier_13.htm#1>
76. Farneste, M. (2006). Benefits and drawbacks of using track changes in peer revision. Hellenic American Union. *EATAW Athens Conference 2005 Proceedings.* [CD-ROM].
77. Ferguson, Charles (1994). Dialect, register, and genre: working assumptions about conventionalization. In: Biber and Finegan (eds.) Sociolinguistic

Perspectives on Register. New York: Oxford, pp. 15-30.
78. Ferrara, K., Brunner, H., Whittemore, G. (1991). Interactive written discourse as an emergent register. *Written Communication*, 8(1), pp. 8-34.
79. Fillmore, C.J. (1981). Pragmatics and the description of discourse. In: Peter Cole (ed.) *Radical Pragmatic.* New York: Academic Press, pp. 143-166.
80. Firth J.R. (1935) The technique of semantics. *Transactions of the Pholological Society* (reprinted in J.R.Firth (1957), pp. 177-189.
81. Firth J.R. (1950) Personality and language in society. In Firth *Papers in Linguistics*, 1957, pp.177-189.
82. Firth J.R. (1951) Modes of meaning. In Firth *Papers in Linguistics* 1957, pp. 190-215.
83. Fitze, M. (2006). Discourse and participation in ESL face-to face and written electronic conferences. *Language Learning and Technology*, 10(1), pp. 67-86.
84. Flower, L. (1989). Cognition, Context and Theory Building. *College Composition and Communication*, 40, pp. 282-311.
85. Flowerdew, L. (2000). Investigating referential and pragmatic errors in a learner corpus. In: L.Burnard, T.McEnery (eds.) *Rethinking Language Pedagogy froma Corpus Perspective.* Frankfurt: Peter Lang, pp. 145-154.
86. Freadman, A., Medway, P (1994). *Genre and the New Rhetoric*. London: Tylor & Francis.
87. Fujimura, O., Erikson, D. (1997) In: W.J. Hardcastle, J.A. Laver (eds.) *Handbook of Phonetic Sciences,* Blackwell Publishers.
88. Garside, R. (1996). The robust tagging of unrestricted text: the BNC experience. In: J. Thomas and M. Short (eds.) *Using corpora for language research: Studies in the Honour of Geoffrey Leech.* Longman: London, pp. 167-180.
89. Garside, R., Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In: Garside, R., Leech, G., and McEnery, A. (eds.) *Corpus Annotation: Linguistic Information from Computer Text Corpora.* Longman: London, pp. 102-121.
90. Gee, J.P. (2002). *An Introduction to Discourse Analysis*. London and New York: Routledge.
91. Goody, J. (1977). *The Domestication of the Savage Mind.* Cambridge: Cambridge University Press.
92. Gosden, H. (1993). Discourse functions of subject in scientific research articles. *Applied Linguistics*, 14(1): pp. 56-75.
93. Granger, S. (1998) *Learner English on Computer.* Longman.
94. Granger, S. (2002). A bird's-eye view of learner corpus research. *In: S.Granger, J Hung, S.Petch-Tyson (eds.) Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching.* Amsterdam Philadelphia: John Benjamins.
95. Granger, S. (2004). Computer learner corpus research: current status and future prospects. In: U.Connor, T.A.Upton (eds.) *Applied Corpus Linguistics*. Amsterdam-New York, NY, pp. 123-145.
96. Granger, S., S. Tyson (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15 (1), pp. 17-27.
97. Greenbaum, S. (1996). *The Oxford English Grammar.* Oxford: Oxford University Press.

98. De Haan, P. (1992). The optimum corpus sample size? In: G. Leitner (ed.) *New Directions in English Language Corpora.* Mouton de Gruyter, Berlin and New York, pp. 3-19.
99. Hale, C., Scanlon, J. (1999). *Wired Style: Principles of English Usage in the Digital Age.* New York: Broadway Books.
100. Hall, R.A. (1964). *Introductory Linguistics*. Philadelphia: Chilton Books.
101. Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
102. Halliday, M.A.K (1970). Language structure and language function. In: (ed.) J.Lyons *New Horizons in Linguistics.* Harmandsworth, Middx: Peguin Books.
103. Halliday, M.A.K. (1985) Spoken and Written Language. Geelong, Vic.: Deakin University Press (reprinted bu OUP, 1989).
104. Halliday, M.A.K. (1996) Literacy and linguistics: a functional perspective. In: R.Hassan and G.Williams (eds.) *Literacy and Society*. London: Longman.
105. Halliday, M.A.K. (2004). *An Introduction to Functional Grammar*. Hoddar Arnold.
106. Halliday, M.A.K. and C.M.I. Matthiessen (2004) *An Introduction to Functional* Grammar, London: Hodder Arnold.
107. Halliday, M.A.K., Hasan, R. (1985, 1991) *Language, Context and Text: Aspects of Language in a Social-semantic Perspective.* Oxford: Oxford University Press.
108. Halliday, M.A.K., McIntosh, A. and Strevens, P. (1964). *The Linguistic Science and Language Teaching.* London: Longman.
109. Van Halteren, H. (1999a). Performance of taggers. In: H. Van Halteren (ed.) *Syntactic Wordclass Tagging*. Dordrecht, Kluwer, pp. 81-94.
110. Van Halteren, H. (1999b). Selection and operation of taggers. In: H.Van Halteren (ed.), *Syntactic Wordclass Tagging*. Dordrecht, Kluwer, pp. 95-104.
111. Harris, Z. (1952). Discourse Analysis. *Language* 28, pp. 1-30.
112. Hatch, E. (1992). *Discourse and Language Education.* New York: Cambridge University Press.
113. Herring, S.C. (2001). Computer-mediated discourse. In: D.Schifrin, D.Tannen, & H.E.Hamilton (eds.), *The Handbook of Discourse Analysis*. Malden, MA: Blackwell, pp. 612-706.
114. Hoey, M. P. (1991). *Patterns of Lexis in Text.* Oxford: Oxford University Press.
115. Hopper, A. (1994). Communications at the desktop. *Computer Networks and ISDN Systems,* 26, pp.1253-1265.
116. Hughes, R. (1996). *English in Speech and Writing*. London and New York: Routledge.
117. Hyland, K. (1996). Nurturing hedges in the ESP curriculum. *System*, 23(4): pp. 477-490.
118. Hyland, K. (1998). *Hedging in Scientific Research Articles*. Amsterdam: John Benjamins.
119. Hyland, K. (2000). *Disciplinary Discourses: Social Interactions in Academic Writing.* Harlow: Longman.
120. Hyland, K. (2002a). Directives: argument and engagement in academic writing. *Applied Linguistic*s 23/2, pp. 215-239.

121. Hyland, K. (2002b). *Teaching and Researching Writing*. Longman: Pearson Education.
122. Hyland, K. (2004). *Disciplinary Discourses.* The University of Michigan Press
123. Hymes, D. (1984). Sociolinguistics: stability and consolidation. *International Journal of the Sociology of Language* 45, pp. 39-45.
124. IDEELS (2006). University of Bremen, Germany. [Online] Available at: <http://www.ideels.uni-bremen.de/> Accessed: 30.01.2007.
125. Jakobson, R. (1960). Linguistics and poetics. In: Th.A.Sebeok (ed.), *Style in Language*. Cambridge, MA: MIT-Press, pp. 350-377.
126. Johansson, S. (1994). Encoding a corpus in machine-readable form: the approach of the text encoding initiative. In: B.T.S. Atkins; A.Zampolli (eds.) *Computational Approaches to the Lexicon.* Oxford: Oxford University Press.
127. Johnes, T., King, P. (eds.) (1991). Classroom Concordancing. *Special Edition: English Language Research Journal* 4.
128. Johnstone, B. (2002). *Discourse Analysis.* Blackwell Publishers.
129. Jones, K. (1990). *Simulations in language teaching.* Cambridge: Cambridge University Press.
130. Kantor, K. J. (1984). Classroom contexts and the development of writing intuitions: An ethnographic case study. In: R.Beach & L.S. Bridwell (eds.), *New directions in composition research*. New York: Guiford Press, pp. 72-94.
131. Kern, R. (1996). Computer-mediated Communication: Using E-mail Exchanges to Explore Personal Histories of Two Cultures. In M. Warschauer (Ed.), *Telecollaboration in Foreign Language Learning. Proceedings of the Hawaii Symposium* Technical Report # 21. Honolulu: University of Hawaii Press, pp. 105-119.
132. Kern, R. (2006). Perspectives on technology in learning and teaching languages. *TESOL Quarterly,* 40(1) pp. 183- 211.
133. Kiesler, S., Siegel, J., McGuire, T.W. (1984). Social psychological aspects of computer-mediated comunication. *American Psychologist,* 39, pp. 1123-34.
134. Kļaviņa S. (1980). *Statistika valodniecībā*. Rīga: LVU.
135. Kramiņa, I. (2000). *Linguo-didactic Theories Underlying Multi-purpose Langauge Acquisition.* Riga: Elpa: 2.
136. Kramiņa, I. Kursa apraksts: *Akadēmiskā angļu rakstu valoda III*. Rīga: LU [Tiešsaiste] pieejams: <http://www.lu.lv>
137. Kress, G., Leeuwen, T. (2001). *Multimodal Discourse: the Modes and Media of Contemporary Communication.* London: Arnold.
138. Lee, Davis Y.W. (2001). Genres, registers, text types, domains, and styles: clarifying the concepts and navigating a path through the BNC jungle. *Language Learning and Technology.* 5(3), pp. 37-72.
139. Leech, G. (1993). Corpus Annotation Schemes. *Literary and Linguistic Computing* 8(4): 275-281.
140. Leech, G., Garside, R., and Bryant, M. (1994). CLAWS4: The tagging of the British National Corpus. In: Proceedings of the 15th International Conference on Computational Linguistics (COLING 94) Kyoto, Japan, pp. 622-628.
141. Levy, M. (1997). *Computer-assisted Langauge Learning: Context and*

Conceptualization, Oxford: Clarendon Press.
142. Lyons, J. (1977) *Semantics,* Vol. 1& 2. Cambridge: Cambridge University Press.
143. Malinowski, B. (1923/46) The problem of Meaning in Primitive Languages. Supplement I to C.K. Ogden and I.A. Richards *The Meaning of Meaning* (8th edition, 1946). New York: Harcourt Brace & World, 296-336.
144. Martin, J. R. (1984). Language, Register and Genre. *In Children's Writing: Reader.* Geelong, Victoria: Deakin University Press.
145. Martin, J. R. (1992). *English Text: System and Structure.* Amsterdam: Benjamins.
146. Martin, J. R. (2003). Cohesion and texture. In: D.Schiffrin, D.Tannen, H.E.Hamillton (eds.), *The Handbook of Discourse Analysis* Blackwell Publishing, pp. 35-53.
147. Martin, J.R., Christine, F. and Rothery, J. (1987). Social processes in education: a reply to Sawyer and Watson (and others). In: I.Reid (eds.), *The Place of Genre in Learning: Current Debates*. Geelong, Australia: Deakin University Press, pp. 46-57.
148. Matsuda, P. K., Canagarajah, A. S., Harklau, L., Hyland, K., Warschauer, M. (2003). Changing currents in second language writing research. *Journal of Second Language Writing*, 12 (2) pp. 159-179.
149. McCarthy, M. (1991). *Discourse analysis for Language Teachers*. Cambridge: Cambridge University Press.
150. McCarthy, M. (2001). *Issues in Applied Linguistics.* Cambridge: Cambridge University Press.
151. McCarthy, M., Carter, R. (1994). *Language as Discourse: Perspectives for Language Teachers.* Harlow: Longman.
152. McEnery, A., Xiao, Z., Tono, Y. (2006). *Corpus-based Language Studies: an Advanced Resource Book*. London: Routledge.
153. Meunier, F. (1998). Computer Tools for the Analysis of Learner Corpora. In: S.Granger (ed.) *Learner English on Computer.* London and New York: London, pp. 19 – 37.
154. Meyer, C.F. (2002). *English Corpus Linguistics*. Cambridge: Cambridge University Press.
155. Miller, C.R. (1984). Genre as spcial action. *Quarterly Journal of Speech*, 70, pp. 157-78, also published in A.Freedman and P.Medway (eds.) (1994) *Genre and the New Rhetoric*. London: Taylor and Francis, pp. 23-42.
156. Milton, J., & E.S.C. Tsang (1993). A Corpus-based Study of Logical Connectors in EFL Students' Writing: Directions for Future Research. In: R.Pemberton & E.S.C. Tsang (eds.), *Studies in Lexis*. Hong Kong: The Hong Kong University of Science and Technology Language Centre, pp. 215-246.
157. Moffett, J. (1982) Writing, inner speech and mediation. *College English*, 44, pp. 231-44.
158. Moran, C., & Hawisher, G. E. (1998). The rhetorics and languages of electronic mail. In: I. Snyder (ed.), *Page to Screen: Taking Literacy into the Electronic Era*. London: Routledge, pp. 80-101.
159. Murray, D. E (1985). *A writer Teaches Writing.* 2nd edn. Boston, MA: Houghton Mifflin.
160. Murray, D. E. (1988). Computer-mediated communication: Implications for ESP. *English for Specific Purposes*, 7, pp. 3-18.

161. Murray, D. E. (1991) *Conversation for Action: The Computer Terminal as a Medium of Communication*. Amsterdam: John Benjamins.
162. Murray, D. E. (1995). *Knowledge Machines*. Singapore: Longman
163. Murray, D. E. (2000). Protean communication: the language of computer-mediated communication. *TESOL Qarterly,* 34, pp. 397-422.
164. Newby, H. (1977). In the fielf: Reflections on the study of Suffolk farm workers. In: C.Bell, H.Newby (eds.) *Doing Sociological Research*. London: Allen and Unvin.
165. Newkirk, T. (1984). Anatomy of a breakthrough: case study of a college freshman writer. In: R.Beach and L.S.Bridwell (eds.), *New Directions in Compoition Research* (pp.131-148). New York: Guiford Press.
166. Nuan, D. (1993). *Introducing Discourse* Analysis. London: Penguin.
167. Nystrand, M., Greene, S., Wiemelt, J. (1993). Where did composition studies come form? An intellectual history. *Written Communication*, 10. pp. 167-133
168. Olson, D., Torrance, N., Hildyard, A. (1985). *Literacy, Language and Learning.* Cambridge: Cambridge University Press.
169. Owston, R. D., Murphy, S., Wideman, H.H. (1992). The effects of word processing on students' writing quality and revision strategies. *Research in the Teaching of English*, 26, 249-276.
170. Par E-universitātes attīstības redzējumu LU - *LU Senāta lēmums* Nr. 38, 26/11/2001.
171. Paul, D., Charney, D. (1995). Introducing chaos (theory) into science and engineering. *Written Communication*, 12(4), pp. 396-438.
172. Pennington, M.C. (1996). *The Computer and the Non-native Writer: A Natural Partnership.* Cresskill, N.J: Hampton Press.
173. Pennington, M.C. (1999). Word processing and beyond: writing in an electronic medium. In: M.C. Pennington (ed.), *Writing in an Electronic Medium: Research with Languge Learners.* Houston, TX: Athelstan, pp. 1-26.
174. Pennington, M.C. (2003). The impact of the computer in second language writing. In: B.Kroll (ed.), *Exploring the Dynamics of Second Language Writing.* New York: Cambridge University Press, pp. 207-310.
175. Pennington, M.C. (2004). Electronic media in second language writing: an overview of tools and research findings. In: S.Fotos (ed.), *New Perspectives in CALL for Second Language Classroom*. Mahwah: Lawrence Erlbaum Associates, pp. 69-92.
176. Poos, D., Simpson, R. (2002). Cross-disciplinary comparison of hedging. In: R.Reppen, S.M.Fitzmaurice, B, Douglas (eds.), *Using Corpora to Explore Linguistic Variation*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 3-48.
177. Poynton, C. (1984) Names as vocatives: forms and functions. *Nottingham Linguistics Circular 13* (Special Issue on Systemic Linguistics), pp. 1-34.
178. *Programma E-Latvija 2005-2008* (2005) Latvijas Republika īpašu uzdevumu ministra elektroniskās pārvaldes lietās sekretariāts. Vidēja termiņa darbības programma laika posmam 2005-2008 Dokuments pieejams <http://www.eps.gov.lv/files/projekti/223progr_e-Latvija.pdf#search=%22E-Latvija%20projekts%22>
179. Quirk, R. S., Greenbaum, G., Leech, Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. London & New York: Longman.

180. Račevska, M., Kristapsone, S. (2000) Statistika psiholoģijas pētījumos. Rīga: Izglītības soļi.
181. Raimes, A. (1983). Tradition and Revolution in ESL Teaching. *TESOL Quarterly*, 17, pp. 535-552.
182. Raimes, A. (1985). What Unskilled ESL students Do as They Write: A Classroom Study of Composing. *TESOL Quarterly*, 19, pp. 229-258.
183. Rozenbergs, J. (2004) *The Stylistics of Latvian*. Rīga: LU Akadēmiskais apgāds.
184. Scott, M. (2004). *WordSmith Tools 4.0* [Online] Available from <http://lexically.net/downloads/version4/html/index.html> accessed: 15.01.2007.
185. Scott, M., & Tribble, C. (2006). *Textual Patterns: keyword and corpus analysis in language education.* Amsterdam: Benjamins.
186. Selfe, C.L. & Meyer, P.R. (1991). Testing claims for on-line conferences. *Written Communication,* 8(2), pp. 163-192.
187. Shaw, P., Liu, E.T.- K. (1998) What develops in the development of second-language writing? *Applied Linguistics* 19/2. Oxford: Oxford University Press.
188. Sinclair, J. (1995). Corpus typology – a framework for classification. In: G.Melchers and B.Warren (eds.) *Studies in Anglistics.* Almqvist and Wiksell International, Stockholm, pp. 17-33.
189. Slade, D. (1999). *Conversation, Grammar, and Genre: Gossiping in English.* Tokyo: AILA August World Congress.
190. Smith, B. (2005). The relationship between negotiated interaction, learner uptake, and lexical acquisition in task-based computer-mediated communication. *TESOL Quarterly,* 39 (1): pp. 33-58.
191. Sotillo, S.M. (2000). Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning and Technology,* Vol.4, No.1 pp. 82-119.
192. Spektors, A. (2001) Latviešu valodas datorfonda izveide. *Latvijas Zinātņu Akadēmijas Vēstis*. Sērija A. Nr.2 74-82. lpp.
193. Spencer, A. (1975). *Noun-Verb Combination in Law*. Birmingham: LSU, University of Aston in Birmingham.
194. Street, B. (1995). Social Literacies. Harlow: Blackweel.
195. Stubbs, M.W. (1983). *Discourse Analysis*. Oxford: Blackwell.
196. Stubbs, M.W. (1996). *Text and Corpus Analysis*. Oxford. Blackwell.
197. Sullivan, N., Pratt, E. (1996). A comparative study of two ESL writing environments: A computer-assisted classroom and a traditional oral classroom. *System,* 29, pp. 491-501.
198. Susser, B. (1993). ESL/EFL process writing with computers. *CAELL Journal*, 4(2), pp. 16-22.
199. Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
200. Swales, J. M., Ahmad,U., Chang, Y-Y., Chavez,D., Dressen D. and Seymour, R. (1998). Consider this: the role of imperatives in scholarly writing. *Applied Linguistics,* 19 (1): pp. 97-121.
201. Swales, J. M., Feak, Ch.B. (2001) *Academic Writing for Graduate Students.* The University of Michigan Press.
202. Sweet, H. (1964.) *The Practical Study of Languages.* London: Oxford University Press (Original work published 1899).

203. Tannen, D. (1982). Oral and literate strategies in spoken and written narratives. *Language*, 58: pp. 1-21.
204. Trask, R.L. (2005). *How to Write Effective Emails*. Penguin Books.
205. Tribble, Ch. (1996). *Writing.* Oxford: Oxford University Press.
206. Tribble, Ch. (2002). Corpora and corpus analysis: new windows on academic writing. In: J.Flowerdew (ed.), *Academic Discourse*. Harlow, Milan: Pearson Education, pp. 131-149.
207. Tribble, Ch., Granger, S. (1998) Lerner corpus data in the foreign language classroom: form-focused instruction and data-driven learning. In: S.Granger (ed.), *Learner English on Computer.* London, New York: Longman, pp.199-209.
208. Trosborg, A. (1997). Text typology: register, genre and text type. In: A.Trosborg (ed.), *Text Typology and Translation*. John Benjamins, pp. 3-23.
209. Valdmanis, J. (1999) Tonis valodas lietojuma reģistrā. *Linguistica Lettica*. 5, Rīga: LU Latviešu valodas institūts, 180-185 lpp.
210. Valdmanis, J. (2002) Register of language use in source and target language. - *Sastatāmā un Lietišķā valodniecība: kontrastīvie pētījumi.* Zinātniskie raksti. XI sēj. Atb. Redaktors Andrejs Veisbergs. Rīga: JUMI, 2002, pp. 63-71.
211. Ventola, E. (1984). *Orientation to Social Semiotics*. Oxford: Basil Blackwell.
212. Vinčela, Z. (2006) Creative Linguistic Variation in the Project IDEELS (Intercultural Dynamics in European Education Through Online Simulation) Simulation / *Zinātnisko rakstu krājums. Radoša Personība IV* . Rīga: RPIVA, 256–262.
213. Vinčela, Z. (2007) Text-based Computer-mediated Communication: Formal or Informal Writing? / *Pasaules Valodu skolotāju federācijas Ziemeļvalstu – Baltijas valstu starptautiskās konferences Zinātniskie raksti* (CD-R disc). – Rīga, Latvija, 290–297.
214. Vinčela, Z. (2008) Corpus analysis: Application of data-driven language studies / *Proceedings of The Third Baltic Conference on Human Language Technologies*. – Kaunas : Vytautas Magnus University, 327–333.
215. Virtanen, T. (1990) On the definitions of text and discourse. *Folia Linguistica* 24, No. 3-4 (1990): pp. 447-455.
216. Warschauer, M. (1996). Comparing face-to-face and electronic communication in the second language classroom, *CALICO Journal*, 13(2), pp. 7-26.
217. Warschauer, M. (1997). Computer-mediated collaborative learning: theory and practice. *Modern Language Journal*, 81, pp. 470-481.
218. Warschauer, M. (1999). *Electronic Literacies: Language, Culture and Power in Online Education.* Mahwah, NJ: Lawrence Erlbaum Associates.
219. White, R., Arndt, V. (1991). *Process Writing.* Harlow: Longman.
220. Whittaker, S. (2003). Theories and methods in mediated communication. In: A. C.Graesser, M.A.Gernsbacher, S.R. Goldman (eds.), *Handbook of Discourse Processes.* Mahwah, New Jersey, London: Lawrence Erlbaum Associates, Publishers, pp. 243-286.
221. Widdowson, H. G. (1979). Directions in the Teaching of Discourse. In: H.G.Widdowson (ed.), *Explorations in Applied Linguistics*. Oxford: Oxford University Press, 89 100. (1st publ. in S.P.Corder & E.Roulet (eds). 1973. Theoretical Linguistic Models inApplied Linguistics. Brussels: AIMAV &

Paris: Didier, pp. 65-76.
222. Widdowson, H. G. (2004). *Text, Context, Pretext.* Oxford: Oxford University Press.
223. Xiao, Z., McEnery, A. (2005) Two aproaches to genre analysis. *Journal of English Linguistics,* Vol.33/No, 1, March, pp. 62-82.
224. Yates, S. (1996). Oral and written linguistic aspects of computer conferencing: a corpus-based study. In: S.Herring (ed.) *Computer-mediated Communication: Linguistic, Social and Cross-cultural Perspectives*. Amsterdam: John Benjamins, pp. 30-46.
225. Zamel, V. (1982). Writing: the process of discovering meaning. *TESOL Quarterly*, 16(2), pp. 195-209.
226. Zamel, V. (1985). Responding to Student Writing. *TESOL Quarterly* 16(2), pp. 79-101.
227. Zamel, V. (1987). Recent Research on Writing Pedagogy. *TESOL Quarterly,* 21 (4), pp. 697-714.

APPENDICES

APPENDIX I: Urcel *claws7* tagset

APPGE	possessive pronoun. pre-nominal (e.g. my, your, our)
AT	article (e.g. the, no)
AT1	singular article (e.g. a. an, every)
BCL	before-clause marker (e.g. in order (that).in order (to))
CC	coordinating conjunction (e.g. and, or)
CCB	subordinating conjunction (e.g. if. Because, unless, so, for)
CSA	as (as conjunction)
CSN	than (as conjunction)
CST	that (as conjunction)
CSW	whether (as conjunction)
DA	after-determiner or post-determiner capable of pronominal function (e.g. such, former, same)
DA1	singular after-determiner (e.g. little, much)
DA2	plural after-determiner (e.g. few, several, many)
DAR	comparative after-determiner (e.g. more, less, fewer)
DAT	superlative after-determiner (e.g. most, least, fewest)
DB	before determiner or pre-determiner capable of pronominal function (all, half)
DB2	plural before-determiner (both)
DD	determiner (capable of pronominal function) (e.g any, some)
DD1	singular determiner (e.g. this, that, another)
DD2	plural determiner (these, those)
DDQ	wh-determiner (which, what)
DDQGE	wh-determiner. genitive (whose)
DDQV	wh-ever determiner. (whichever, whatever)
EX	existential there
FO	formula
FU	unclassified word
FW	foreign word
GE	germanic genitive marker - (' or's)
IF	for (as preposition)
II	general preposition
IO	of (as preposition)
IW	with. without (as prepositions)
JJ	general adjective
JJR	general comparative adjective (e.g. older, better, stronger)
JJT	general superlative adjective (e.g. oldest, best, strongest)
JK	catenative adjective (able in be able to, willing in be willing to)
MC	cardinal number.neutral for number (two, three..)
MC1	singular cardinal number (one)
MC2	plural cardinal number (e.g. sixes, sevens)
MCGE	genitive cardinal number. neutral for number (two's, 100's)
MCMC	hyphenated number (40-50. 1770-1827)
MD	ordinal number (e.g. first, second, next, last)
MF	fraction.neutral for number (e.g. quarters, two-thirds)
ND1	singular noun of direction (e.g. north, southeast)
NN	common noun. neutral for number (e.g. sheep, cod, headquarters)
NN1	singular common noun (e.g. book, girl)
NN2	plural common noun (e.g. books, girls)
NNA	following noun of title (e.g. M.A.)
NNB	preceding noun of title (e.g. Mr., Prof.)
NNL1	singular locative noun (e.g. Island, Street)
NNL2	plural locative noun (e.g. Islands, Streets)
NNO	numeral noun. neutral for number (e.g. dozen, hundred)
NNO2	numeral noun. plural (e.g. hundreds, thousands)
NNT1	temporal noun. singular (e.g. day, week, year)
NNT2	temporal noun. plural (e.g. days, weeks, years)
NNU	unit of measurement. neutral for number (e.g., in., cc)
NNU1	singular unit of measurement (e.g. inch, centimetre)
NNU2	plural unit of measurement (e.g. ins., feet)

NP	proper noun. neutral for number (e.g. IBM, Andes)
NP1	singular proper noun (e.g. London. Jane, Frederick)
NP2	plural proper noun (e.g. Browns, Reagans, Koreas)
NPD1	singular weekday noun (e.g. Sunday)
NPD2	plural weekday noun (e.g. Sundays)
NPM1	singular month noun (e.g. October)
NPM2	plural month noun (e.g. Octobers)
PN	indefinite pronoun. neutral for number (none)
PN1	indefinite pronoun. singular (e.g. anyone, everything, nobody, one)
PNQO	objective wh-pronoun (whom)
PNQS	subjective wh-pronoun (who)
PNQV	wh-ever pronoun (whoever)
PNX1	reflexive indefinite pronoun (oneself)
PPGE	nominal possessive personal pronoun (e.g. mine, yours)
PPH1	3rd person sing. neuter personal pronoun (it)
PPHO1	3rd person sing. objective personal pronoun (him. her)
PPHO2	3rd person plural objective personal pronoun (them)
PPHS1	3rd person sing. subjective personal pronoun (he. she)
PPHS2	3rd person plural subjective personal pronoun (they)
PPIO1	1st person sing. objective personal pronoun (me)
PPIO2	1st person plural objective personal pronoun (us)
PPIS1	1st person sing. subjective personal pronoun (I)
PPIS2	1st person plural subjective personal pronoun (we)
PPX1	singular reflexive personal pronoun (e.g., yourself, itself)
PPX2	plural reflexive personal pronoun (e.g., yourselves, themselves)
PPY	2nd person personal pronoun (you)
RA	adverb. after nominal head (e.g., else, galore)
REX	adverb introducing appositional constructions (namely, e.g.)
RG	degree adverb (very, so, too)
RGQ	wh- degree adverb (how)
RGQV	wh-ever degree adverb (however)
RGR	comparative degree adverb (more, less)
RGT	superlative degree adverb (most, least)
RL	locative adverb (e.g., alongside, forward)
RP	prep. adverb. particle (e.g., about, in)
RPK	prep. adv.. catenative (about in be about to)
RR	general adverb
RRQ	wh- general adverb (where, when, why, how)
RRQV	wh-ever general adverb (wherever, whenever)
RRR	comparative general adverb (e.g., better, longer)
RRT	superlative general adverb (e.g., best, longest)
RT	quasi-nominal adverb of time (e.g., now, tomorrow)
TO	infinitive marker (to)
UH	interjection (e.g., oh, yes, um)
VB0	be. base form (finite i.e. imperative, subjunctive)
VBDR	were
VBDZ	was
VBG	being
VBI	be. infinitive (To be or not... It will be ..)
VBM	am
VBN	been
VBR	are
VBZ	is
VD0	do. base form (finite)
VDD	did
VDG	doing
VDI	do. infinitive (I may do... To do...)
VDN	done
VDZ	does
VH0	have. base form (finite)

VHD	had (past tense)
VHG	having
VHI	have (infinitive)
VHN	had (past participle)
VHZ	has
VM	modal auxiliary (can, will, would, etc.)
VMK	modal catenative (ought, used)
VV0	base form of lexical verb (e.g., give, work)
VVD	past tense of lexical verb (e.g., gave, worked)
VVG	-ing participle of lexical verb (e.g., giving, working)
VVGK	-ing participle catenative (going in be going to)
VVI	infinitive (e.g. to give... It will work...)
VVN	past participle of lexical verb (e.g., given, worked)
VVNK	past participle catenative (e.g. bound in be bound to)
VVZ	-s form of lexical verb (e.g. gives, works)
XX	not. n't
ZZ1	singular letter of the alphabet (e.g. A, b)
ZZ2	plural letter of the alphabet (e.g. A's, b's)

APPENDIX II: Summary of the Applied Dimensions, Linguistic Features and The Examples From *NNSE* Corpus

Dimensions	**Linguistic felements**	**Examples from the students' corpus**
A **Informational versus involved production**	**Elements with positive loadings** + (1) Private verbs: *anticipate, assume, believe, conclude, decide, demonstrate, determine, discover, doubt, estimate, fear, feel, find, forget, guess, hear, hope, imagine, imply, indicate, infer, know, learn, mean, notice, prove, realize, recognize, remember, reveal, see, show, suppose, think, understand.*	They **see** this issue as vital. I **hope** I have shown you the main points.
	+ (2) Contractions excluding possessive form: *n't, 'll, 'd, 're, 've, 's, 'm*	**That's** what we meant. **Let's** keep in touch.
	+ (3) Present-tense verbs: base forms and third-person singular present verb forms.	Some people **prefer** traditional learning. Traditional learning **gives** students the possibility of face-to-face communication.
	+ (4) Second-person pronouns: *you, your, yourself, yourselves, yours*	All that **you** need is to have **your** computer with **you**. Shall we include **your** statements?
	+ (5) *Do* as proverb.	I'll **do** it. You **did** a great job!
	+ (6) Analytic negation *not*	People do **not** have to waste so much time on searching for information. We do **not** consider it to be a division.
	+ (7) Demonstrative pronouns not followed by a noun: *this, that, these, those*	By **this** I mean formal and informal writing. There are specialists who take care of **that**.
	+ (8) General emphatics: *for sure, a lot, such, real, just, really, most, more, so, do+verb*	It takes hours and hours to find something **really** useful. It was a pleasure to have **such a** nice discussion.
	+ (9) First person pronouns: *I, my,we, our, myself, ourselves, mine, ours*	While we are reading your paper you can read **ours**. **I** think if every student has at least one question it is very useful.
	+ (10) Pronoun *it*.	Yes, **it** would be a brilliant idea. Your idea of dividing the responsibilities of the agency is really very useful as **it** could make the data operation process more accurate and elaborate.
	+ (11) *Be* as main verb.	Hiking and travelling by bicycle **are** similar. What **is** your opinion on this matter?
	+ (12) Causative subordination: *because*	Students have all the opportunities to do a good job **because there is a variety of source materials**. We were 20 people and it was easier to get things done **because** we could share the tasks.
	+ (13) Discourse markers: *well, now, anyway, anyhow, anyways*	**Well,** we could, but it isn't needed. **Anyway**, studying at school is easier than at university. **Anyway,** we need to communicate with people.
	+ (14) Indefinite pronouns: *none, one, anyone, someone, somebody,*	**Everything** depends on a person. **One** has to try really hard.

	anybody, nobody, everything, nothing	**Nobody** can say that one method is better than the other. Good morning **everybody**!
	+ (15) General hedges: *about* [not as a preposition] *something like, more or less, almost, maybe, sort of, kind of*	It is **kind of** reasonable that people choose travelling by land. All these source materials are **more or less** reliable.
	+ (16) Amplifiers: absolutely, altogether, completely, enormously, entirely, extremely, fully, greatly, highly, intensely, perfectly, strongly, thoroughly, totally, utterly, very	We **completely** agree with the DAP policy vision of Northland. We **fully** support this point. Our delegation wants to thank you for a **very** productive and fruitful teleconference.
	+ (17) Sentence relatives/non-restrictive relative clauses.	The last point, **which is also important**, is the safety. In contrast to the individual work, **which is a characteristic feature of e-learning**, the collaborative conventional activities can provide new skills.
	+ (18) WH questions	**What** do you mean by government operations? **Why** do you think so? **Where** are you from?
	+ (19) Possibility modals: *can, could, may, might*	You **can** find all the information you need in this virtual laboratory. It **may** seem very comfortable and modern.
	+ (20) Nonphrasal coordination.	It was great **and we gained new experience!** Thank you for your greeting message, **and we are also looking forward to negotiations.**
	+ (21) WH clauses.	I didn't know **what to do with my independence**. Would you please specify **what you mean by this question**?
	+ (22) Final prepositions.	That is what I am thinking **about**. It allows visiting places you are mostly interested **in**.
	Elements with negative loadings - (23) Other nouns: all nouns excluding nominalizations (nouns ending in *-tion, -ment, -ness, -ity*)	At **schoo**l, **children** are taught how to write **letters**. A **person,** therefore, perhaps should have a special **talent** to acquire new **skills**.
	- (24) Word length	Average word length.
	- (25) Prepositions: *at, by, in, of* etc.	The data **on** a person or a group should be taken **from** the Database. Being a part **of** the Eutropean Federation, we expect to get support **from** other Eutropean countries.
	- (26) Type/token ratio	Standardized type/token ratio.
B **Explicit versus situation dependent reference**	**Elements with positive loadings** + (27) WH relative clauses	People **who choose to travel on their own** have to decide how they will reach their destination. Anyone **who suffers damage** is entitled to compensation.
	+ (28) Pied piping constructions	It does not matter **in which** particular field of life these writing skills will be used. We truly look forward to cooperation with the

		other members of UN and expect messages with shared interests **on which** we could base a common resolution.
	+ (29) Phrasal coordination	Informal papers have a lot of **verbs, short sentences and words** that show **emotions and feelings**. The information is stored by the public institutions, for example, **hospitals and banks**.
	+ (30) Nominalizations: all nouns ending in *-tion, -ment, -ness, -ity*	They have to read quite a lot to gain the **information**. In fact, the traditional learning and technology-enhanced one have the same **functions.**
	Elements with negative loadings + (31) All adverbs of time.	**Nowadays**, students have all the opportunities. We are glad to start our negations **today**. It is a pleasure to study from books printed **recently**.
	+ (32) All adverbs of place.	We will be **there.** I am **here**. Linda, are you **there**?
	+ (33) Other adverbs.	Writing and speaking are **closely** connected. Travelling by sea and by air provide people with an opportunity to get to the destination **quickly**.
C **Abstract versus nonabstract information**	**All elements with positive loadings** + (34) Conjuncts: *alternatively, altogether, consequently, conversely, eg, e.g., else, furthermore, hence, however, i.e., instead, likewise, moreover, namely, nevertheless, nonetheless, notwithstanding, otherwise, rather, similarly, that is, therefore, thus, viz, in (comparison, contrast, particular, addition, conclusion, consequence, sum, summary, any event, any case, other words), for example (instance), by contrast (comparison), as a result, (consequence), on the contrary, (other hand)*	**As a result,** we discovered that subjects at the University are more interesting. **Moreover,** e-learning is mostly one's own individual work. **On the other hand**, the tasks that are provided at school and university are very different. **However**, these regulations apply only to information held by public authorities.
	+ (35) Agentless passives.	Formal communication **is** basically **used** at work. The commission **will be divided** into several specific departments.
	+ (36) BY-pasives.	These stylistic differences **are studied by linguists**. The celebration was organized **by the non-governmental organizations**.
	+ (37) Past participal WHIZ deletions.	Formal writing is a kind of writing **developed** for definite purpose. Education should not ignore the opportunities **offered** by modern technologies. There are buildings **designed** in Baroque and Classical styles.
	+ (38) Other adverbial subordinators: *since, while, whilst, whereupon, whereas, whereby, such that, so that, inasmuch as, forasmuch as, insofar as, insomuch as, as long as, as soon as*	E-studies require a lot of individual work **since** all the theory is studied on one's own. These materials range from books to electronic sources, **so that** one can choose the bets and the most suitable.

APPENDIX III

Descriptive Statistics: Collaborative Statements

Dimension	Features	Mean	Std dev.	Min.	Max.	Range	Score	Dimension Scores
A Involved versus Informational Production production	+ (1) Private verbs	5.76	1.84	3.88	7.55	3.68	-8.85	
	+ (2) Contractions	0	0	0	0	0	-0.95	
	+ (3) Present tense v.	45.02	1.19	44.12	46.38	2.25	-1.58	
	+ (4) Second person pr.	0	0	0	0	0	-1.11	
	+ (5) Do as a pro-verb	0.21	0.37	0	0.64	0.64	-0.7	
	+ (6) Analytic negation	3.45	0.36	3.23	3.87	0.63	-1.17	
	+ (7) Demonstrative pr.	0.79	0.24	0.64	1.07	0.43	+1.06	
	+ (8) General emphatics	2.59	1.12	1.29	3.24	1.95	-0.98	
	+ (9) First person pr.	8.13	4.8	2.58	11.03	8.44	-1.05	
	+ (10) Pronoun *it*	4.53	1.28	3.23	5.81	2.57	-1.51	
	+ (11) *Be* as a main verb	4.24	0.32	3.89	4.52	0.62	-1.05	
	+ (12) Causative sub.	0.21	0.37	0	0.64	0.64	-1.05	
	+ (13) Discourse markers	5.75	1.83	3.87	7.55	3.67	-0.46	
	+ (14) Indefinite pr.	0.57	0.54	0	1.07	1.07	-1.48	
	+ (15) Hedges	0	0	0	0	0	-1.15	
	+ (16) Amplifiers	0	0	0	0	0	-1.78	
	+ (17) Sentence relatives	0.43	0.37	0	0.64	0.64	-0.86	
	+ (18) WH questions	0	0	0	0	0	-0.47	
	+ (19) Possibility modals	3.52	1.52	2.15	5.16	3.01	-1.65	
	+ (20) Non-phrasal coord.	8.98	2.23	6.48	10.78	4.29	-0.58	
	+ (21) WH clauses	0.57	0.54	0	1.07	1.07	-0.83	
	+ (22) Final prepositions	0	0	0	0	0	-0.37	
	- (23) Other nouns	235.09	17.91	218.98	254.38	35.39	+1.15	
	- (24) Word length	5.38	0.06	5.31	5.44	0.13	+1.80	
	- (25) Prepositions	115.04	17.46	96.89	131.73	34.83	+1.34	
	- (26) Type/token ratio	31.35	1.62	30.2	32.5	2.3	-1.16	**-31.70**
B Explicit versus Situation Dependent Reference	+ (27) WH relative clauses	3.81	1.07	2.58	4.54	1.95	+1.35	
	+ (28) Pied piping constr.	1.22	0.66	0.64	1.94	1.3	+0.47	
	+ (29) Phrasal coordin.	24.46	5.22	21.31	30.49	9.18	+0.67	
	+ (30) Nominalizations	93.14	8.28	88.25	102.71	14.45	+2.01	
	- (31) Time adverbs	0.21	0.37	0	0.64	0.64	-1.52	
	- (32) Place adverbs	1.79	0.44	1.29	2.15	0.85	-0.53	
	- (33) Other adverbs	44.08	12.43	34.51	58.13	23.61	-1.66	**+8.21**
C Abstract versus Nonabstract Information	+ (34) Conjuncts	4.45	0.64	3.89	5.16	1.27	-0.17	
	+ (35) Agentless passives	12.15	0.72	11.62	12.97	1.35	+0.42	
	+ (36) BY-passives	1.22	0.65	0.64	1.93	1.28	+0.51	
	+ (37) Past part. WHIZ d.	7.12	2.32	5.19	9.7	4.51	+1.81	
	+ (38) Other adv. Sub.	0	0	0	0	0	-1.1	**+1.47**

APPENDIX IV

Descriptive Statistics: Collaborative E-Letters

Dimension	Features	Mean	Std dev.	Min.	Max.	Range	Scores	Dimension scores
A Involved versus Informational Production production	+ (1) Private verbs	13.32	15.94	2.05	24.59	22.60	-0.45	
	+ (2) Contractions	1.53	1.69	0.34	2.73	2.39	-0.57	
	+ (3) Present tense v.	47.39	10.24	40.15	54.64	14.49	-1.46	
	+ (4) Second person pr.	27.75	26.44	9.05	46.44	37.39	+0.9	
	+ (5) Do as a pro-verb	1.53	1.69	0.34	2.73	2.39	+0.08	
	+ (6) Analytic negation	10.5	8.33	4.61	16.39	11.78	+0.24	
	+ (7) Demonstrative pr.	3.41	2.89	1.36	5.46	4.09	-0.24	
	+ (8) General emphatics	1.62	1.56	0.51	2.73	2.21	-1.24	
	+ (9) First person pr.	31.16	25.47	13.15	49.18	36.02	-0.08	
	+ (10) Pronoun *it*	7.34	1.2	6.49	8.19	1.7	-0.97	
	+ (11) *Be* as a main verb	15.71	2.9	13.66	17.76	4.1	-0.49	
	+ (12) Causative sub.	0.08	0.12	0	0.17	0.17	-1.12	
	+ (13) Discourse markers	13.32	15.93	2.05	24.59	22.53	-0.46	
	+ (14) Indefinite pr.	1.87	1.2	1.02	2.73	1.7	-1.04	
	+ (15) Hedges	1.53	1.69	0.34	2.73	2.39	+0.2	
	+ (16) Amplifiers	3.07	3.38	0.68	5.46	4.78	-0.59	
	+ (17) Sentence relatives	1.87	1.2	1.02	2.73	1.7	+0.53	
	+ (18) WH questions	0.08	0.12	0	0.17	0.17	-0.46	
	+ (19) Possibility modals	5.8	0.48	5.46	6.15	0.68	-0.91	
	+ (20) Non-phrasal coord.	12.8	8.93	6.49	19.12	12.63	+0.38	
	+ (21) WH clauses	1.53	1.69	0.34	2.73	2.39	-0.02	
	+ (22) Final prepositions	0	0	0	0	0	-0.37	
	- (23) Other nouns	212.65	12.24	203.99	221.31	17.31	+0.43	
	- (24) Word length	5.01	0.16	4.9	5.13	0.23	+0.64	
	- (25) Prepositions	105.81	1.04	105.07	106.55	1.48	+0.9	
	- (26) Type/token ratio	36.12	0.87	35.5	36.74	1.24	+0.27	**-8.14**
B Explicit versus Situation Dependent Reference	+ (27) WH relative clauses	2.56	0.24	2.39	2.73	0.34	+0.36	
	+ (28) Pied piping constr.	2.98	0.36	2.73	3.24	0.51	+2.78	
	+ (29) Phrasal coordin.	13.32	7.25	8.19	18.45	10.25	-0.47	
	+ (30) Nominalizations	57.05	22.72	40.98	73.12	32.14	+0.55	
	- (31) Time adverbs	1.53	1.69	0.34	2.73	2.39	-0.91	
	- (32) Place adverbs	2.30	0.6	1.87	2.73	0.85	-0.24	
	- (33) Other adverbs	52.35	7.1	47.32	57.37	10.05	-1.37	**+5.74**
C Abstract versus Nonabstract Information	+ (34) Conjuncts	5.8	0.48	5.46	6.15	0.68	+0.25	
	+ (35) Agentless passives	14.34	8.7	8.19	20.5	12.3	+0.83	
	+ (36) BY-passives	0	0	0	0	0	-0.9	
	+ (37) Past part. WHIZ d.	5.55	3.74	2.9	8.19	5.29	+1.2	
	+ (38) Other adv. Sub.	0	0	0	0	0	-1.1	**+0.28**

APPENDIX V

Descriptive Statistics: Collaborative Messages of Teleconferences

Dimension	Features	Mean	Std dev.	Min.	Max.	Range	Scores	Dimension Scores
A Involved versus Informational Production production	+ (1) Private verbs	12.39	0.76	11.52	12.97	1.45	-0.62	
	+ (2) Contractions	1.81	0.6	1.4	2.51	1.1	-0.5	
	+ (3) Present tense v.	79.36	4.21	75.29	83.71	8.41	+0.1	
	+ (4) Second person pr.	14.29	2.48	12.66	17.15	4.48	-0.08	
	+ (5) Do as a pro-verb	0.76	0.06	0.7	0.83	0.13	-0.37	
	+ (6) Analytic negation	4.45	0.55	3.84	4.92	1.08	-0.97	
	+ (7) Demonstrative pr.	2.44	0.42	2.11	2.92	0.81	-0.55	
	+ (8) General emphatics	1.96	1.39	0.83	3.51	2.68	-1.15	
	+ (9) First person pr.	43.82	7.86	36.4	52	15.67	+0.46	
	+ (10) Pronoun *it*	8.55	1.23	7.74	9.98	2.24	-0.74	
	+ (11) *Be* as a main verb	24.96	1.85	23.04	26.74	3.7	-0.04	
	+ (12) Causative sub.	0.39	0.38	0	0.76	0.76	-0.96	
	+ (13) Discourse markers	12.38	0.76	11.52	12.97	1.44	0	
	+ (14) Indefinite pr.	2.02	0.32	1.67	2.3	0.63	-0.99	
	+ (15) Hedges	0.63	0.18	0.41	0.76	0.34	-0.59	
	+ (16) Amplifiers	3.89	1.66	2.09	5.37	3.28	-0.27	
	+ (17) Sentence relatives	0.37	0.35	0	0.7	0.7	-0.92	
	+ (18) WH questions	4.08	0.21	3.84	4.22	0.38	+0.04	
	+ (19) Possibility modals	8.5	1.5	7.03	10.04	3	-0.04	
	+ (20) Non-phrasal coord.	7.6	1.5	6.14	9.14	3	-0.92	
	+ (21) WH clauses	1.28	0.46	0.76	1.67	0.9	-0.23	
	+ (22) Final prepositions	0.13	0.24	0	0.41	0.41	-0.32	
	- (23) Other nouns	197.1	9.59	190.7	208.14	17.43	-0.06	
	- (24) Word length	4.98	0.17	4.78	5.12	0.34	+0.54	
	- (25) Prepositions	97.47	7.05	90.79	104.85	14.06	+0.51	
	- (26) Type/token ratio	39.25	4.09	34.65	42.5	7.85	+1.33	**-11.98**
B Explicit versus Situation Dependent Reference	+ (27) WH relative clauses	3.59	0.55	3.07	4.18	1.11	+1.17	
	+ (28) Pied piping constr.	0.9	0.3	0.7	1.25	0.55	+0.05	
	+ (29) Phrasal coordin.	12.22	0.47	11.71	12.66	0.95	-0.58	
	+ (30) Nominalizations	62.68	10.89	56.29	75.26	18.97	+0.78	
	- (31) Time adverbs	4.82	0.6	4.18	5.37	1.19	+0.61	
	- (32) Place adverbs	2.54	0.25	2.3	2.81	0.51	-0.11	
	- (33) Other adverbs	64.4	6.04	57.7	69.45	11.75	-0.93	**+1.85**
C Abstract versus Nonabstract Information	+ (34) Conjuncts	2.79	0.28	2.51	3.07	0.56	-0.7	
	+ (35) Agentless passives	14.59	3.08	11.96	17.99	6.02	+0.88	
	+ (36) BY-passives	1.53	0.13	1.4	1.67	0.26	+0.87	
	+ (37) Past part. WHIZ d.	3.61	1.36	2.3	5.02	2.71	+0.44	
	+ (38) Other adv. Sub.	0.76	0.06	0.7	0.83	0.13	+0.18	**+1.67**

APPENDIX VI

Descriptive Statistics: Collaborative Messages of Post-Simulation Conferences

Dimension	Features	Mean	Std dev.	Min.	Max.	Range	Scores	Dimension scores
A Involved versus Informational Production production	+ (1) Private verbs	37.13	9.03	26.72	42.86	16.14	+3.96	
	+ (2) Contractions	11.19	3.85	7.60	15.27	7.66	+1.8	
	+ (3) Present tense v.	108.42	16.63	92.86	125.95	33.10	+1.63	
	+ (4) Second person pr.	39.92	7.03	32.14	45.80	13.66	+1.78	
	+ (5) Do as a pro-verb	4.99	2.26	3.57	7.60	4.03	+2.1	
	+ (6) Analytic negation	11.11	3.34	7.63	14.29	6.65	+0.36	
	+ (7) Demonstrative pr.	7.46	0.28	7.14	7.63	0.49	+0.91	
	+ (8) General emphatics	11.11	3.35	7.60	14.29	6.68	+1.24	
	+ (9) First person pr.	68.23	13.49	53.43	79.84	26.41	+1.5	
	+ (10) Pronoun *it*	9.92	2.00	7.63	11.40	3.77	-0.47	
	+ (11) *Be* as a main verb	77.13	8.95	71.42	87.45	16.02	+2.52	
	+ (12) Causative sub.	3.65	3.57	0	7.14	7.14	+0.71	
	+ (13) Discourse markers	4.92	1.92	3.80	7.14	3.34	+2.47	
	+ (14) Indefinite pr.	8.72	2.33	7.14	11.40	4.26	+1.26	
	+ (15) Hedges	3.73	0.14	3.57	3.82	0.25	+2.15	
	+ (16) Amplifiers	4.99	2.26	3.57	7.60	4.03	-0.3	
	+ (17) Sentence relatives	0	0	0	0	0	-1.28	
	+ (18) WH questions	23.73	9.21	17.85	34.35	16.49	+2.53	
	+ (19) Possibility modals	11.27	4.06	7.14	15.27	8.12	+0.84	
	+ (20) Non-phrasal coord.	11.11	3.35	7.60	14.29	6.68	-0.04	
	+ (21) WH clauses	0	0	0	0	0	-1.31	
	+ (22) Final prepositions	8.57	5.39	3.81	14.28	10.47	+2.38	
	- (23) Other nouns	134.10	48.69	201.52	290.07	88.58	+1.12	
	- (24) Word length	4.49	0.16	4.3	4.61	0.31	-1.1	
	- (25) Prepositions	72.94	14.16	64.64	89.29	24.65	-0.64	**+27.36**
B Explicit versus Situation Dependent Reference	+ (27) WH relative clauses	0	0	0	0	0	-1.67	
	+ (28) Pied piping constr.	0	0	0	0	0	-1.14	
	+ (29) Phrasal coordin.	3.73	0.14	3.57	3.82	0.25	-1.44	
	+ (30) Nominalizations	11.20	3.85	7.60	15.27	7.66	-1.31	
	- (31) Time adverbs	6.19	2.07	3.82	7.60	3.79	+1.23	
	- (32) Place adverbs	6.27	2.34	3.57	7.63	4.06	+1.97	
	- (33) Other adverbs	116.82	2.21	114.29	118.32	4.03	+0.9	**-9.66**
C Abstract versus Nonabstract i Information	+ (34) Conjuncts	0	0	0	0	0	-1.60	
	+ (35) Agentless passives	0	0	0	0	0	-1.82	
	+ (36) BY-passives	1.27	2.20	0	3.80	3.80	-0.9	
	+ (37) Past part. WHIZ d.	0	0	0	0	0	-0.97	
	+ (38) Other adv. Sub.	0	0	0	0	0	-1.1	**-6.39**

APPENDIX VII

Descriptive Statistics: Individual Messages of Asynchronous Discussions

Dimension	Features	Mean	Std dev.	Min.	Max.	Range	Scores	Dimension Scores
A Involved versus Informational Production production	+ (1) Private verbs	19.40	5.17	15.71	26.95	11.24	+0.68	
	+ (2) Contractions	6.07	3.47	3.05	10.96	7.91	+0.54	
	+ (3) Present tense v.	76.45	5.46	70.25	80.86	10.61	+0.01	
	+ (4) Second person pr.	15.64	7.61	9.29	28.71	19.42	+0.02	
	+ (5) Do as a pro-verb	2.00	0.97	1.39	3.67	2.28	+0.35	
	+ (6) Analytic negation	12.06	6.67	5.50	21.75	16.25	+0.54	
	+ (7) Demonstrative pr.	2.27	1.17	0.61	3.55	2.94	-0.61	
	+ (8) General emphatics	4.67	1.27	3.20	5.89	2.70	-0.44	
	+ (9) First person pr.	34.85	6.79	25.65	42.76	17.12	+0.08	
	+ (10) Pronoun *it*	17.81	1.18	16.61	19.55	2.94	+1.04	
	+ (11) *Be* as a main verb	15.87	2.23	12.33	18.33	5.99	-0.48	
	+ (12) Causative sub.	2.76	1.63	1.07	4.94	3.87	+0.26	
	+ (13) Discourse markers	0.72	0.83	0.00	1.83	1.83	-0.03	
	+ (14) Indefinite pr.	4.79	1.82	2.44	7.54	5.09	-0.06	
	+ (15) Hedges	1.60	0.64	1.07	2.31	1.24	+0.27	
	+ (16) Amplifiers	7.02	2.25	5.54	11.00	5.46	+0.97	
	+ (17) Sentence relatives	1.25	0.60	0.46	1.96	1.51	-0.08	
	+ (18) WH questions	1.67	0.60	0.61	2.06	1.44	+0.33	
	+ (19) Possibility modals	8.84	1.05	7.62	10.05	2.43	+0.06	
	+ (20) Non-phrasal coord.	10.42	3.52	7.50	16.49	8.99	-0.22	
	+ (21) WH clauses	2.93	1.20	1.49	4.28	2.79	-2.1	
	+ (22) Final prepositions	0.14	0.21	0.00	0.46	0.46	-0.32	
	- (23) Other nouns	181.15	10.73	164.32	191.60	27.28	-0.57	
	- (24) Word length	4.67	0.27	4.34	4.91	0.57	-0.22	
	- (25) Prepositions	78.97	6.91	72.64	89.46	16.83	-0.35	
	- (26) Type/token ratio	35.88	3.10	33.00	40.32	7.32	+0.27	**+1.68**
B Explicit versus Situation Dependent reference	+ (27) WH relative clauses	1.31	0.58	0.61	2.06	1.44	-0.64	
	+ (28) Pied piping constr.	0.85	0.48	0.46	1.49	1.03	-0.02	
	+ (29) Phrasal coordin.	12.04	3.16	7.77	14.66	6.89	-0.6	
	+ (30) Nominalizations	32.85	10.62	21.24	43.93	22.69	-0.44	
	- (31) Time adverbs	2.92	1.09	1.30	4.28	2.98	-0.28	
	- (32) Place adverbs	3.45	1.27	1.83	4.47	2.64	+0.39	
	- (33) Other adverbs	106.30	16.62	87.68	124.04	36.36	+0.56	**-2.37**
C Abstract versus Nonabstract Information	+ (34) Conjuncts	5.03	1.86	3.39	7.94	4.55	0	
	+ (35) Agentless passives	12.00	3.40	7.87	16.79	8.92	+0.4	
	+ (36) BY-passives	0.65	0.46	0.00	1.12	1.12	-0.14	
	+ (37) Past part. WHIZ d.	2.14	0.53	1.54	2.68	1.14	-0.14	
	+ (38) Other adv. Sub.	0.65	0.38	0.19	1.23	1.05	-0.01	**+0.11**

APPENDIX VIII

Descriptive Statistics: Academic Pre-essays

Dimensions	Features	Mean	Std dev.	Min.	Max.	Range	Scores	Dimension scores
A Involved versus Informational Production production	+ (1) Private verbs	20.7	4.30	16.23	25.33	9.10	+0.90	
	+ (2) Contractions	4.96	2.14	2.36	6.96	4.60	+0.27	
	+ (3) Present tense v.	86.85	10.97	77.82	103.72	25.90	+0.54	
	+ (4) Second person pr.	15.94	8.02	2.88	24.45	21.58	+0.04	
	+ (5) Do as a pro-verb	1.03	0.46	0.26	1.43	1.17	-0.23	
	+ (6) Analytic negation	12.74	3.99	9.95	19.67	9.73	+0.68	
	+ (7) Demonstrative pr.	9.54	1.10	8.56	11.421	2.86	+1.63	
	+ (8) General emphatics	9.10	2.04	7.19	12.48	5.29	+0.72	
	+ (9) First person pr.	46.35	21.71	19.50	74.33	54.83	+0.56	
	+ (10) Pronoun *it*	18.19	2.48	15.56	21.46	5.90	+1.12	
	+ (11) *Be* as a main verb	25.15	4.75	21.37	33.01	11.64	-0.03	
	+ (12) Causative sub.	3.90	0.69	2.75	4.54	1.79	+0.85	
	+ (13) Discourse markers	0.16	0.98	0.00	0.26	0.26	-0.37	
	+ (14) Indefinite pr.	7.40	1.47	5.41	9.10	3.68	+0.82	
	+ (15) Hedges	0.92	0.36	0.52	1.43	0.91	-0.44	
	+ (16) Amplifiers	5.55	1.69	3.07	7.33	4.26	+0.37	
	+ (17) Sentence relatives	2.81	0.29	2.45	3.09	0.63	+1.43	
	+ (18) WH questions	0.70	0.41	0.40	1.40	0.99	-0.37	
	+ (19) Possibility modals	10.11	2.35	6.17	12.23	6.06	+0.49	
	+ (20) Non-phrasal coord.	14.66	4.65	9.26	20.46	11.20	+0.86	
	+ (21) WH clauses	2.34	0.60	1.69	3.01	1.32	+0.67	
	+ (22) Final prepositions	0.51	0.10	0.41	0.67	0.26	-0.2	
	- (23) Other nouns	167.67	11.93	152.66	185.32	32.65	-1.01	
	- (24) Word length	4.48	0.32	4.00	4.77	0.77	-0.78	
	- (25) Prepositions	57.09	4.65	51.41	62.39	10.98	-1.38	
	- (26) Type/token ratio	32.60	1.67	30.00	34.00	4.00	-0.63	**+14.06**
B Explicit versus Situation Dependent Reference	+ (27) WH relative clauses	2.33	0.45	1.70	2.86	1.16	+0.18	
	+ (28) Pied piping constr.	0.59	0.23	0.27	0.92	0.65	-0.29	
	+ (29) Phrasal coordin.	27.90	10.06	16.77	41.89	25.12	+0.91	
	+ (30) Nominalizations	37.76	7.09	30.32	45.94	15.62	-0.23	
	- (31) Time adverbs	4.90	1.72	3.81	7.92	4.11	+0.64	
	- (32) Place adverbs	1.76	1.40	0.52	4.16	3.64	-0.54	
	- (33) Other adverbs	114.60	5.91	108.35	121.01	12.66	+0.86	**-0.55**
C Abstract versus Nonabstract Information	+ (34) Conjuncts	4.96	2.30	2.45	8.67	6.22	-0.01	
	+ (35) Agentless passives	7.83	4.51	4.16	13.87	9.71	-0.37	
	+ (36) BY-passives	0.32	0.20	0.00	0.52	0.52	-0.53	
	+ (37) Past part. WHIZ d.	0.50	0.70	0.00	1.44	1.44	-0.69	
	+ (38) Other adv. Sub.	0.92	0.28	0.52	1.31	0.78	+0.43	**-1.17**

APPENDIX IX

Descriptive Statistics for Post-essays

Dimension	Features	Mean	Std dev.	Min.	Max.	Range	Scores	Dimension scores
A Involved versus Informational Production production	+ (1) Private verbs	12.10	2.29	9.62	14.69	5.07	-0.69	
	+ (2) Contractions	0.57	0.45	0.09	1.18	1.09	-0.81	
	+ (3) Present tense v.	74.29	6.92	62.67	79.7	17.03	-0.1	
	+ (4) Second person pr.	4.57	4.08	0.54	10.18	9.64	-0.78	
	+ (5) Do as a pro-verb	0.10	0.09	0	0.18	0.18	-0.77	
	+ (6) Analytic negation	8.03	1.32	5.96	9.31	3.35	-0.26	
	+ (7) Demonstrative pr.	2.70	0.73	1.43	3.25	1.82	-0.48	
	+ (8) General emphatics	9.32	0.77	8.36	10.19	1.83	+0.77	
	+ (9) First person pr.	5.66	3.88	2.21	11.78	9.57	-1.16	
	+ (10) Pronoun *it*	11.65	1.89	8.97	13.96	4.99	-0.14	
	+ (11) *Be* as a main verb	22.70	4.26	15.81	26.62	10.81	-0.15	
	+ (12) Causative sub.	2.50	0.9	1.04	3.2	2.16	+0.12	
	+ (13) Discourse markers	0.16	0.16	0	0.36	0.36	-0.36	
	+ (14) Indefinite pr.	6.10	1.04	4.57	7.42	2.84	+0.38	
	+ (15) Hedges	0.99	0.44	0.52	1.45	0.93	-0.27	
	+ (16) Amplifiers	4.73	1.01	3.27	5.84	2.58	+0.05	
	+ (17) Sentence relatives	1.64	0.41	1.17	2.21	1.04	+0.30	
	+ (18) WH questions	0.38	0.3	0.09	0.87	0.78	-0.43	
	+ (19) Possibility modals	9.78	3.19	5.4	13.72	8.32	+0.36	
	+ (20) Non-phrasal coord.	11.90	2.31	8.36	13.94	5.58	+0.15	
	+ (21) WH clauses	1.11	0.39	0.87	1.8	0.93	-0.38	
	+ (22) Final prepositions	0.09	0.09	0	0.18	0.18	-0.34	
	- (23) Other nouns	201.79	28.48	152	222.47	70.52	+0.09	
	- (24) Word length	4.86	0.10	4.75	4.97	0.22	+0.19	
	- (25) Prepositions	100.40	9.83	83.03	106.36	23.33	+0.64	
	- (26) Type/token ratio	34.70	2.09	31.47	37.12	5.65	-0.10	**- 4.17**
B Explicit versus Situation Dependent Reference	+ (27) WH relative clauses	1.84	0.59	1.08	2.47	1.39	-0.21	
	+ (28) Pied piping constr.	0.36	0.21	0.18	0.65	0.47	-0.3	
	+ (29) Phrasal coordin.	21.31	4.83	14.57	25.74	11.18	+0.63	
	+ (30) Nominalizations	32.42	8.6	21.76	41.15	19.39	-0.44	
	- (31) Time adverbs	3.15	1.58	1.44	5.67	4.23	-0.17	
	- (32) Place adverbs	1.72	0.52	1.17	2.33	1.16	-0.56	
	- (33) Other adverbs	93.81	6.34	88.79	104.55	15.76	+0.11	**+0.30**
C Abstract versus Nonabstract Information	+ (34) Conjuncts	9.43	2.27	5.4	10.76	5.37	+1.41	
	+ (35) Agentless passives	9.60	3.42	5.05	13.46	8.4	-0.04	
	+ (36) BY-passives	0.67	0.67	0.18	1.82	1.64	-0.13	
	+ (37) Past part. WHIZ d.	1.60	0.74	0.9	2.62	1.72	-0.34	
	+ (38) Other adv. Sub.	1.42	0.6	0.72	2.35	1.63	+1.25	**+2.15**