UNIVERSITY OF LATVIA
FACULTY OF MODERN LANGUAGES

NATAĻJA CIGANKOVA

LINGVISTISKAS VARIĀCIJAS ANGĻU VALODAS ELEKTRONISKAJĀ AKADĒMISKAJĀ DISKURSĀ

LINGUISTIC VARIATION IN ENGLISH COMPUTER MEDIATED ACADEMIC DISCOURSE

Doctoral Dissertation

Advisor - Dr.habil.phil.,
professor I.Kramiņa

RIGA, 2009

TABLE OF CONTENTS

ABBREVIATIONS

(Alphabetically)

ADV	adverb
ADVS	adverbial subordinator
AL	applied linguistics
ALPASS	agentless passive
AMP	amplifier
ANEG	analytic negation
ATADJ	attributive adjective
BE	*be* as main verb
BYPASS	*By* passive
CHAT	synchronous computer-mediated conference (chat)
CMAD	computer-mediated academic discourse
CLC	clausal coordination
CMC	computer-mediated communication
CMD	computer-mediated discourse
CMDA	computer-mediated discourse analysis
COND	conditional adverbial subordinator
CONJ	conjunct
CONTR	contraction
CSUB	causative subordination
DEMP	demonstrative pronoun
DISC	asynchronous computer-mediated discussion
DPART	discourse particle
EMAIL	academic email
FPR	final (stranded) preposition
GENEM	general emphatic
GER	gerund
HED	hedge
HTEXT	academic hypertext
INDP	indefinite pronoun
INF	infinitive
IT	pronoun *it*
MSL	mean syntactic length
MWL	mean word length
N	noun
NMOD	necessity modal
NOM	nominalization
NPR	number of prepositions

ABBREVIATIONS
(Continued)

PAV	perfect aspect verb
PHC	phrasal coordination
PLADV	place adverbial
PMOD	possibility modal
PPC	past participial clause
PRADJ	predicative adjective
PRMOD	prediction modal
PRO 1	first person pronoun
PRO 2	second person pronoun
PRO 3	third person pronoun
PRPCL	present participial clause
PRTV	present tense verb
PTV	past tense verb
PUBV	public verb
PVERB	private verb
RTHAT	*that* relative
SAUX	split auxiliary
SEM	online academic seminar
SFL	Systemic Functional Linguistics
SREL	sentence relative
SUV	suasive verb
SYNEG	synthetic negation
TADV	time adverbial
THATD	*that* deletion
TTR	type/token ratio
WEBL	academic weblog
WHC	WH clause
WHQ	WH question
WHRCO	WH relative on object position
WHRCS	WH relative on subject position

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ACKNOWLEDGMENTS

I wish to thank first and foremost my adviser professor Ingrīda Kramiņa for her invaluable help, academic guidance and constant support.

I must acknowledge that my theoretical perspective on CMAD and the research approach applied in the present study have developed owing to insights of many prominent scholars whose lectures and seminars I had attended during the time of my doctoral studies at the University of Latvia, Southern University in Odense and the University of Aalborg in Denmark, the University of New Anglia in Norwich, the UK, and Nordic Graduate School of Language Technology at the University of Helsinki in Finland. I am particularly grateful to James Hurford, Gunter Kress, Ron Scollon, Simon Kirby and Sigrid Norris, who provided the necessary theoretical insights at the initial stages of developing the framework for the present research.

I owe a debt to my colleagues at the Department of English Studies, the University of Latvia, for constantly supporting and encouraging me in the process of writing this paper. I am particularly indebted to Irina Surkova and Dace Liepiņa and other colleagues for the critical reading of the parts of this dissertation and providing valuable advice.

I would like to acknowledge the Head of the Chair of Management of the Faculty of Management and Economics at the Transport and Telecommunications Institute professor R. Kopitovs and the leading specialist in the Laboratory of Applied Systems J. Juškeviča for the help with statistical data processing during the approbation of my research results.

I thank my students, friends and family for their patience.

INTRODUCTION

In recent years, the role of the English language as *lingua franca* for international academic communication has significantly increased. Moreover, the area of its application has further dramatically expanded with the emergence of new information and communication technologies. Recently, new ways of communicating ideas in the academic world have appeared. For example, the ideas are now communicated at on-line seminars and lectures, in academic discussion forums and e-mails, academic weblogs and hypertexts. As a result, a new type of academic discourse has appeared – the use of academic language mediated by computer. In the research presented in the dissertation, I distinguish this new type of discourse from other types of English discourse and investigate its specific characteristics. To refer to this type of academic discourse in the present research, I introduce a new term *computer-mediated academic discourse*. The Latvian term *elektroniskais akadēmiskais diskurss* is used as the equivalent. The scope of its meaning in the present study is restricted to only the type of academic discourse that is mediated by computer as a communication medium.

It is important to investigate English computer-mediated academic discourse and to distinguish its specific characteristics for the following reasons. The new technological facilities for academic communication are now available not only for native speakers of English but also for non-native speakers, offering them new opportunities for study and promoting their research worldwide. Latvia, as the European Union member state, is actively participating in the integration process in the academic world. Hence, researchers and students in Latvia have become a part of the international academic community and participate in international academic discourse, which is now often mediated by computer. Thus, academic discourse competence that the students of academic English in Latvia need to develop should include the knowledge of specific linguistic characteristics of academic discourse mediated by computer and the ability to distinguish the differences among its types.

With the increase of the role of computer in human communication, the interest in the study of the use of language mediated by computer has dramatically grown in recent years. Many linguistic studies reveal considerable changes in the use of linguistic means on the Internet (e.g. Ferrara, Brunner and Whittemore, 1991; Collot and Belmore, 1993; Yates, 1996). Moreover, the researchers investigating electronic varieties of language have noticed that new electronic communication facilities have a high potential to change also the nature of academic discourse (Joyce, 1995; Dowling 1998; Snyder, 1998, 2002; Edminster and Moxley, 2002; Kress, 2003; Baron, 2005). The scholars analysing trends in current development of academic research genres have noticed some already observable changes and predict further

transformations caused by the emerging new technology even in the most conservative academic genre – the doctoral dissertation. According to Swales,

> It thus appears that the doctoral dissertation (like many other academic genres) is in a state of considerable flux, and this will certainly continue, partly as a result of technological change. The growing worldwide move toward ETDs [electronic theses and dissertations] introduces possibilities of colour images, sound and video files, and external links long denied to the traditional text dissertation.
>
> Swales (2004: 110)

Moreover, continuing 'englishisation' (Phillipson, 1992; Yates, 1996; Posteguillo, 2002) of the Internet, especially at the tertiary level of education, and the trend towards globalisation and 'digitalisation' in academic publishing set new demands for teaching the English language for academic *computer-mediated communication* (CMC) to help young scientists and scholars to study and promote their research worldwide.

Nevertheless, computer-mediated academic discourse has not yet been investigated theoretically and empirically. The reason for that is that academic language use has been regarded as rather conservative and not considerably changing under the influence of the computer medium (e.g. Crystal, 2001). Almost the same linguistic means are thought to be used for writing an academic e-mail, a message in an academic discussion forum, a weblog entry or a traditional academic text. As a result, computer-mediated academic discourse has not yet been sufficiently studied. Moreover, there have been no linguistic studies that would investigate and compare the characteristics of different types of English computer-mediated academic discourse in order to apply this knowledge to teaching and learning academic language at university. No such studies have been found conducted in Latvian context either.

However, the idea that the use of academic language is not influenced by recently developed electronic means of communication raises considerable doubts. If there were no significant differences in the use of academic language mediated by computer, communicators would not have difficulties in switching from one type of academic discourse to another. A number of recent studies, however, report on the major problems encountered by advanced language learners participating in computer-mediated academic discourse while they did not have such difficulties in traditional academic discourse. Chen (2006), for example, describes the struggle of a Chinese learner of English for finding appropriate linguistic means in academic e-mail communication with professors in the United States. Toyoda and Harrison (2002) provide further examples of the difficulties experienced by Japanese students in synchronous academic computer-mediated communication. One more illustration is provided by Belz (2003), who reports that e-mail communication between German students of English and the students in the United States was not sustained because

the participants failed to understand and observe the linguistic norms appropriate in different academic cultures.

To illustrate the existing problem, the following is one of the first messages received by the author of the present paper soon after joining an on-line academic community and after activating personal e-mail communication with other members on the website *academici.com* (the place is supposed to be of a high academic status, as to be allowed to join that academic community one must be at least a Ph.D. student and needs a special invitation from a current member holding a doctoral degree)[1]:

> hi natalie,
> thanx for allowing me to be in your network.
> may i have some more conversation with u.
> can we be friends>>>?? if yes then we have to share more thoughts.
> i would like to exchange more with u??
> well my yahoo id is
> xxxxxxxx@yahoo.com
> and yours??? may i know???
> ok bye
> looking forward for your reply
> XXXXXX

The example above may be regarded as an extreme case of inappropriate choice of register for academic communication, use of unconventional spelling and punctuation. However, it is a good illustration to the problem that may be caused by lack of experience and unawareness of academic writing conventions or inability to differentiate between the academic communication and an informal exchange of e-mails with friends and relatives. A similar problem of the proper language use in English computer-mediated academic discourse has been earlier reported (Cigankova, 2004) as encountered by university students in Latvia.

The provided examples imply that linguistically competent communicators choose different linguistic means available in the English language in different types of academic discourse mediated by computer. Therefore, the types of computer-mediated academic discourse, supposedly, should have specific linguistic characteristics distinguishing them from each other and other types of English discourse. The present study aims to distinguish the objective, statistically computable differences in the frequency of occurrence of linguistic features in six types of computer-mediated academic discourse. The English term *variation* (Biber, 1988) is used throughout the dissertation to denote fluctuations in the frequency of occurrence of linguistic features in texts (Latvian statistical term *variācija* (Raščevska and

[1] Presented here with gratitude to the author of this message for the permission to use it in the present paper.

Kristapsone, 2000) is used as the equivalent). To measure the differences in the frequency of occurrence of linguistic features is important for testing the theoretical model of linguistic variation in computer-mediated academic discourse elaborated in the present study. Thus, the dissertation raises topical, timely and important problem and provides objective statistical data for its solution.

The research subject

Linguistic variation in English computer-mediated academic discourse.

The research object

Oral and written computer-mediated academic discourse compiled in a corpus of authentic texts. The corpus was collected in the international on-line academic community of education professionals in 2003-2007.

The research hypothesis

There is a significant variation in linguistic characteristics across text types in English computer-mediated academic discourse, reflecting the differences in the functional use of linguistic features in its different types.

The research goals

1) To study English computer-mediated academic discourse;

2) To obtain objective statistical data on the variation in linguistic characteristics across its types for further applied linguistic research and language education.

The research objectives

1. To conduct the review of the relevant literature in order to trace the development of the linguistic concepts of *discourse, discourse analysis, academic discourse, computer-mediated discourse* and *linguistic variation* necessary for the theoretical conceptualisation of the notion of *computer-mediated academic discourse*.
2. On the basis of the theoretical exploration, to conceptualise the notion of *computer-mediated academic discourse* for the empirical study of linguistic variation in it.
3. To provide a comprehensive typology and description of the text types in English academic discourse mediated by computer with the purpose to systematise the knowledge that has been accumulated in the studied literature.
4. To compile a corpus of texts representing the types of English computer-mediated academic discourse for its empirical study.
5. To determine the distinguishing linguistic characteristics of the types of computer-mediated academic discourse in order to reveal the differences and measure variation.

6. To measure the linguistic variation in English computer-mediated academic discourse empirically in order to identify the most characteristic features of its different types.
7. To conduct the triangulation of data and instruments by applying the corresponding statistical procedures and tests in order to increase the internal validity of the study.
8. On the basis of revealed linguistic variation in the compiled corpus, to give an objective generic classification and description of the text types in English computer-mediated academic discourse for further applied linguistic research and language education.

Theoretical research methods

The author conducted a review and analysis of the relevant literature on the research problem. The **theoretical foundation** of the present research is built on the following:

1. The insights from the seminal works of Malinowski (1923), Firth (1957), Halliday (1973, 1978), Halliday and Hasan (1976, 1990), Halliday and Matthiessen (2004) defined the functional-linguistic orientation of the conducted research work. The "non-autonomy postulate" (Givón, 1995: xv) was applied as the main theoretical principle. It states that language cannot be adequately investigated as an autonomous system and should be studied as discourse, i.e. language in use in a situational context.
2. The discourse-analytic approach was informed by the works of Brown and Yule (1983), Schiffrin (1994), Hyland (2000), Herring (2001), Načisčione (2001) and Smith (2004).
3. The perspectives of genre and register theories on language variation in discourse, applied to the study of CMAD in the present dissertation, have been drawn from the works of Bakhtin (1986/1999), Chafe and Danielewicz (1987), Swales (1990, 2004), Bhatia (1993, 2004), Bazerman and Prior (2004), Biber (2003, 2006), Eggins and Martin (1997), Martin (2001), Giménez-Moreno (2006).
4. Biber's (1988) study of the variation across speech and writing provided further theoretical grounds for empirical investigation of linguistic variation in English computer-mediated academic discourse.
5. Additionally, the present thesis incorporates the views on the Internet language expressed in the works of Crystal (2001), Shortis (2000), Inman (2004) Jewitt (2006), Posteguillo (2002), Snyder (1998, 2002) and Wysocki (2004).

Empirical research methods

1. Quantitative multidimentional analysis of the frequency of linguistic features, originally elaborated by (Biber, 1988), was applied to reveal the linguistic variation across text types in English computer-mediated academic discourse.

2. To increase the internal validity of the study, the triangulation of the quantitative research data and instruments was carried out by applying the following statistical procedures:
   1) *Factor Analysis* was applied to identify the number of functional dimensions along which the variation takes place in English computer-mediated academic discourse.
   2) *General Linear Models* (ANOVA) was applied to calculate the statistical differences between CMAD types along the textual dimensions identified in the Factor Analysis.
   3) *Scheffé's Test* was conducted to analyze the pairs of mean values to reveal where exactly the differences lie.
3. The study was complimented by the elements of qualitative analysis applied to reveal the characteristics of computer-mediated academic discourse overlooked in the quantitative study.

The research novelty

1. The study has revealed a statistically significant variation in computer-mediated academic discourse that had been previously considered not significantly varying.
2. The study has distinguished a new type of English discourse – *computer-mediated academic discourse* (CMAD), proposed its original conceptual definition and provided the theoretical basis for the empirical exploration of linguistic variation in it.
3. The author has elaborated a new classification of texts representing English CMAD, based on the systematisation of the knowledge accumulated in the studied literature.
4. The author compiled and examined a new corpus of texts representing English CMAD. The authentic texts comprising the corpus had not been previously studied.
5. A well-established research methodology (Biber, 1988) was applied in the present research to a new linguistic material – English CMAD.
6. In the course of the statistical analysis, the author has obtained new quantitative data on linguistic characteristics of each CMAD type which enabled its objective typological classification.
7. The author has distinguished new lexical, grammatical and generic characteristics of CMAD types.
8. The author has made a new contribution to the study of the varieties of the English language in use by determining the position of each type of CMAD among other types of English discourse on each functional dimension. This was done by comparing the results of the present study with the results provided by Biber and other researchers.

Theoretical significance

1. The author has traced the historical development and the use of the concepts of *discourse, discourse analysis, academic discourse, computer-mediated discourse* and *linguistic variation* in applied linguistics, which is significant for the conceptualisation of the notion of *computer-mediated academic discourse* as a contribution to the linguistic theory.
2. The author proposes the conceptualisation of the notion of *computer-mediated academic discourse* and the analysis of the causes and mechanisms of variation in it, which is theoretically significant as a basis for further empirical research.
3. The proposed theoretical model of linguistic variation in CMAD, tested and corrected in the course of empirical research, may be used as a methodological basis for further linguistic theoretical and empirical research.
4. The listed above theoretical inferences set a theoretical basis for linguistically informed methodological approaches to language education.

Practical significance

1. The conceptual framework built in the study may be applied as a basis for elaboration of new approaches to the development of students' communicative language competence in English as the language for international academic communication mediated by computer.
2. The results of the study may be applied to designing the teaching materials for distance education.
3. The author has distinguished new lexical, grammatical and generic characteristics of CMAD types that are practically significant for the development of communicative language competence in language education.
4. The results of this study may be further applied to designing new electronic university courses for MA in English Philology (applied linguistics) students at the University of Latvia**,** e.g. *English Computer-Mediated Communication* and *Computer-Assisted Language Learning*.

The approbation of the research

1. The theoretical framework developed in the dissertation and the results of the empirical research have been discussed and approved at the meetings of the Department of English Studies at the Faculty of Modern Languages of the University of Latvia.
2. The statistical procedures applied in the present research have undergone approbation in the Laboratory of Applied Systems at the Transport and Telecommunications Institute.
3. The results of the research and the theoretical framework elaborated in the present work have been applied to designing language teaching materials within the framework of the

E-university Project at the University of Latvia. During the time of developing the present doctoral research the following electronic teaching materials (WebCT/Moodle courses) have been created:

1) Cigankova, N. (2008) *English Computer-Mediated Communication* (English Philology MA programme, part B).
2) Kalnberziņa, V., Ozoliņa, S. and N. Cigankova, (2005). *Role of Language in Intercultural Communication* (English Philology MA programme, part B), University of Latvia Exemplary Course Competition 2006 First Prize winner.
3) Cigankova, N. (2003). *Applied Communication II (*English Philology BA programme, part A), University of Latvia Exemplary Course Competition 2003 Second Prize winner.
4) Kramiņa, I., Cigankova, N., and Z. Vinčela, (2003). *Modern Information Technology in English Language Teaching* (English Philology MA programme, part B).
5) Cigankova, N. (2002). *Applied Communication I* (English Philology BA programme, part A).

4. The results of the present study have been reported in the following seven internationally refereed publications:

1) Cigankova, N. (2008) Academic Culture on the World Wide Web: Implications for Teaching Academic Writing. In Nikševič-Batričevič, A., Kneževič, M. (eds.) *Culture-Bound Translation and Language in the Global Era.* Cambridge: Cambridge Scholars Publishing, 117-136.
2) Cigankova, N. (2007) Between Pragmatics and Creativity: Writing Academic Hypertext in L2. *Innovations in Language Teaching and Learning in the Multicultural Context*. Research papers: International Nordic-Baltic Conference of the World Federation of Language Teacher Associations (FIPLV), Riga, Latvia 15th -16th June, 2007, CD-ROM (9 pages).
3) Cigankova, N., Nukševica I. (2007) Creativity in English Academic Hypertext. *Radoša Personība*, Riga: Jumi, 247-254.
4) 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.
5) Balčiūnaitienė, A., Cigankova, N., Voronova, L. (2005). EAP Learners' Perspective on On-line Presentation of Academic Texts. *Teaching Writing Online and Face to Face.* Athens: Hellenic American Union, CD ROM (9 pages).
6) Cigankova, H. A. (2005) Polycontextuality of Electronic Academic Discourse. *Texts and Contexts: the Movement of Language Selected Papers.* Kaunas 2005: VU Press, 583-89.
7) Cigankova, N. (2002) Students Guiding Students through the Web. Pulverness (Ed.) *IATEFL 2002 York Conference Selections,* UK, York: IATEFL, 58-59.

5. The theoretical considerations on the research theme and empirical findings obtained in the present study have been presented at eighteen international conferences in Riga and Daugavpils in Latvia and in Athens (Greece), Kaunas (Lithuania), Nikšič (Montenegro),

York (United Kingdom). The following international conference reports on the research topic have been presented:

1) Cigankova, N. (2008) Linguistic Variation in Computer-Mediated Academic Discourse. International Conference *Texts and Contexts: Interactive Perspectives,* 16th—17th October, 2008, Kaunas Faculty of Humanities, Vilnius University, Lithuania.
2) Cigankova, N. (2007) Variation in Linguistic Dimensions of English Computer-Mediated Academic Discourse in Latvian Context. *The 2nd Congress of Letonics*, 31st October, 2007, Latvian Academy of Science, Riga, Latvia.
3) Cigankova, N. (2006) A Comparison of Academics' and Students' Computer-Mediated Discussions. International Conference *Language, Culture and Technologies,* May 19th – 21st, 2006, Kaunas University of Technology, Kaunas, Lithuania.
4) Ozoliņa, S., Kalnberziņa, V. and N. Cigankova (2006). Designing and Managing New WebCT Courses *Language Awareness* and *Role of Language in Intercultural Communication* for MA students', International Conference *Educational Management and School Development*, 3rd February, 2006, University of Latvia, Riga, Latvia.
5) Zigrīda Vinčela, Vita Kalnberziņa and Nataļja Cigankova (2006). An Overview of the Software for Learner Language Analysis in Applied Linguistics, 64th Annual University of Latvia Scientific Conference, Riga, Latvia, 1st February, 2006, University of Latvia, Riga, Latvia.
6) Vinčela, Z. Cigankova, N. (2005) Developing Young Learners' Foreign Language Communication Skills with Computers*, IATEFL Young Learners Conference*, 28th August, 2005, University of Latvia, Riga, Latvia.
7) Cigankova, N. (2005) What Can Non-humans Teach Humans? - Developing Language Fluency by Dialogue Games with 'Chatbots'. International scientific conference *Humanities in New Europe*, 13th–14th January 2005, Kaunas, Lithuania.
8) Cigankova, N. (2005) Learner Language Adaptation to Electronic Environment. *LU 63rd Scientific Conference,* 4th February, 2005, Riga, Latvia.
9) Cigankova, N. (2004) Metalinguistic Aspect of On-line Academic Discourse. *LU 62nd Scientific Conference,* 3rd February, 2004, Riga, Latvia.
10) Cigankova, N. (2003) WebCT Presentation and Assessment Tools: Teaching Experience. *LU 61st Scientific Conference,* 5th February, 2003, Riga, Latvia
11) Cigankova, N. (2002) Using the Internet for Teaching English. *LU 60th Scientific Conference,* 1st February, 2002, Riga, Latvia.

6. Series of workshops for university and secondary school teachers have been conducted at the University of Latvia and in the tertiary educational establishments in Latvia, and Lithuania. The following workshops practically applying the results of the present research have been conducted:

   1) *Developing Young Learners' Foreign Language Communication Skills with Computers*, 28th August, 2005, University of Latvia, Faculty of Modern Languages.

2) *Developing Learner Native Tongue- and Culture-friendly Electronic Language Teaching Materials*, 22nd April, 2005, **Ventspils Graduate School, Latvia.**
3) *Electronic Discourse and Artificial Intelligence: Implications for Language Teaching*, 4th March, 2005, Vytautas the Great University, Kaunas, Lithuania.
4) *Developing Electronic Materials for Language Teaching and Testing*, 3rd March, 2005, Vytautas the Great University, Kaunas, Lithuania.

Short Outline of the Dissertation

The doctoral dissertation (160 pages) comprises an introduction, two parts (six chapters), conclusions, bibliography (252 items), table of contents, tables, figures, a list of conference reports and publications of the author, a glossary of the terms used in the study and appendices.

Part 1 of the dissertation consists of three chapters. It is devoted to a review of the literature on the main concepts investigated in the present study.

Part 2, comprising two chapters, describes the research methodology and presents the obtained results and discussion on the findings.

The last section of the dissertation presents the main conclusions and further research perspectives.

Glossary contains the definitions of the terms used in the dissertation.

Appendices present samples of CMAD text types, samples of computer program output, and the tables with the results of computerised analysis of the research samples.

1. THE LITERATURE REVIEW ON LINGUISTIC VARIATION IN COMPUTER-MEDIATED ACADEMIC DISCOURSE

The present study is devoted to the investigation of English computer-mediated academic discourse (CMAD), i.e. the English language use by language teaching professionals in computer-mediated seminars, synchronous conferences, asynchronous discussions, e-mails, weblogs and hypertexts that are used in academia, that is at the tertiary-level of education, for teaching and learning the disciplinary subject matter in the field of education. The analysis of the literature in the present research is restricted to only the sources that are relevant to linguistic analysis of variation in CMAD, excluding the discussion on other types of English discourse, the use of other types of language on the Internet and the disciplinary research approaches other than linguistic.

In the first part of the present study, the author approaches the task of defining the main research concept – *computer-mediated academic discourse*. Chapter 1.1 offers the conceptual definition of CMAD developed for the present research and a classification and a comprehensive linguistic description of the main characteristic features of six CMAD types. Chapter 1.2 presents the author's view on the notion of linguistic variation in discourse, its types and the factors influencing it. An overview of recent empirical research and the rationales for the present study are provided in Chapter 1.3.

1.1 CONCEPTUAL DEFINITION AND CLASSIFICATION OF COMPUTER-MEDIATED ACADEMIC DISCOURSE

The definition of CMAD as a basic theoretical concept for the present study requires the use of the terms *discourse, academic discourse* and *computer-mediated discourse* as building blocks for its construction. This presupposes the existence of a conceived, unambiguous and widely accepted in contemporary linguistics definition of *discourse*. However, the discussion in the relevant literature on the concept of discourse and approaches to discourse analysis reflects that, as in many other new fields of academic studies, the terminology of discourse studies in linguistics is still in the process of development. Further, the author of the dissertation will trace how the meaning of the term varies in different theoretical approaches in linguistics and explains the choice of the conceptual definition of *discourse* for the present research. The concepts of *academic discourse* and *computer-mediated discourse* are scrutinised as constituent parts ('ingredients') of the concept of CMAD. The author provides a conceptual definition of CMAD and discriminates between CMAD and other contesting

terms, e.g. *electronic discourse, interactive networking, interactive written discourse, electronic language, electronic networked discourse,* etc.

1.1.1 Concepts of Discourse and Discourse Analysis

Although the study of discourse is a relatively new area of linguistic research, it has attracted much attention and continues to raise considerable interest of scholars from various schools of thought and sub-disciplines of linguistics (Harris, 1952; Brown and Yule, 1983; Schiffrin, 1994; Gutiérrez, 1995; Gee, 1996; Carter, 1997; van Dijk, 1997; Hyland, 2000; Herring, 2001; Chafe, 2001; Flowerdew, 2001; Kress and van Leeuwen, 2001; Schiffrin, Tannen, and Hamilton, 2001; Scollon and Scollon, 2001; Celce-Murcia and Olshtain, 2005). The definition of discourse as a term, however, poses serious difficulties. It is complicated by the use of the term across many different disciplines other than linguistics. According to Lemke (1995), the term *discourse* is used to mean a range of different phenomena from very specific, e.g. spoken language, to overly general, e.g. communication as a social process. Further, only the definitions of discourse in the field of linguistics that are informative for the present study are discussed.

The first scholar who introduced the term *discourse* in linguistics was Harris (1952a). He applied a structural approach to study the syntax of units of communication larger than words or sentences and attempted to do a formal analysis of what he understood as "connected discourse" (1952b: 474), i.e. language in sequence beyond the sentence. Since that time, *discourse analysis* has been understood in formal linguistics as the study of stretches of language longer than a clause or sentence. Nevertheless, because of the idealisation of the linguistic data, syntactic structures are typically seen without their connection with the real world. The analysis of language in this linguistic tradition does not take into consideration the context of language use in the texts. Consequently, this approach cannot be considered satisfactory informative for the present research, as it overlooks many factors that influence language use in CMC. The present study aims at investigation of computer-mediated language used in academic social contexts; therefore, the concept of context is regarded in it as crucially important.

The tradition of discourse analysis that locates language in the real world and regards it in the context of communicative situation has been started in the works of Halliday (1973) and Halliday and Hasan (1976: 2), who define it as a "unit of language in use", the linguistic form of which is determined by the functions that the linguistic features in the text perform in communication context. Martin and Ringham (2000: 51) summarise systemic-functional

approaches to discourse in the following definition: discourse is "a unit of language larger than a sentence and which is firmly rooted in a specific context," e.g. academic, political, legal or media discourse. Thus, in systemic-functional tradition, discourse is a unit of language beyond sentence level used in a situational context. Although this interpretation of the concept of discourse is very common in functional-linguistic research, it is rather ambiguous, as it depends on the understanding of the terms a *unit* of language, language *use* and *context* which are also rather ambiguous terms (van Dijk, 1997).

The first word in the systemic-functional definition of discourse that needs clarification is a *unit* of language, in terms of whether a unit of language formed by utterances in speech or by sentences in writing is meant as a unit of analysis. Although the term *text* in systemic-functional linguistics is applied to refer to "any passage, spoken or written, of whatever length, that does form a unified whole" (Halliday and Hasan, 1976: 1), spoken and written texts are treated differently in systemic-functional research. Only spoken texts are typically regarded as discourse. The term *text*, not *discourse*, is used to refer to a unit of written language above sentence level in its context of use. This happens because written discourse is generally regarded as a product, while spoken discourse – as a process. The differentiation is the result of the distinction between the ways in which spoken and written discourse are realised. While language in spoken communication unfolds to a listener in sequence – utterance per utterance, written language 'appears' in a form of a whole text, though it is read by a reader sentence by sentence. Hence, depending on whether spoken or written variety is used in situational context, discourse is regarded as either a process of language production or a product emerging as a result of communication process.

Although Halliday (1985: 290) warns against "artificial polarization of speech versus writing", different discriminating speech and writing contemporary approaches to discourse and discourse analysis have developed out of the systemic-functional approach. In the studies of spoken language, e.g. *conversation analysis* (Sacks, Schegloff, and Jefferson, 1974), only spoken language in its dialogic forms is defined and studied as discourse. An illustration is the definition by Thompson (1984: 74), who writes that *discourse analysis* is "... a rapidly expanding body of material which is concerned with the study of socially situated speech ... united by an interest in extended sequences of speech and a sensitivity to social context." Written language, in contrast, is studied by the discipline of *text linguistics* where the term *discourse* is used to refer "... to the level of meaning above the lexicogrammar, the level concerned with relations of meaning across a text" (Eggins, 2004: 24). This division can hardly inform a comparative analysis of different spoken and written types of CMAD undertaken with the purpose to apply the results of the research to English language teaching.

If the author regarded the division described above, the present research would have to apply conversation analytical procedures to dialogic/ polylogic types of CMAD (e.g. synchronous conferences, discussion forums) and text linguistics methods to its predominantly monologic types (e.g. e-mails, weblogs, academic hypertexts).

The purpose of the present study, however, is to identify the most salient characteristics of each type of CMAD distinguishing it from other types of English discourse for the purposes of language learning and teaching. A profound, exhaustive analysis and description of every particular type of CMAD, however useful for the description of language as a system in general it could be, is beyond the scope of the present study. On the contrary, the focus of it is on comparing and on revealing salient differences between the CMAD types, which could be applied in language learning and teaching. Therefore, the uniformity in analytical procedures applied to the investigation of each type of CMAD is crucial for revealing the linguistic variation among its different types.

The second word that makes the definition of discourse in systemic-functional linguistics rather ambiguous is *use* of language. According to Halliday,

> A functional approach to language means, first of all, investigating how language is used: trying to find out what are the purposes that language serves for us, and how we are able to achieve these purposes through speaking, listening, reading and writing. But it also means more than this. It means seeking to explain the nature of language in functional terms: seeing whether language itself has been shaped by use, and if so, in what ways – how the form of language has been determined by the function it has evolved to serve…"
>
> Halliday (1973: 7)

Thus, it is important to regard "*who* uses language, *how, why* and *when*" (van Dijk, 1997:2, emphasis in the original). The social status and relations of the language users, their purposes of communication and intended audience all influence the linguistic preferences that they make. To achieve various goals in the process of communication, language users address a particular social group, e.g. academic users of language, adjusting their individual use of language to cultural norms, values and conventions of language use appropriate to the group. Moreover, the study of discourse, either as spoken or written text, also involves consideration of common social knowledge about language use in a particular social group – general knowledge of the rules of grammar and principles of discourse organization, coherence, management and interpretation. This knowledge is implicitly or explicitly shared by all linguistically competent members of society. This implies cognitive aspects, such as personal knowledge, beliefs, attitudes, constraints on memory of the language users. This also implies that social situation in which language serves as a means of communication is regarded as a factor causing context-dependent variation in the choice of linguistic features by language

users in a particular communicative situation. Thus, the social and cultural context shapes the use of language causing variation in functional use of linguistic features in texts. Consequently, functional discourse analysis aims to investigate the correlation between linguistic form and function, revealing how lexicogrammatical forms are determined by their use in particular contexts.

The question, however, is how autonomous the language users are from the society when they establish meanings to linguistic forms in discourse. Selecting which aspects of social and cultural life of humans should be regarded as decisive factors influencing language use in a particular communicative situation depends on the extent to which individuals are regarded dependent or independent of the society as language users. The notion of language use, therefore, heavily depends on the breadth of the scope of the notion of relevant social context.

Context, thus, is the third ambiguous word that should be clarified in order to understand properly the definition of discourse. Functional approaches to discourse analysis discern between different understandings of the notion of *context of language use* (Malinowski, 1923; Firth, 1957). On the one hand, it can be understood as *intratextual context*, *co-text* (Halliday, 1973), i.e. as co-occurring text surrounding a particular linguistic feature in it. On the other hand, it can also be understood as *extratextual context*, i.e. as broader social and cultural processes influencing language use. Thus, discourse may be regarded as a complex of phenomena, including social and psychological, in addition to linguistic. However, a broader social context of language use, although regarded as a factor influencing linguistic form, is not regarded in systemic-functional linguistics as a constituent part of discourse. Individual language users are viewed as relatively autonomous from the society in establishing meanings to linguistic forms. Therefore, emphasizing that language use is affected by context, systemic-functional discourse analysis is only concerned with the influence of immediate context on language use. Thus the notion of *discourse* does not embrace the use of language as an instrument applied by people to *do* things, i.e. to achieve their goals.

To summarise, the clarification of the terms that comprise the systemic-functional definition of discourse reveals the following limitations of it:

- Firstly, the notion of *discourse* is regarded as static while the use of language is a process in which various "units of language use" are realised as products.
- Secondly, in the case of written language, discourse actually equals to *text*. However, the use of language cannot be delimited to text (at least not to only verbal text), as it is

an instrument used for communication, and any communication is multimodal (Kress and van Leeuven, 2001).

- Thirdly, the ambiguity is caused by the fact that the notion of relevant context is restricted to only immediate situational context. The users of language are viewed as relatively autonomous from the society in making meaning, and a broader social context is not seen as a part of discourse.

Thus, the limitations of both presented above structural approaches have led to further attempts to define discourse made by scholars looking at it from different theoretical perspectives. While the concept of context is rather restricted in formal and systemic-functional approaches to discourse, other disciplines attach a much broader meaning to it. The importance of the role of language in social and cultural life in human society is recognized and studied by many different disciplines other than linguistics: philosophy, psychology, human anthropology, rhetoric, sociology, and cultural studies. The terms *discourse* and *discourse analysis* may mean different things and activities depending on with what aspects of language use the discipline is concerned. For example, discourse analysts interested in content, i.e. emergence and circulation of ideas in society, see discourse as "interrelated set of texts, and the practices of their production, dissemination, and reception, that brings an object into being" (Phillips and Hardy, 2002: 3). They are concerned with ideological implications of language use on society and culture, as for example in the disciplines of *sociology of language* and *critical discourse analysis,* which aim to investigate why people say or write certain things and what social changes and how this may cause (Gee, 1996; Fairclough, 1993, 2000; van Dijk, 1997).

Linguistics, however, is primarily concerned with the study of language as a phenomenon/system while other disciplines interested in the study of discourse regard things other than language as primary objects of their research, e.g. society, human personality and behaviour, communication technology etc. Therefore, it is reasonable that linguistic discourse analytic research, in order not to step in other's disciplinary shoes, should limit the scope of the notion of context of language use to what is relevant to the main goal of linguistics – the study of the language system and its discourse realisations. Seeking to increase the knowledge about the society and its members should be left to other disciplines. For example, the borders of the notion of relevant context in the present research are restricted by its linguistic target: the increment of empirical knowledge about medium-specific linguistic variation in computer-mediated academic discourse. Thus, the ideological, socio-cultural, religious etc. aspects of language use in CMC, as well as a broader critical view with the purpose to 'improve' the society, are beyond the scope of the present research. Therefore, a further

search for a definition of the concept of discourse for the present study goes back to the field of linguistics.

The definition of discourse that gives priority in discourse analysis to the linguistic goals has been introduced by Brown and Yule (1983). They apply Halliday's functional perspective on discourse analysis to perform the work traditional for descriptive linguists: "to give an account of how forms of language are used in communication" (ibid: ix[2]). However, in contrast to Halliday, Brown and Yule define discourse as "a dynamic process in which language was used as an instrument of communication in a context by a speaker/ writer to express meanings and achieve intentions" (ibid: 26). In this definition, discourse is not a "unit of language" but "a dynamic process" in which language is only "an instrument", i.e. a means of communication. Thus, discourse is understood as a process of language use; however, it is realised in collectable and analysable linguistic data. These *linguistic realisations* of discourse as a process are (spoken or written) texts exchanged by the members of society. How language is used in the texts is determined by the purposes of the use of linguistic features in them. According to Brown and Yule,

> The analysis of **discourse** is, necessarily, the analysis of language in use. As such, it cannot be restricted to the description of linguistic forms independent of the purposes or functions which those forms are designed to serve in human affairs.
>
> Brown and Yule (1983: 1, emphasis in the original)

Brown and Yule insist that discourse analysis is primarily a linguistic discipline with linguistic research methodology. Linguistic data in discourse analytic methodology are treated as a record (text) of the process of language use (discourse) and are used to find and describe the patterns in the use of linguistic features that people choose from available language resources to achieve their goals in communication. The scholars propose a pragmatic approach that puts speakers/ writers at the centre of communication process. As Brown and Yule insist,

> …it is people who communicate and people who interpret. It is speakers/ writers who have topics, presuppositions, who assign information structure and who make reference. It is hearers/ readers who interpret and who draw inferences (ibid: ix).

Linguistic features in discourse are analysed in their interconnection to investigate "what people using language are doing" (ibid: 26). They are regarded as means employed by language users to achieve their goals. This methodological approach opposes traditional hermeneutic methodology with its "…individual (or idiosyncratic) approach to the interpretation of each discourse fragment" (ibid: x). Instead, Brown and Yule adopt a "compromise position" and suggest combining the study of linguistic forms and "the

[2] References to pages throughout the present dissertation refer to the sixteenth reprinted edition (2006) of the book first published in 1983.

regularities of their distribution" (ibid.) and the principles of their interpretation. For this reason, Brown and Yule utilise Halliday's assumption that discourse function determines a linguistic form in a particular context and that "configurations of functions" (Halliday, 1985: x) determine the patterns of use of linguistic features in texts as linguistic realisations of discourse (also *discourse representation* in Brown and Yule, 1983: 5). According to Brown and Yule, the aim of the study of discourse is to use linguistic data to "describe regularities in the linguistic realisations used by people to communicate those meanings and intentions" (ibid.). The regularities in the use of linguistic features in texts (as linguistic representation of discourse) are determined by similarities in discourse functions that linguistic features perform in them. Hence, the study of discourse for Brown and Yule is concerned with the study of language functions.

The difference between the formal structural (Harris) and the systemic-functional (Halliday) and pragmatic (Brown and Yule) approaches to conceiving the concept of discourse is summarised by Schiffrin,

> Discourse has often been viewed in two different ways: a structure, i.e. a unit of language that is larger than the sentence; and the realization of functions, i.e. as the use of language for social, expressive, and referential purposes.
>
> Schiffrin (1994: 339)

Thus, from the structural perspective, discourse is viewed as a sequence of utterances or sentences and the analysis of discourse is concerned with functional analysis of the use of linguistic features in them (structure + function). From the pragmatic perspective, discourse is regarded as the process of language use in communication and is analysed as the result of it. At one end of the continuum between these two theoretical positions is an extreme use of the term di*scourse* to mean a single *text*; at the other end – the use of the term discourse to mean *context*. The former approach may reveal itself in such an extreme form as, for example, the definition of discourse by Stubbs, in which *discourse* is simply "language above the sentence or above the clause" (Stubbs, 1983: 1), that completely disregards the context of language use. The latter approach is illustrated by a definition given by Ochs (1990: 289), who writes that *discourse* is 'a set of norms, preferences and expectations relating language to context, which speaker-hearers draw on and modify in producing and making sense out of language in context." This definition moves a primary attention of a linguist from the study of language to the study of "norms, preferences and expectations," i.e. to the study of social organisation, processes, human psychology and ideologies. The tradition of seeing language as "a site through which ideologies are produced" (Carter, 1997: 80) and, therefore, as an ideological battlefield goes back to Foucault (1972). Further development of this view of discourse leads

to an overwhelmingly broad approach to the study of language, as, for example, one that is proposed by Carter:

> Our models of language ... have to take account therefore of rules which are variable according to such factors as status, power and ideology and which recognise the fact that language involves systems which are both static (decontextualised) and dynamic (contextualised), rule-bound and rule-breaking, structured and fractured.
>
> Carter (1997: 81)

However, further stretching the definition of discourse as a process of language use by extending the view of communicative context to countless social processes and ideologies takes discourse studies far away from the discipline of linguistics. This leads to the emergence of such extreme, from the linguistic point of view, definitions of discourse as, for example, a characterisation of discourse as a communicative event by van Dijk (1997). In his view, discourse as a communicative event is "a part of more complex social events, for instance, in such specific situations as an encounter with friends, a phone call, a lesson in the classroom, a job interview, during a visit to the doctor, or when writing or reading a news report" (ibid: 2). The concept discourse in this interpretation goes beyond the view of language as an independent semiotic system and is seen as a part of a more general social process. Discourse is viewed as verbal interaction during such events that has "three main dimensions: (a) *language use,* (b) the *communication of beliefs* (cognition), and (c) *interaction* in social situations" (ibid.). The attempts to study discourse along all these three different 'dimensions' have resulted in losing the original linguistic focus in discourse studies, as language is regarded as only one out of three 'dimensions' of verbal interaction. Another view of discourse as social construction of reality, expressed by Coupland and Jaworski (2001), takes defining discourse even farther from linguistics' concerns. The scholars summarise the use of the term discourse in the humanities and social sciences and propose an integrated approach that they define in constructivist terms:

> ...the unifying insight that discourse analysis offers is that important aspects of our social lives are constructed in and through language, whether in the moment-to-moment social interchanges of everyday talk or in the beliefs, understandings and principles that structure our lives. Discourse analysis is therefore the attempt to observe, unravel and critique these acts of construction.
>
> (Coupland and Jaworski, 2001: 134)

Such approaches to discourse evidently aim at targets other than purely linguistic as the main concern of discourse analysis while intervention into the disciplines other than linguistics causes further differences in the use of the term discourse in the scholarly world. Schiffrin, Tannen and Hamiltton (2001: 1) summarise the abundant definitions of discourse in recent discourse studies and put them into three main categories: "(1) anything beyond the sentence,

(2) language use, and (3) a broader range of social practice that includes non-linguistic and nonspecific instances of language."

The increase in the number of different uses of the term discourse has resulted in the multiplication of approaches to discourse analysis. Six different approaches (speech act theory, pragmatics, ethnomethodology, interactional sociolinguistics, ethnography of communication, and variation theory) compared and contrasted by Schiffrin (1994) have expanded to about forty, as described in Schiffrin, Tannen and Hamilton (2001). Most of the approaches to define discourse can be situated on the continuum proposed earlier in this section. To define discourse, scholars choose the end of the continuum that suits their research interest and the purpose of their study. Coupland and Jaworski see the difference between definitions of discourse situated at the opposite ends of the continuum as the difference between 'local' and 'global' dimensions of discourse, concluding that, "The most incisive approaches to discourse are those that combine the detailed analysis of language, in particular instances of its use, with the analysis of social structure and cultural practice" (Coupland and Jaworski, 2001: 134).

With the emergence of written interactive forms of synchronous and asynchronous computer-mediated communication, it is important to understand how language functions are used in the process of CMC. Therefore, restricting the concept of discourse only to structures "beyond the sentence" is unreasonable and unproductive. However, stretching it to unmanageable by the disciplinary means of linguistics all-embracing study of social and cultural life of humans is unpractical, as such conceptual definition can hardly be used for developing an operational definition of discourse for the empirical research in the present study. A further search for a definition of discourse that would, on the one hand, enable easy operationalisation of CMAD for the present empirical research and, on the other hand, would not be a too abstract oversimplification of such a complex phenomenon as discourse leads back to the previously discussed definition of Brown and Yule and to Latvia where the term *discourse analysis* has been defined by Načisčione:

> Discourse analysis is a discipline, which attempts to identify and describe linguistic regularities and irregularities in utterances which cannot be accounted for at sentence level (see Carter 1995: 39), study language *in use* across sentence boundaries and explore the organisation of texts (see op.cit.: 40).
>
> Načisčione (2001:3)

Načisčione applies this definition of discourse analysis to the study of English phraseological units in discourse, which allows her "to discern features occurring across a wide stretch of text" (ibid.) that play the main role in manifestation of textual and interpersonal meanings of linguistic features in discourse. Although Načisčione does not directly give a definition of

discourse, the use of specific terminology in the definition above implies that she attempts to combine a structural and a pragmatic view on the concept of discourse. On the one hand, "language in use" is regarded as structure and studied "across sentence boundaries." On the other hand, "regularities and irregularities" in utterances and "organisation of texts" as linguistic realisations imply that discourse is regarded as realisation of functions in the process of language use in communication. This is in line with the definition of discourse provided by Brown and Yule discussed earlier in this section. Following this discourse analytic tradition, the author of the present research applies the integrated view of discourse to the study of CMAD and synthesises Brown and Yule's and Načisčione's definitions of discourse.

Thus the definition of discourse used in the present study is as follows:
Discourse is a dynamic process of functional use of language as a means of communication in a situational context. It is realised in semantically connected and meaningful to the communicating users of language verbal instances of spoken or written language longer than a sentence.

There are a number of important reasons for adopting this definition in the present study.

1. This definition emphasises that discourse, as a *dynamic process,* is expressed by *action* rather than by *state* of being. It is a verbal flow characterised by a continuous activity of language users and a change in the use of verbal means that they employ in communication. It means that the flow of discourse occurs either as a continuous action (*transaction*) or a series of actions (*interaction*) undertaken by language users to achieve their goals in communication. The former is defined by Brown and Yule (1983) as *transactional* in contrast to *interactional* characteristics of discourse. This dynamic process "unfolds" (Martin, 2001) in a linear manner – utterance per utterance, sentence per sentence – in real-time communication, e.g. spoken face-to-face conversation or synchronous computer-mediated communication. It can also be "unfolded' in a non-linear or recursive manner by readers of written texts and, especially, hypertexts with links to other parts of the text and other texts, allowing them to break the linear sequence of reading. Moreover, the process of producing written language is rarely linear, as writers constantly re-read and revise their writing.
2. The process of language use is defined as *functional*. This word performs two important roles in the definition. First of all, it implies that language in discourse has three metafunctions (Halliday, 1973): ideational, interpersonal and textual. Furthermore, it signifies that discourse is itself a function of (i.e. depends on) a number of factors

influencing any language use, e.g. purpose and intended audience, social background of language users, physical settings, etc. This dependence of the role that language performs and of the factors influencing it reveals itself at all linguistic levels – phonetic/ phonological, lexical, morphological, syntactic – in a variation in the use of linguistic features in discourse (i.e. in the process of the use of language above sentence level in context) and in organisation of verbal instances longer than a sentence (i.e. texts) as discourse realisations.

3. The definition restricts the scope of the notion of context. *Situational context* is often viewed as anything that is not verbal but is relevant, i.e. having influence on the process of language use, e.g. a social group of language users, software that they use for exchanging messages, their cognitive model of the type of communication in which they are engaged, etc. Context, however, is not everything but only what is at the foreground of communication, i.e. what is the factor influencing language use in each particular case (Bušs, 2003).
4. The proposed definition states that language is used as a *means of communication.* It designates that, in all three Hallidayan metafunctions that language performs in communication, it is used as an instrument for making meaning (i.e. *semiosis*). In this capacity, it represents the whole semiotic system of verbal signs that language users have at their disposal. To achieve a particular communicative goal, language users make linguistic choices that depend on the situational context.
5. *Meaningful* to communicating language users verbal instances are those which they foreground at each particular moment out of the discourse flow. For example, verbal signs sent by working radio or television at the background of a face-to-face dialogue are not a part of discourse unless some of them are noticed and moved to the foreground of the conversation, e.g. news items. In synchronous multi-user communication, such as that taking place in electronic 'chatrooms', one particular language user cannot maintain communication with all one hundred or more people engaged in communication at the same time. This language user would communicate with only a few people and only their utterances would make sense for him or her. Verbal instances produced by the rest of communicatiors would be moved to the background and would not be meaningful to this particular language user. An important point here is the unity in meaning that holds together the linguistic elements in which discourse is realised. Although discourse is not Hallidayan "unit of language use", it is a process realised in such units.
6. Discourse as a process is realised in *verbal instances longer than a sentence*. The obligatory condition under which any string of words longer than a sentence becomes

discourse is a semantic unity, i.e. that they are fixed together as a unit by logical coherence. Realisation means "the process of linking one level of organization with another" (Halliday and Matthiessen, 2004: 26), i.e. organising sounds or letters into words, words into utterances or sentences and further to form a discourse unit. In functional discourse analysis, these discourse realisations are texts, as *text* is defined in functional terms as "any instance of language, in any medium, that makes sense to someone who knows the language" (Halliday and Matthiessen, 2004:3). "Any medium" means that verbal instances used in discourse as a process may be spoken and transmitted by air or wire or written on paper, a computer screen or any other material carrier.

The proposed definition is adopted as the conceptual definition of the construct *discourse* in the present research. It serves as a basis for defining other key terms for the present study, such as the concepts of *academic discourse*, *computer-mediated discourse* and *computer-mediated academic discourse* which are derived from this a more general term.

Computer-mediated academic discourse is viewed in the present research as one type the process of language use that is delimited by situational context in two ways: by the social settings of communication, i.e. an academic community of language users, and by the use of computer medium for communication. Correspondingly, the two following sections of the present dissertation deal with each particular type of discourse: one that takes place in academic settings – *academic discourse* – and another which is mediated by computer – *computer-mediated discourse*. All six types of CMAD that are under investigation in the present research are classified as various types of academic discourse, since the context in which the process of language use takes place is restricted to only academic settings, i.e. when written or spoken language is used for presenting, discussing or learning the disciplinary subject matter in education at tertiary-level. The conceptual framework for academic discourse is discussed in the following section.

1.1.2 Definition and Characteristics of Academic Discourse

A survey of the literature indicates increasing interest in the research into contemporary academic discourse (Halliday and Martin, 1993; Berkenkotter and Huckin, 1995; Dowling, 1998; Hyland, 2000; Swales, 1990, 2004). Hyland writes about a clear consensus among scholars from different academic disciplines on the importance of academic discourse in the tertiary-level educational establishments, i.e. recognition that understanding the disciplines involves understanding their discourses" (Hyland, 2000: 2). Although spoken forms of academic discourse have also been given attention recently (Flowerdew, 1994; Swales and

Malcewski, 2001; Biber, 2003, 2004, 2006; Csomay, 2004, 2006, 2007), most of the studies have focused on written products – academic texts in different disciplines, such as academic articles, theses and dissertations, academic essays and reports. Researchers typically focus on the rhetorical structure and generic specificity of academic texts or the conventions of academic writing in different academic discourse communities (see an overview of research into academic discourse in Grabe and Kaplan, 1996; Kaplan and Grabe, 2002). In Latvia, English written academic discourse has been studied from applied linguistics' perspective by Kramiņa (2000, 2007, 2009), Karapetjana (2007, 2008), Iļjinska (2007), Surkova (2008) and spoken discourse – by Ozola (2007). However, computer-mediated academic discourse has not yet been investigated.

Definition of Academic Discourse

The definition of academic discourse is complicated by its broad and, at times, informal use in the discipline of *academic writing,* which positions itself across university disciplines (Jordan, 1997; Young, 2006). There is an assumption among scholars that academic discourse is simply something that distinguishes an educated person from an uneducated one, irrespectively of the academic discipline. It is considered to be a special way in which academics communicate among themselves in writing (Swales, 1990; Elbow, 1998).

To find a more specific/scientific definition of academic discourse in the literature appears to be a rather difficult task. The specialists "teaching academic writing across the curriculum" (Young, 2006) either avoid giving definitions at all or tend to give all-inclusive definitions embracing all the areas related to teaching written English for educational purposes. This can be illustrated by a broad range of the use of the word *academic*, which may include or exclude lower levels of education (secondary and even primary) apart from the university level. At times, the scope of the word overlaps with *scientific discourse* or *professional discourse*, as the boundaries are rather fuzzy.

Moreover, the task of defining academic discourse is further complicated by the differences in meaning in the use of the term *discourse* (previously discussed in the present paper in section 1.1.). For example, in an ostensive definition given by Elbow (1998), academic discourse is "a discourse the academics use when they publish for other academics." This definition implies the existence of explicit or implicit agreement among academics on the conventions of writing that may be traced by "regarding linguistic features as regularities of academic style" (Hyland, 2000: 1). Other scholars agree that such conventions are decisive and that "...successful academic writing depends on the individual writer's projection of a shared professional context" (Hyland, 2000: 1). It is not clear, however, what is meant by the

term *discourse* in Elbow's definition: the use of linguistic means, the rhetoric, the process of communication or the ways to practice social hegemony. For example, in the fields of the *sociology of language* and *critical discourse analysis*, academic discourse is ascribed a decisive role in social knowledge construction:

> Academic discourse is the principle means by which knowledge is constituted in the world today and English is the globalized language in and through which such knowledge most often gets constructed and transmitted. Be it in the form of specialized books, disciplinary journals, international congresses or university lectures, the influence and power of such discourse is enormous.
> (Silver, 2006: 1)

Academic discourses are seen by Hyland as "collective social practices" in the disciplinary cultures and texts – as "the most concrete, public and accessible realisation of these practices" (Hyland, 2000: 1). Academic texts are viewed as "the lifeblood of the academy" through which *academic discourse communities* re-establish themselves, "as it is through the public discourses of their members that disciplines authenticate knowledge, establish their hierarchies and reward systems, and maintain their cultural authority" (Hyland, 2000: 1). Thus, academic discourse is seen as "community-specific," i.e. depending on the conventions elaborated in a specific academic discourse community formed out of academics working in one particular discipline.

However, academic discourse communities are not viewed as "static and deterministic" or "monolithic and unitary" (Hyland, 2000: 1). While socio-cultural differences between participants and their communication purposes influence their use of language, these differences are not exclusive. The members of one community may successfully travel to another academic community, or they can participate in collaborative projects and multidisciplinary research, creating conditions for the emergence of new academic discourses. According to Harris (1989: 17), "The borders of most discourses are hazily marked and often travelled, and ... the communities they define are thus often indistinct and overlapping." This especially is true at the time of dramatic changes caused by globalisation of the academic world, spread of new forms of interdisciplinary collaboration in research and internalisation of high education. Academic discourse is sometimes viewed by scholars as a global phenomenon. This implies that similar academic writing conventions are observed among academic discourse communities all over the world and academics comprise a global academic discourse community (Coxhead, 2000).

Another definition that may illustrate such a view is a definition offered by Cazden (1988), who defines *academic discourses* as "modes of language and interaction that help identify someone as a competent member of an educated community." The use of the word "competent" in this definition explains why the study of academic discourse has mostly

focused on the study of 'universal' academic writing conventions and the analysis of academic texts produced by students which are compared with professionally written exemplary texts. Furthermore, the emphasis in this definition on social interaction among academics by using different "modes of language" implies that not only written but also spoken and computer-mediated modes of language are regarded. However, the term *mode* of interaction, in its broadest sense, may have various meanings. Broadly understood, it may designate all modes of interactive communication among humans, including computer-mediated inter-changes (Herring, 2001). Notwithstanding the evident benefits of Cazden's definition for the study of CMAD, it cannot be accepted as the conceptual definition in the present dissertation due to its excessive breadth and imprecision. The scope of the word *discourse* in the Cazden's definition equals to *mode* of language or *mode* of interaction. Thus, this definition brings the understanding of *discourse* back to as a narrow view of it as just spoken language or text or to as a broad one as understanding discourse as human communication.

Discourse, however, has already been defined in the present dissertation as a dynamic process of language use in situational contexts. Applying this definition, the author of the present research defines *academic discourse* as a dynamic process of functional use of language as a means of communication in a situational context marked 'academic.' It means that the language users in academic discourse are the members of an academic discourse community, i.e. educated members of society that represent academic disciplines. Academic discourse is realized in semantically connected and meaningful to the members of a particular academic discourse community (Hyland, 2000) verbal instances of spoken or written language longer than a sentence.

This definition helps to build the conceptual basis for empirical investigation of computer-mediated academic discourse as it enables to see how language forms are determined by their discourse functions. The use of linguistic features in academic text types is determined by the functions they are used to perform in the texts. For example, the use of conjuncts is explained by their function to mark explicitly logical relations in academic texts. Another illustration is agentless passive that is used to make an impression of impersonal, objective writing. The same target is aimed at by avoiding personal pronouns in academic texts. Thus, different linguistic features are used to perform the same or similar discourse function to create the texts acceptable for a particular academic community. The process of academic language use has a number of specific linguistic characteristics, which are discussed in the following section.

Characteristics of Academic Discourse

It has been confirmed by many researchers that linguistic characteristics of academic discourse depend on the *mode of communication* (Swales, 2004), i.e. whether written or spoken language is used. As a process of language use in academic context, academic discourse may take the form of spoken or written discourse. Further, the characteristics of written and spoken academic discourse are discussed.

Characteristics of Written Academic Discourse

Scholars investigating language used in academic settings typically agree on the importance of academic discourse research for the English language learning and teaching at university and its general characteristics. However, there is no consensus on specific characteristics and the nature of linguistic variation in written academic discourse. In "a purely formal view of academic writing" (Hyland, 2000:4), dominating in English for Academic Purposes (EAP) approach to teaching academic writing, disciplinary and other contextual variations in language are disregarded for the reasons of practical teaching. According to Hyland (ibid.), "By ignoring context it was possible to ignore variation and to marginalise language itself as simply a set of skills for clearly communicating ideas from one person to another." Academic discourse is viewed as static, conceived and unchangeable aggregate of idealised academic genres (thus, not a process) represented in samples of published academic texts, which are traditionally viewed as "objective, rational and impersonal" (ibid.). These texts are set as examples for students who are supposed to develop generic skills for reproducing such texts. In summary, such a view of academic discourse is presented by Jordan, who sees it as

> formal in an impersonal or objective style (often using impersonal pronouns and phrases and passive verb forms); cautious language is frequently used in reporting research and making claims; vocabulary appropriate for particular academic contexts is used (this may involve specialist or technical words); the structure of the writing will vary according to the particular type (genre), for example, essay, report, etc. … [It] contains references to other writers' publications, sometimes including quotations.
>
> Jordan (1999: 88)

A principally different view on the characteristics of written academic discourse is expressed by Hyland (2000). He draws attention of scholars towards "textual variation, not only in the content of the text we examine in a particular discipline, but in the structure of those texts and the kinds of rhetorical strategies they allow" (ibid: 4). Hyland has studied

social interaction in published academic writing, i.e. "relationships between people, and between people and ideas." The focus of his interest is on "why members of specific disciplines use language in the ways they do" (ibid: 1). Texts are seen and studied as outcomes of social interactions, in which "...individual and social purposes interact with discourse features at every point of choice and in every genre" (ibid: 12). The recognition of the variation in academic discourse is an advantage of the approach of Hyland. However, the scholar applies only a qualitative, descriptive approach to the study of linguistic variation in academic texts across disciplines. The drawback of his research is that it is rather impressionistic and takes a predominantly critical view on disciplinary academic discourses instead of providing robust empirical evidence of linguistic variation in characteristic features.

Another view on variation in linguistic characteristics of academic texts is presented by Smith (2004), who introduces and defines a *mode of discourse* as a linguistic category for the local level of analysis, similar to *text type* in the field of rhetoric, attempting "to right the balance" distorted, in her opinion, by recent over-popularity of lexical and pragmatic qualitative approaches to the study of texts. According to Smith,

> It has sometimes seemed, though, that nothing at all is conveyed by linguistic forms, while everything is due to pragmatics or lexical content. I attempt to right the balance here, at least in part. I propose a local level of discourse, the Discourse Mode, which has linguistic properties and discourse meaning. I posit five modes: Narrative, Report, Descriptive, Information, and Argument.
>
> Smith (2004: 1)

Significantly, the title of Smith's book is *Modes of Discourse: The Local Structure of Text*, signalling that the author attempts to return the study of text as close to the sentence level as possible. Smith's "modes of discourse" are characterised by their linguistic features, which are regarded by the scholars as "grammatical forms with consistent interpretations" (ibid.). Although the linguistic features of the "modes of discourse" are not overtly marked in academic texts, Smith defines them as linguistic categories that have characteristic patterns of distribution in different academic texts.

Researchers studying academic discourse seek to distinguish the linguistic features that are common in academic texts at different levels of language analysis, e.g. at the levels of *register* and *genre*. The study of academic texts has been undertaken by many prominent scholars investigating generic variation in academic texts (e.g. Chafe and Danielewicz, 1987; Swales 1990; Bhatia, 1993, 2004; Bazerman and Prior, 2004). Text and genre analyses approaches to academic discourse aim "to take account of the conventions that govern such genres" (Hoey, 2001). The researchers working in the field of academic genre analysis form many different groups and schools of thought. The rhetorically-oriented approach to genre

analysis represent, for example, Berkenkotter and Huckin (1995). The second approach, routed in applied linguistics and language teaching, is represented by Swales (1990), who has applied the method of analysis of functional *moves* to the study of introductions to academic articles and Australian linguistic school (Halliday, 1985; Christie and Martin, 1997). The method has been further developed by Bhatia (1993, 2004) as well as Dudley-Evans (1994), Paltridge (1996), and Flowerdew (2001).

Contrary to a traditional view, which is characteristic to the discipline of academic writing, that academic genres are relatively stable, genres have been found by the researchers to be dynamic, changing with time (Devitt, 1993). For example, Mair and Hundt (1999) have discovered that written academic genres have become more informal and have acquired more features of spoken genres over time. Hoey mentions two factors that may cause the appearance of *merging genres* (i.e. breaking genre conventions): the creativity of the authors and their attempt to affect the readers. According to Hoey (2001: 6), "Text does have patterning, and genres do conform to convention, but it is always possible to deviate from the expected, buck the convention. Text is one of the places where we can show most creativity." Surprisingly, the study of generic variation in computer-mediated academic discourse has not yet attracted sufficient attention of scholars.

In contrast, quite a number of studies has analysed register-specific differences between academic texts applying corpus-linguistics research methods (Biber, 2003, 2004, 2006; Conrad, 2001; Csomay, 2004, 2006). Most of the researchers report that the use of linguistic features in written academic discourse differs from that in non-academic writing. Biber et al. (1999), Chih-Hua (1999) have studied personal pronouns in academic texts. According to Biber, the use of personal pronouns with human reference is the least frequent in academic prose in comparison to conversation, fiction and news reports. He has found that the most frequent pronoun in academic texts is *it*, followed by *we* and *they*. Other personal pronouns are reported to be rare in academic prose.

Other grammatical features occurring in academic texts have been in the focus of studies of many scholars. Halliday (1988), for example, has found the use of nominalisations and complex noun phrase structure characteristic to written academic discourse. Both short (agentless) and long (*by*-) passives have been found more frequent in written academic discourse that in spoken discourse, fiction and news. Short passives are more frequent than long passives. The syntactic position of the former with dynamic verbs is more frequent than with stative verbs. The latter have been found characteristic to academic prose and extremely infrequent in spoken non-academic discourse (ibid: 938).

Other scholars have reported the findings on the use of pragmatic features in academic texts, such as politeness markers (Myers, 1989), hedging devices (Grabe and Kaplan, 1997; Holmes, 1990; Hyland, 1994, 1998). Discourse markers have been found rather frequent in academic texts. Halliday and Hasan (1976) give a detailed account on many discourse markers functioning in academic texts as cohesive devices. Schiffrin (1987) has also investigated discourse markers and has found that their function is to aid coherence and cohesion in an academic text. Siepmann (2005) has studied meaning and functions of multi-word second-level discourse markers, such as exemplifiers, reformulators, resumers and inferrers. He has provided a functional taxonomy of discourse markers and contrastive analysis of their use in non-native academic writers' production. In Latvian context, discourse markers, as contextualization cues, have been studied by Brēde (2004).

Researchers have also identified lexical characteristics of writing (including academic writing), which is characterised by high lexical richness, and originality (Read, 2000; Surkova, 2008). Lexical specificity (TTR + MWL), studied by Biber 1988, has also been found characteristic to formal, academic texts "that have greater precision of meaning and a high lexical diversity" (Malvern et al., 2004: 193). Other studies report on specific use of private and public verbs in academic prose characteristic to research articles (Hunston, 1995; Thompson and Yiyun, 1991). Coxhead (2000) has distinguished a list of words that are the most frequent academic words across many different academic disciplines.

In general, the researchers studying written academic discourse report a relatively high frequency of the linguistic features that characterise academic texts as abstract, impersonal, and uninvolved (Biber, 2004).

Specific characteristics of spoken academic discourse

Spoken academic discourse has attracted close attention of scholars only recently when the development of modern information-communication technology has enabled the study of different forms of spoken interactions in academic settings (Biber et al., 2004). Spoken texts, such as university lectures, conference presentations and other predominantly monologic oral academic presentations are usually first written and then orally produced. Researchers traditionally study them by recording, transcribing and then analyzing as written texts. Nowadays, technology makes it possible to study a range of spoken and written academic registers used in universities.

The question that has attracted the most of attention of scholars is whether there is a binominal dichotomy between academic speech and academic writing (i.e. spoken discourse versus written discourse) or a "gradation from conversation-like one-to-one research speech to writing-like formal lectures" (Swales, 2004: 27). To investigate this question, many

researchers have focused on the overall organization and discourse management in university lectures (Flowerdew, 1994; Swales and Malcewski, 2001). Vague language, e. g. indefinite pronouns, and hedging, e.g. *at about*, (Poos and Simpson, 2002), known as characteristic to spoken communication, are reported as widely used in academic university lectures for the reasons of modesty or uncertainty or "as a way of socializing students into their particular discipline" (Swales, ibid.). Spoken academic genres are reported to contain reflexive language, transition signals and different kinds of other discourse markers (Swales and Malcewski, 2001). Researchers who study spoken academic discourse found it to be rather informal. Biber et.al. (2004) report that their corpus-based study conducted in the United States universities revealed that academic spoken discourse is characterized by a higher degree of involvement than corresponding written discourse. Summarising his review of the research into academic spoken discourse, Swales (2004: 28) states that "academic speech turns out to be much 'more like' ordinary conversation than academic prose."

There have been found, however, some differences between informal spoken communication and academic speech. Cazden, (1988) reports that the use of rhetorical devices in academic and non-academic spoken discourse differs in that there are different question-response patterns in them. More differences have been found in the corpus of British (Cambridge and Nottingham Corpus of Discourse in English – CANCODE), as Carter (2001) reports. For example, idioms and metaphors are relatively rare in academic speech, in comparison with non-academic speech. Swales presents a comparison of fifteen the most frequent nouns in the corpus of English research articles provided by Hyland and by the corpus of Research Speech events (MICASE). He has found that, although the lists overlap to some extent, "the most frequent of the high frequency nouns in research speech are drawn from what we call the common stock" (2004: 29). He concludes that the most common nouns in research articles are those that represent the process of research, e.g. *study, result, effect*, while the most common nouns in academic speech represent verbal activities, e.g. *word*, *question, language* (ibid: 30). The use of rhetorical devices in spoken academic discourse has also attracted attention of Cazden, (1988). In sum, although academic spoken discourse has been found by researchers considerably less formal than academic written discourse, it is significantly different from non-academic spoken discourse.

The studies mentioned above have investigated linguistic features of academic discourse across a range of spoken and written registers and genres. Nevertheless, the types of texts representing computer-mediated academic discourse have not been sufficiently researched. Meanwhile, in other linguistic subfields, the interest to the study of language mediated by computer has been growing (Crystal, 2001; Fox, 2003; Shortis, 2000), especially in the field

of *computer-mediated discourse* (CMD) introduced by Herring (2001). The next section is devoted to the discussion on CMD.

1.1.3 Definition and Characteristics of Computer-Mediated Discourse

The studies into computer-mediated discourse (Ko, 1996; Warschauer, 1996; Yates, 1996; Davis and Brewer, 1997; Snyder, 1998; 2002; Biber, 2003) emphasise the impact of computer on language use, especially in synchronous text-type computer-mediated communication (Herring, 2001). The scholars, while generally agree on the importance of the study of the phenomenon, demonstrate a broad range of opinions on the terminology and the research approaches to its study.

Notwithstanding a relevantly young age of the field, the process of the study of computer-mediated discourse has been fast and intensive. However, it is complicated by partial incompatibility of different disciplinary approaches to scientific conceptualisation of CMD, as the area is at the crossroads of research interests of linguists, sociologists, psychologists, linguistic anthropologists, specialists in communication studies etc. Not only have these disciplines different from linguistics' research questions to answer, different purposes to study CMD and different goals to achieve, they all tend to apply the terminology and research methods appropriate in their own disciplinary research tradition. Recognising the importance of multidisciplinary approaches to the study of such a complex phenomenon related to language use by humans as CMD, the author of the present research, however, seeks to define it in primarily linguistic terms. Therefore, the following short overview of the scholars' attempts to define and classify the process of language use mediated by computer is restricted to the works of linguists.

The study of CMD began soon after the emergence of it in late 1960s. Computer-mediated communication started as an area restricted in the interests of US national defence and initially was available only to computer scientists and a limited number of organisations. The first global commercial network – the Internet – was started in early 1980s (Herring, 2001). It remains the largest public electronic network by today and is constantly growing.

During a relatively short time following the birth of the field, numerous attempts have been undertaken to conceptualise CMD in applied linguistics. The influence of the use of computer to mediate human-to-human verbal communication on the process of language use was investigated at that time by Baron (1984), Ferrara et al. (1991), Collot and Belmore (1993), Schiffrin (1994), Hale (1996), Warschauer (1996), Ferris (1997), Davis and Brewer (1997), Hawisher and Selfe (1998). Many researchers have written extensively also in the 21st

century, investigating how language is shaped by a wide range of new technologies (Shortis, 2000; Crystal, 2001; Herring, 2001; Posteguillo, 2002; Aitchinson and Lewis, 2003; Fox, 2003; Inman, 2004; Boardman, 2005; Herring and Paolillo, 2006; Nelson, 2006).

There have been numerous attempts to define the concept of the process of language use mediated by computer as a new linguistic phenomenon. In the early studies of CMD, a variety of terms were introduced to denote it: *interactive networking* (Baron, 1984), *interactive written discourse* (Ferrara, Bruner and Whittemore, 1991), *electronic language* (Collot and Belmore, 1993), *electronic discourse* (Davis and Brewer, 1997), *electronic networked discourse* (Hawisher and Selfe, 1998), *telecollaboration* (Warschauer, 1996, Belz, 2003). The variety of terms introduced to denote CMD was caused by incomplete understanding of the nature of the phenomenon at the time, as well as by a variety of the definitions of the term *discourse* itself in the disciplines concerned.

Ferrara et al. (1991) were among the first scholars who attempted to term the new linguistic phenomenon. They referred to it as to homogeneous "interactive written discourse", as the only possible mode of computer-mediated communication at the time was the synchronous mode. In such a mode of communication, the communicators take turns exchanging messages that they type on the keyboard. The messages then appear on the computer screen of the other party. That the messages are typed and exchanged between the participants explains the appearance of the words *interactive* and *written* in the term proposed by Ferrara et al. What is missing in the proposed term is a word to denote the medium that is used for communication, as "interactive written discourse" may also be the one which makes use of other media, e.g. paper, chalk and board etc.

For that early period of CMD studies, bold claims were characteristic of researchers, such as that, for example, made by Collot and Belmore in the introduction to their paper published in 1993: "This paper describes a new variety of English, which we have called *Electronic Language"* (1993: 41, italics in the original). Collot and Belmore come from the discipline of language technology, a sub-discipline of computational linguistics, which is concerned with computer languages and human language processing by computer. Nevertheless, they announced the 'discovery' of a 'new' human language, referring to the process of language use in human-to-human communication in specific situational context, i.e. human language mediated by computer. Although they used a well-developed research design and method (based on the methodology of Biber, 1988), the lack of strong linguistic background, evidently, prevented Collot and Belmore from a more careful approach to determining their terms. Consequently, both of the words *electroni*c and *language* in the name

of the 'new language' are too broad in meaning to denote CMD, which is not a new language but the process of language use in context mediated by computer.

More than a decade ago, Davis and Brewer also attempted to define the discourse emerging when people use computers for communication. They, however, restricted the scope of the meaning of the term in their study to only one dimension. The scholars proposed the following definition: "Electronic discourse is one form of interactive electronic communication" (Davis and Brewer, 1997: 1). This definition has become recently rather inaccurate, because it is not specific enough. Firstly, the meaning of the term *electronic* has expanded along with advances in modern technology and may mean not only language in use mediated by computer but also by other electronic devices, e.g. mobile phones, pagers etc. Secondly, the attempt to equate *discourse* and *communication* also leads to unjustified broadening the scope of the term. This is because the term *communication* implies much more than the process of the use of verbal language (discourse), which, in its turn, leads to veering to the research directions that abandon purely linguistic interests and intrude in the tenets of other disciplines. Acknowledging all the benefits of interdisciplinarity in research, the author of the present paper states that the primacy of linguistics proper interests and goals should not be forgotten in linguistic research, as in spite of all the efforts of thousands of linguists, there still are more questions than answers concerning the language faculty as an exclusively human property.

Another inconsistency in the attempt to conceptualise CMD by Davis and Brewer is in the difference between what they propose "electronic discourse" to be and what they have actually studied. The researchers use the term 'electronic discourse' to refer to "the two-directional texts in which one person using a keyboard writes language that appears on the sender's monitor and is transmitted to the monitor of a recipient, who responds by keyboard" (ibid.). Although the researchers recognise that the electronic discourse of this kind, emerging among large or small groups of humans as well as among individuals, could be 'multifaceted and complex', they themselves limit their investigation to the study of "textual artefact resulting from electronic discourse" (ibid.). On the one hand, the researchers recognise the importance of social interaction among the participants of computer-mediated communication and the role of context of such participation; on the other hand, they focus on only written forms of language though claim that they are interested in language in general. Thus, the extent of meaning implied by the term "electronic discourse" seems to be superfluous for the research object in Davis and Brewer' research.

An attempt to name the new linguistic phenomenon has been undertaken by Hawisher and Selfe (1998: 7), who refer to the "discourse of the nets, more accurately called electronic

networked discourse." As in the previous definition, the use of the term *electronic* used in their study has become too broad in the course of technological development. Moreover, the term *networked* is also rather ambiguous: it is not clear whether it emphases the role of the medium, i.e. a net of computers (the Internet) or a social dimension of discourse, i.e. social net, or both. Notwithstanding the ambiguity of the term, it has gained certain popularity, especially in the field of CALL – computer-assisted language learning (Warschauer and Kern, 2001).

Another term that is used in CALL to denote CMD is 'telecollaboration', which Belz and Reinhardt (2004) use to refer to the process of language use in computer-mediated communication between the subject of her research – an American advanced learner of German – and his German communication partners. The scholars state, "Telecollaboration involves the use of Internet communication tools by internationally dispersed groups of language learners in institutionalized settings for the purposes of foreign language linguistic development and the development of intercultural competence" (Belz and Reinhardt, 2004: 325). Thus, Belz and Reinhardt make an attempt to introduce the use of a contesting term denoting, however, only one of further divisions of CMD, i.e. institutionalised purposeful 'bilingual electronic exchanges' for target language learning. Other examples of the use of this term, as cited in Belz and Reinhardt, are Kinginger (2002) and Warschauer (1996). However, as the survey of the most recent literature shows, the term has not found that much support in the academic community as the term *computer mediated discourse* or its orthographic variant *computer-mediated discourse* (Herring, 2001), the traditional abbreviation for both being CMD (or CmD in Belz, 2004). The term 'telecollaboration' is, definitely, narrower in meaning than CMD, distinguishing a small area of study within the field of CMD. However, in contrast to CMD, it makes the emphasis on collaborative language learning in the electronic environment, thus, referring to the field of language pedagogy rather than to linguistic discourse analysis.

In this study, the term CMD is used throughout the paper to widen the application area beyond language pedagogy and to specify the area of linguistic interest – the analysis of language in actual use in the context of computer-mediated communication. The term *computer-mediated discourse* was coined by Herring only in 1995. Herring provides the following definition of CMD:

> Computer-mediated discourse is the communication produced when human beings interact with one another by transmitting messages via networked computers.
> (Herring, 2001: 626)

This definition suffers the same problem of overgeneralization that the definition of Davis and Brewer discussed earlier in the present dissertation (p. 34). The problem is caused by equating *discourse* and *communication*, thus unjustifiably broadening the scope of meaning of the term *discourse*. Although the term communication has a broader meaning than the term discourse in linguistics and includes verbal, nonverbal and graphic types of communication, Herring uses the term *computer-mediated discourse* to only one type – written or "text-based" CMC (Herring, 2001). Nevertheless, narrowing the application of the term to only written (typed) computer-mediated messages is regarded by Herring as an advantage for empirical research into CMD. The nature of computer-mediated discourse is a consequence of the influence of many different factors, for example, characteristics of the medium, mode of communication, linguistic and computer literacies of communicators, etc. Following Crystal (2001), Herring sees "text-based" CMD as a unique environment "in which to study verbal interaction and the relationship between discourse and social practice" (Herring, 2001: 612), as it is "free from competing influences from other channels of communication and from physical context" (ibid.). Herring states that all the various forms of CMD have at least one common feature: "…the activity that takes place through them is constituted primarily – in many cases, exclusively – by visually presented language" (ibid.). With further technological advances, the forms of computer-mediated communication expand to include also spoken/oral CMC (voice mail, stream audio, etc.), as well as the technological facilities for their research. Thus, restricting the scope of the term CMD to only written (typed) language seems unjustified.

Another limitation of the scope of the meaning of the term proposed by Herring is that she restricts the study of CMD to only "reciprocally interactive" forms of it, admitting that the discourse properties of various Web 'pages' "deserve of study on its own terms" (Herring, 2001: 626) as they are monologic and prepared in advance and "constitute a separate phenomenon" (ibid.). However, recently emerging new forms of CMD, e.g. weblogs, which are Web pages in a constant flux, being updated daily and presenting personal journals with the elements of a dialogue between the author of the weblog and the readers: the readers' messages are usually posted on the weblog and are followed by the comments of the weblog owner. The weblogs maintained by scholars, teachers or Ph.D. students often present academic discussions on topical research issues. Another illustration is such a new form of collectively created text-based CMD as on-line encyclopaedias, e.g. Wikipedia, where the entries emerge as a result of collective efforts of many contributors and editors of the contributed texts. On the contrary, many on-line discussions that started as interactive CMC then are systematised and stored as unchangeable archived documents. Other forms gradually

change, differing by the speed of change (different frequency with which they are updated). All this implies that the dichotomy between interactive text-based CMD, in the understanding of Herring, and non-interactive CMD blurs and becomes not so obvious when the border between them is merging, all of the types of CMD becoming interactive to some extent. Therefore, it is difficult to see the way and reason nowadays for restricting the research of CMD to the study of only interactive computer-mediated discourse. Therefore, the present dissertation includes the study of samples of various text-based forms of CMD representing computer-mediated discourse, from the most interactive (chats) to the least interactive (hypertexts). However, it is limited to the investigation of only academic discourse as opposed to non-academic discourse (e.g. personal communication, work place communication etc.).

The author of the dissertation proposes the following conceptual definition of CMD based on the definition of discourse provided in Chapter 1.1: *Computer-mediated discourse is a type of discourse that is mediated by computer. It is a dynamic process of functional use of language as an instrument in different modes of CMC, e.g. chats, discussion forums, e-mails, and Internet publishing, e.g. weblogs, hypertexts. It is realized in textual artefacts (semantically connected and meaningful for the communicating users of language verbal instances of spoken or written language longer than a sentence) that are produced by language users applying computer as a tool for communication.*

The present study is mainly concerned with the investigation of linguistic properties of computer-mediated texts as discourse realisations. Therefore, the study of the process of interaction, the behaviour, gender differences and social relationships of the interactants are beyond the scope of it. Instead, the main focus is placed on the variation in the use of linguistic features in text types produced in the process of language use in academic settings, i.e. for communicating on the subject matter using different types of communication technology.

Medium-specific characteristics of CMD

Computer-mediated discourse can be spoken and written. The present study investigates text-based computer-mediated discourse discussed above. Whether it is presented in the written form on the World Wide Web or exchanged in the form of written messages via the Internet it is the main form of CMD prevailing in universities. Additionally, the study addresses CMD in its 'spoken' form – orally produced at on-line academic seminars and transmitted via computer. This form of CMD is also text-based. First, the text of presentation is written by a presenter. Then, while presented, it is digitally recorded and transmitted to

listeners. The sound files are then transcribed back into a written form, this time, however, including all the speakers' 'deviations' from the original written text notes, including spoken formulas, specific lexical bundles, discourse markers, hesitations, repetitions etc.

Scholars investigating CMD, have noticed that as *a medium* of communication, computer is not neutral, which means that it influences the process of language use (Segerstad, 2002). Herring (2001) reports that CMD is distinct from spoken and written face-to-face discourse and, thus, should not be regarded as simply writing transmitted via electronic network. She enumerates four main reasons for distinguishing CMD from the discourse transmitted via other media. According to Herring, the reasons are the following:

1. The process of language use is faster than in written but slower than in spoken communication.
2. Any number of participants can be engaged in CMD simultaneously.
3. Participants use specific to only the electronic medium ways of communication due to the cognitive constraints on attending more than one exchange at a time.
4. Specific to the electronic medium *public/private* factor, i.e. it makes an impression of direct and 'private' communication, while there are unseen multiple other participants in communication.

(Herring, 2001: 614)

Computer as a medium may transmit one or more channels of communication. Contrary to face-to-face communication, which is always multimodal (Norris, 2004), CMD is a "lean" medium (Herring, 2001), at least in the forms that are prevailing at the tertiary level of education. This means that in face-to-face communication multiple channels – visual, auditory, gestural, etc. – are used to transmit information. In contrast, information in CMC is often limited to written text available only through the visual channel. Language users have to invent new means to compensate for the information that is supposed to be transmitted by the missing channels. For example, writers use graphical signs (emoticons) or glosses, i.e. verbal descriptions of actions or feelings (Crystal, 2001).

According to the type of interaction, CMD is represented by two distinct ways of exchanging verbal messages by participants: synchronous and asynchronous interchanges. Synchronous computer-mediated discourse is the process of language use in real-time, simultaneous (or semi-simultaneous) computer-mediated conversations that may be nowadays transmitted via different Internet channels: video conferencing, audio conferencing, audio-graphic (Hampel, 2003) or text-based chats (Herring, 2001). It is realized in linguistic production that language users participating in CMC produce by typing or speaking.

Asynchronous computer-mediated discourse is such a process of language use in human-to-human interaction transmitted over computer networks when the messages posted by interlocutors are separated in time. The characteristic feature of asynchronous CMD that

distinguishes it from synchronous CMD is that the communicators are not logged into the system simultaneously but post and read messages at the time convenient to them. It is realized in linguistic production that language users participating in CMC provide by typing text messages and sending them to other participants. The literature review on synchronous and asynchronous CMD in academic settings is provided in 1.1.4.3.1.

1.1.4 Computer-Mediated Academic Discourse as the Research Object

The emergence of a new vast area (CMD on the Internet), where new context for academic language use may have an impact on the language as a system, has attracted attention of many prominent scholars: Aitchinson and Lewis (2003), Boardman (2005), Chrystal (2001) Inman (2004), Ferris (1997), Fox (2003), Hale (1996), **Nelson (2006),** van Waes, Leijten and Neuwirth (2006), Posteguillo (2002), Shortis (2000). A number of corpus-based studies into register variation in CMD (Biber, 2003, 2006; Collot and Belmore, 1993; Ferrara, Brunner and Whittemore, 1991; Yates, 1996) report on the findings that reveal the tendency to what Shortis (2000: 24) named 'informalisation' – occurrence of features characteristic to informal/spoken language in written computer-mediated communication discourse. The changes in traditional genres, including academic genres on the Internet (e.g. electronic theses and dissertations), have been noticed by many researches (Bhatia, 1993, 2004; Eggins and Martin, 1997; Hoey, 2001; Shepherd and Watters, 1998; Swales, 1990, 2004), who have raised the question of adequate teaching those genres at university. There have also been registered morpho-syntactic and lexical changes in the language used in electronic academic and scientific texts noticed by Crystal (2004), Posteguillo (2002), and Carter (2004). The emergence of specific compounds, e.g. *cyberspace*, *netlinguistics*, *e-zines*, *hypertext* etc. word collocations, e.g. *upload/download the article, to scroll up/down the text, etc.* and longer lexical bundles, e.g. *back to top, see the enlarged image, view Table 1 in another window, click here to view the table* have been registered by Cigankova (2008). These features appear in academic texts at different linguistic levels in the texts across disciplines.

As a possible reason for the changes in academic discourse mediated by computer, the studies in academic discourse (Flowerdew, 2001; Gutiérrez, 1995; Hyland, 2000; Swales, 1990) emphasise the role of new context in formation of academic texts. As computer utilizes different combination of media and modes of communication (Kress and van Leeuwen, 2001; Norris 2004) and offers specific facilities to support a variety of communication ways and styles, it is fair to suppose that it may also influence computer-mediated academic discourse. Therefore, many scholars regard computer as a new context that influences the process of

language use in academic texts circulating in CMC (Cummins, 2000; Dowling, 1998; Edminster and Moxley, 2002; Snyder 1998). They report on the changes observed in academic discourse functions on the Internet and in academic writing practices: new ways of writing, interacting, and reading, and analysed their implications for teaching written academic language (Broady, 2000; Warschauer, 2002; Wysocki et al., 2004). However, CMAD as a new linguistic phenomenon has not been defined in the literature yet, and the characteristic features of it have not been systematically studied.

1.1.4.1 A Conceptual Definition of CMAD

The author of the present research proposes a conceptual definition of CMAD as a type of discourse mediated by computer in academic settings that is derived from the previously defined concept of *discourse* (see section 1.1 of the present dissertation), synthesising the concepts *academic discourse* (section 1.2) and *computer-mediated discourse* (section 1.3) as 'ingredients', since it has the characteristics of both types of discourse. CMAD is defined as *a computer-mediated process of functional use of language as a means of communication in academic context that is realised in semantically connected and meaningful to the communicating users of language verbal instances of spoken or written language longer than a sentence.*

This definition allows the author to approach the systematisation of what is known about CMAD, starting with its classification.

1.1.4.2 A Typological Classification of CMAD

Further in this section, the classification of the most conceived to the moment of writing of the present dissertation types of CMAD is presented. It is an attempt of the author of the present text to systematise the knowledge about CMAD types that has been obtained by previous researchers. Scholars studying academic discourse base their classifications of various types of it on different levels of variation in discourse (the discussion on the concept of linguistic variation in 1.2). The present study investigates the variation of academic register in CMAD caused by computer as a medium of communication.

The author of the present research proposes the following classification of CMAD types as a model of linguistic variation in it (Fig. 1.1). Each type of CMAD is viewed as a unique combination of

- the *type of discourse* (e.g. transactional or interactional)

- the *mode of discourse* (spoken or written)
- the *mode of interaction* in CMC (synchronous or asynchronous)
- the *type of software* used (e.g. e-mail protocol, discussion forum, etc.).

For example, the type of discourse marked as 'discussion forum' in the present study is classified as asynchronous interactional written CMAD, and the type of discourse marked as 'on-line seminar' is characterised as synchronous transactional spoken CMAD. Further, each level of division in the proposed classification of CMAD types is explained.

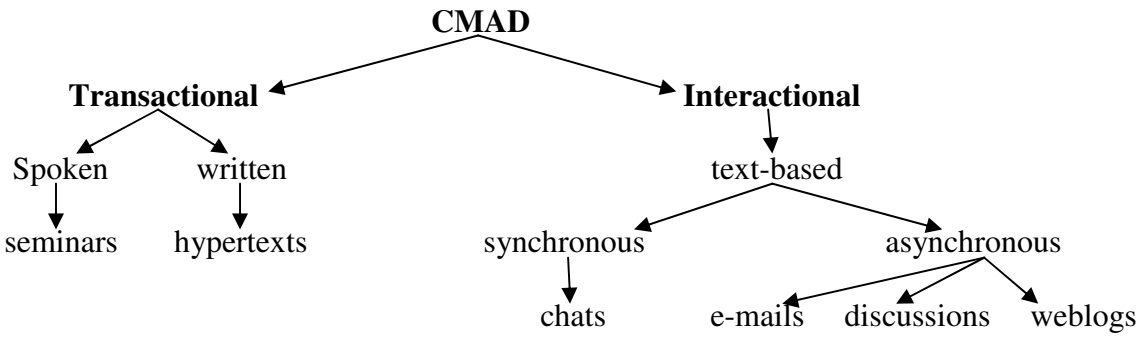

Figure 1.1 Classification of CMAD types proposed by the author of the present study.

The division of CMAD into *transactional and interactional types* has been done according to the type of discourse they represent, i.e. transactional and interactional discourse (Brown and Yule, 1983). Interactional CMAD is a computer-mediated process of academic language use in which text messages are exchanged among the participants of CMC in a form of a dialogue or a polylogue (i.e. a number of simultaneously occurring dialogues in CMC). Language serves an interactional function in this type of CMAD by "expressing social relations and personal attitudes" (ibid: 1). According to Brown and Yule, this language function aims at establishing and maintaining social relationships among the members of computer-mediated academic discourse community: establishing common ground on and sharing points of view on topical academic issues, negotiating the agenda of academic meetings, collaborating in studies and research, etc. Transactional CMAD, in contrast, is such a process of computer-mediated academic language use in which the transactional function of language is performed by expressing 'content', i.e. conveying factual information. English transactional academic discourse is informational in nature, reader or listener oriented, context independent and characterised by precision and clarity of expression (ibid.). Transactional types of CMAD have been further sub-divided into spoken (e.g. on-line academic seminars) and written (e.g. published academic hypertexts) transactional CMAD .

The formation of each type of CMAD also depends on the *mode of interaction* (Norris, 2004), which is understood in the present work as the way in which the process of language use is realised by material means. Mode of interaction applied in each type of CMAD is a unique combination of several material means of communication and the speed of interaction. Interactional types of CMAD have been further divided into two groups: synchronous and asynchronous CMAD (Fig. 1.1). The criterion for the division is a specific characteristic of the mode of interaction in CMC – *synchronicity* (Collot and Belmore, 1993). Synchronous interactional CMAD, represented by synchronous academic conferencing – 'chats', is the process of language use characterised by an immediate exchange of messages or by relatively short time between them. In asynchronous interactional CMAD, the time between messages is usually much longer though, occasionally, the interchanges may be very fast. This type of CMAD is further sub-divided into three groups according to the type of the software used for communication: academic e-mails, discussion forums and weblogs. To be precise, the latter are the representations of a transitional type between interactional and transactional CMAD. Weblogs are only partially interactive. They originated as personal academic Web journals and would be considered as representations of transactional CMAD if the interactive facility for reader response provided by the software were not so actively used. As the owner of the weblog usually comments on the readers' responses daily, the discourse of weblogs is rather interactional than transactional. Additional reasons for classifying weblogs as interactional type of CMAD are that the structure of the text is usually not planned by the author and the writing appears on the screen without much reviewing and editing, thus resembling natural, unprepared conversation on the studied topics between university lecturers and students.

The difficulty with the classification of academic weblogs shows that the presented classification is rather subjective. An attempt to eliminate this drawback and to provide a more objective division of CMAD into types and genres, based on statistical analysis of internal linguistic qualities of CMAD texts, will be presented in Part 2. Meanwhile, the classification provided above was necessary to systematise the knowledge of specific characteristics of CMAD types that had been accumulated before the present research in order to determine the place of the present research among the relevant empirical studies.

1.1.4.3 Description of Linguistic Characteristics of CMAD Types

1.1.4.3.1 Interactional Types of CMAD

Interactional CMAD is a computer-mediated process of academic language use in which text messages are exchanged among the participants of CMC in a form of a dialogue or a *polylogue* (i.e. a number of dialogues or threads of conversation that occur simultaneously in CMC). Language serves an interactional function in this type of CMAD by "expressing social relations and personal attitudes" (Brown and Yule, 1983: 1). This language function aims at establishing and maintaining social relationships among the members of academic discourse community using CMC: establishing common ground on and sharing points of view on topical academic issues, negotiating the agenda of academic meetings, collaborating in studies and research etc. It also includes *phatic* use of language to open or close conversation. Interactional function of language, e.g. negotiation of social roles, peer-solidarity, taking turns in a conversation, politeness and saving face, has been studied by conversational analysts, e.g. Sacks, Schegloff and Jefferson (1974), Brown and Levinson (1978). In the present research, interactional CMAD is represented by synchronous academic chats and asynchronous academic e-mails, discussion forums and weblogs.

Synchronous modes of interactional CMAD

In synchronous modes of CMAD, e.g. real-time Web-based chats (WebCT chatrooms, Blackboard), MOOs (Multi-User Object-Oriented environments), MUDs (Multi-User Dungeons/Dimensions), or chat facility provided by HorizonLive presentation software, the communicators are simultaneously logged into the same network system from different remote locations. They are supposed to respond to the sender's message immediately. The size of the computer programme window in which the readers can see the messages sent by all the participants of a polylogue is limited: the messages move up each time a new message appears in the window while the messages sent earlier quickly disappear from the sight. In the chat rooms with multiple participants sending messages simultaneously, the text that is to be read by readers appears in the window for only a very short time. This imposes specific constraints on communication to which language users try to adjust. Firstly, the message should be as short as possible to be read very quickly; hence, frequent use of abbreviations, short forms etc. Secondly, there are also some constraints on the complexity of the language used: simple sentences rather than complex or compound are typed. Moreover, long sentences are sometimes divided into parts and sent separately to 'keep the line busy'.

Time constraints on communication are the cause of emergence of the specific characteristics of synchronous CMAD: unconventional spelling, frequent uncorrected typographic and grammar mistakes (no time for repair), the use of informal expressions, colloquial forms, etc. There have also been changes observed in text organisation of

messages. Kern (1995), for instance, has noticed lack of coherence in synchronous CMD because of the fast speed of interchanges, in addition to grammatical inaccuracy. Werry (1996) has also observed frequent breakdowns in synchronous CMC experienced by novice users caused by the fact that many interlocutors simultaneously maintain communication using the same dialogue window of the computer programme. The scholar examines the use of linguistic and interactional features in Internet Relay Chat (IRC) – a special computer programme that enables synchronous CMC over the Internet. He finds that the peculiarities of synchronous CMD may be explained by the intention of the participants to imitate spoken discourse. According to Werry (1996: 61), "Communication on IRC is shaped at many different levels by the drive to reproduce or stimulate the discursive style of face-to-face spoken language."

Many other researchers of synchronous CMAD also mention its hybrid nature that is intermediate between spoken and written language (e.g. Pellettieri, 2001, Zitzen and Stein, 2004). Although the texts of messages are written (typed), the choice of lexis, syntactic and discourse structures and rhetorical devices are characteristic to spoken language. Ulrich (2007) has investigated sentence coordination in real-time chats and found that it resembles the coordination in spoken communication.

A number of researchers report that the use of synchronous communication improves speaking skills of language learners. Abrams (2003) has found a positive effect of synchronous CMC on oral language performance. Chun (1998) has discovered an improvement in oral fluency in face-to-face classroom communication in the group of learners practicing text-based synchronous CMC (academic chats) in comparison with the learners in the control group who practiced face-to-face communication. Weininger and Shield (2001) argue that "while the language produced in synchronous, text-based CMC can be regarded as "written speech", constrained by, for example, spelling conventions and lack of extra- and paralinguistic features, it may also be considered in terms of its communicative effect rather than of its channel or medium of production." The researchers state that synchronous CMD "offers an opportunity for learners to acquire fluency in the oral register." The study of Tudini (2003) seems to support this claim while the researcher focuses on repairs, variety of speech acts, and the presence of discourse markers in synchronous academic discourse. Pellettieri (2001) has obtained the evidence on developing grammatical competence of language learners participating in synchronous academic interchanges.

Toyoda and Harrison (2002) have identified the difficulties that non-native speakers of English experience when they participate in synchronous CMD: inability to recognize a new word and negotiate its meaning with the interlocutor, misunderstanding caused by a word

polysemy or wrong spelling, grammatical errors, the use of abbreviations, sudden topic change, slow response and intercultural differences in communication practices.

Nevertheless, other researchers find synchronous CMD useful for developing language learners' communication skills. Carr-Chellman and Duchastel (2000: 236) claim that synchronous CMD is characterised by "greater social pressure for conformity in participation." The researchers consider that the real-time nature of communication in synchronous CMD and social pressure caused by it enhance developing communication skills for discussion, team-building and collaborative knowledge construction. Some scholars have studied gender issues in synchronous CMD and noticed that synchronous CMD encourages active participation in discussions of female language learners who often avoid it in face-to-face classroom discussions dominated by more active male learners (New and Green, 2001, Herring, 1996). Jeong (2005) has studied software tools that could help to measure and identify the differences in participation and linguistic patterns in message-response exchanges in synchronous CMC.

In terms of internal linguistic characteristics, the language used in synchronous CMAD differs from the language used in both asynchronous CMD and in oral face-to-face discourse. Having compared synchronous CMD and face-to-face interchanges, Ko (1996) describes synchronous CMD as characterized by lower lexical richness, i.e. a combination of lexical diversity and complexity, than spontaneous speech. Sotillo (2000) has compared linguistic production of English language learners in synchronous and asynchronous CMD. The researcher has discovered that synchronous academic discourse resembled oral face-to-face conversations in terms of quantity and types of discourse functions. The language output in synchronous CMAD has been found less lengthy and containing more sentence fragments. Additionally, the language produced in synchronous interchanges has been found less syntactically complex, e.g. comprising less subordinate and embedded subordinate clauses, than the language used in asynchronous academic discourse. The scholar concludes that the two types of CMAD should be used in ELT for different purposes: synchronous chats for developing fluency in spoken communication and asynchronous discussions for improving accuracy in academic writing.

A radically different view has been expressed by Zitzen and Stein (2004), who point out that chat and conversation "are related, but different genres." According to Zitzen and Stein, synchronous CMD is not spoken language written (or typed) but a new genre of written English that has its own medium-specific characteristics. As such, it is characterized by different from either spoken or written language choice of linguistic features by language users. In particular, the scholars mention the increased use of specific metalanguage for

pragmatic purposes (such as glosses, e.g. <rolling on the floor> or *lol* – abbreviation for *laughing out loud*), which is explained by the need to compensate for the absence of body language and other non-linguistic information in synchronous text-based CMD. As Zitzen and Stein state,

> In digidiscourse, a heavier functional load is placed on linguistic signalling and linguistic choices that is instrumental in steering the options in terms of discourse strategies and the creation of distance or closeness perceptions in, for example, chat.
>
> Zitzen and Stein (2004: 1017)

Zitzen and Stein express a doubt that synchronous CMD can be successfully used for developing language skills that are fully applicable in spoken communication. They point out that a possible effect of using chats for developing speaking skills is "totally unexplored in terms of empirical study." This is because the choice of linguistic features in synchronous CMD may be totally different from that in spoken discourse that is rich in contextual cues. Responding to the plea of the researchers to investigate synchronous CMD empirically from the applied linguistics perspective, the author of the present dissertation has compiled a specialized corpus of synchronous academic chats and studied them as the research object (2.1.4).

Asynchronous modes of interactional CMAD

Three modes of asynchronous CMD that are widely used in the academic settings have recently attracted the most attention of scholars: *academic e-mail, discussion forums* and *weblogs*.

Academic e-mail

Electronic mail (e-mail) has recently become a ubiquitous interpersonal communication medium, particularly in academic settings, because of its speed and relevant ease of use. There exist a large number of different computer programmes enabling e-mail exchange that insignificantly differ in some respect. Modern e-mail software allows sending animated e-mails and texts with incorporated video and audio files, hyperlinks, emoticons, etc. However, these facilities are rarely used in academic settings while are common in, for example, e-commerce. Academics usually avoid wasting time and effort on electronic "bells and whistles" where the essence of the message can be expressed in words. The computer programme is usually equipped with a style guide for writing letters which can be consulted by a writer. This partially explains why e-mail messages resemble conventional letters.

A number of scholars have studied the discourse of e-mail (Yates, 1996; Baron, 1998, 2003; Emmerson, 2004; Moran and Hawisher, 1998; Gimenez, 2000; Sims, 1996) including the language use in academic settings (Chen, 2006). E-mail has been investigated from different perspectives in different disciplines: sociology, pedagogy, anthropology, psychology

etc. In applied linguistics, the main focus has been on the language in its actual use in e-mail (Herring, 1996, 2001; Sunderland, 2002) or on the use of e-mail for teaching a foreign language (Liaw, 1998; Naya, 1999) and developing literacy (Moran and Hawisher, 1998). In language teaching, e-mail is considered to be very important means for increasing student-instructor and student-student communication in on-line and hybrid university courses (Pincas, 1998). According to Carr-Chellman and Duchastel (2000: 236), the traditional e-mail function "is extremely useful for student-instructor communication, for instance with respect to assignments, progress, feedback, and administration."

The empirical research that has been conducted by researchers since e-mail emerged in 1990s has revealed the characteristics of this type of CMAD. A number of scholars have noticed that e-mail encourages a less formal way of communication among language users in corresponding academic settings.

Yeats (1996) has published the corpus-based study of the language of academic e-mails, comparing samples of written and spoken languages with samples of electronic discourse along such continua as type/token ratio, lexical density, degree of personal presence and modal auxiliary use, claiming that the results of his study suggest that computer discourse is neither writing nor speaking but a new variety of language that needs further studying.

Ma (1996) has analysed the e-mail messages exchanged among 25 students in East Asia and in North America and reports that e-mail facilitates the intercultural aspect of CMC, in particular, "a more egalitarian and information-oriented experience than FTF [face to face] intercultural communication" (1996: 179).

Luuka (1998) has analysed the language used in Finnish-Estonian scholarly e-mail communication. The researcher reports that the mode of discourse was rather written than spoken-like; however, the means of achieving coherence more resembled spoken than written interaction. The tenor of discourse is reported to have been equal, co-operative, factual and non-personal.

Sunderland (2002) reports on the results of the empirical study of the e-mail language of fourteen Romanian Ph.D. students in Applied Linguistics in Lancaster University. She has applied language-focused content analysis, coding the messaged into two groups: responding and initiating, in her research of 164 e-mail messages sent by the doctoral students to the programme coordinator. She has identified a number of medium-specific peculiarities of e-mail language: "First is the amount of vocabulary of an affective nature: *hope, glad, finally, nice, addicted, desperate, great, excited, hate, happy"* (Sunderland, 2002: 237). The second surprising thing discovered by the researcher was that all the messages consisted of grammatically correct sentences and did not contain the features of spoken language, in

contrast with earlier reports on the use of language in e-mail by English native speakers. Other features of the language of e-mail revealed by Sunderland were the use of metalanguage referring to the medium itself and the use of special expressions to create the sense of immediacy, also reported earlier by Moran and Hawisher (1998). Sunderland also reports on the intensity and lexical richness of the e-mail messages and presents the breakdown of the messages by the functional use of e-mail by the students (ibid.).

Ancarno (2005) has published the results of a corpus-based study investigating the pragmatics of academic e-mail. The researcher has investigated 86 e-mail messages of both native and non-native speakers' of English and compared them with 66 conventional paper-format academic letters. As Ancarno reports, his findings suggest that academic e-mail is comparable in formality with traditional academic letters while academic e-mails written by native speakers of English are more formal than those written by non-native speakers. Ancarno has found the evidence that academic e-mail, although still being influenced by the style of conventional academic letter, is in the process of formation as a new type of academic discourse. However, the classification of academic e-mail as a new genre by Ancarno is arguable, as the computer programme enabling sending e-mails does not determine the register or the generic characteristics of the messages sent. The register is determined by the choice and functional use of linguistic features in particular academic context, and the genre of the message – by the intended audience and the purpose of writing rather than the choice of the software. Thus, as e-mail genres could be regarded investigated by the researcher four e-mail types: e-mails of request, of offer, of apology and e-mails expressing gratitude. The author of the present study argues that academic e-mail should be classified as a type of CMAD, not as a genre. As such, it could be further characterized as employing academic register and could be divided into sub-groups according to the discipline they represent (disciplinary discourse) and further sub-divided according to the purpose of writing (genre), e.g. e-mails maintaining relationships, informational e-mails, invitations, e-mails of application, etc.

The problems in the use of language in e-mail messages experienced by non-native speakers of language in academic communication with academic authorities have been addressed by Chen (2006 on-line). The researcher has traced the process of development of e-mail literacy in the target language in the case study of one Chinese doctoral students' correspondence with university authorities in the United States. The study has revealed that the experience in informal peer-to-peer communication did not help the student to avoid the problems in her communication with professors. Chen has identified a number of pragmatic problems that students experience in academic communication by e-mail:

- "delayed purpose statements with many irrelevant details,
- requests framed from a student-oriented perspective and with a strong help-needed tone,
- failure to demonstrate status-appropriate politeness,
- ineffective use of reasons or explanations as supportive moves" (ibid.)

The scholar insists that students should be taught explicitly in the language classroom how to use language in e-mail, as neither the experience in writing formal letters nor in participating in informal peer-to-peer e-mail communication seems to be sufficient for effective academic communication via e-mail.

Academic discussion forums

Another asynchronous type of CMAD is the process of language use in on-line academic discussion forums (also known as on-line conferences, bulletin boards, delayed-time mainframe conferences).

Many researchers emphasise the highly mediated nature of interactions in asynchronous CMAD (e.g. Warschauer, 1996; Warren and Rada, 1998; Light and Light, 1999; Motteram, 2000; Murray, 2000, 2005; Thomas, 2002; Abrams, 2003; Savignon and Roithmeier, 2004). Walther (1996) has identified impersonal, interpersonal, and hyperpersonal patterns of interaction in asynchronous CMAD. Beuchot and Bullen (2005) and Jeong (2005) have investigated interpersonality and interaction patterns by analysing message-response sequences in asynchronous CMAD.

Some scholars have studied the implications of the use of synchronous discussion forums for collaborative language learning, insisting that they enhance the learning process. Schrire (2006) has studied asynchronous CMAD at different levels: discussion forum as a whole, the discussion threads, the messages, and the exchanges and moves among the messages. The researcher has revealed the correlation between the variables important for understanding the process of knowledge building in asynchronous discussion groups, e.g. the type of interaction and move in the exchange structure. Moore and Marra (2005) have analysed asynchronous CMAD of thirty-seven graduate students using electronic discussion software in a university course. The results of their research reveal that the structure of the discussion influence the process of communication and knowledge building, which has implications for the use of asynchronous discussion forums to improve the quality of learning. Veermans and Cesareni (2005), having investigated synchronous and asynchronous collaboration in CMD of the university students and teachers in Finland, Greece, Italy, and the Netherlands, point out that computer-mediated collaboration enhances learning. The researchers have applied the case study methodology to the investigation of the nature of the students' and teachers' computer-mediated discourse. They claim that their findings show

considerable differences in the students' perception of the nature of the discourse and difficulties in CMC.

A number of scholars have noticed a positive effect of learners' participation in asynchronous discussions on the development of intercultural and pragmatic competences. Belz and Vyatkina (2005), for example, report on a corpus-based research of the use of German modal particles in asynchronous CMAD. They have analysed the development of intercultural communicative in the process of CMC between native and non-native speakers. Morand and Ocker (2003) have revealed specific politeness strategies in asynchronous CMAD.

The issues of the influence of CMAD of the language learner identity have attracted the attention of researchers. Spiliotopoulos and Carey (2005) have investigated the relationship between language and identity of language learners who were using CMD (WebCT) to improve their academic writing skills in English in Canada. The researchers claim that monologic and dialogic writing tasks that are applied in asynchronous discussion forums in electronic courses have important implications for student participation, motivation, and inter-cultural awareness. Nguyen and Kellogg (2005) have also studied the issues of language learners' identity construction and participation in asynchronous discussions. They applied discourse analysis and ethnographic observations to analyze electronic bulletin board postings. The finding suggest that asynchronous discussions increase participation and personal disclosure of the participants, thus increasing the amount of written language produced by the learners in the classroom.

Surprisingly, the linguistic characteristics of the texts of messages posted in asynchronous discussion forums have not recently attracted much attention of researchers, although discussion forums as asynchronous CMAD has been studied by many scholars from psychological, pedagogical and sociological perspectives. Herring (1998) has found that the messages posted on professional discussion lists contained few grammatical errors and demonstrated syntactic complexity and lexical sophistication in comparison with synchronous CMD. Davis and Thiede (2001: 87) write about style shifting in discussion forum messages. The researchers have examined "what happens to different features of discourse when EFL learners must move to function in an ESL situation." They concentrate on lexical and syntactic indications of style shift, which they call 'stylistic emulation,' in the corpus of electronic discourse they have compiled in one United States' university. Davis and Thiede have designed special measures – syntactic complexity and lexical density – to compare the linguistic production of language users participating in synchronous discussions in order to reveal the process of students' integration into academic writing community. Their findings

show that the participants change the level of syntactic complexity according to the communication purpose and audience (e.g. for non-native speakers). In terms of lexical density, i.e. the ratio of lexical words to the total number of words, the researchers report that it has been dropping considerably while the participants were engaged in communication, which indicates the increase of functional words over lexical. Thus, the researchers conclude that there have been style shifts each time the communicative function has changed. Similar shifts in register over time have been registered by Cigankova (2004).

No studies have been found that investigate a wide range of linguistic features in asynchronous discussion forum messages for language learning and teaching purposes.

Academic Weblogs

One more interactional type of CMAD that is investigated in the present research is represented by academic *weblogs* (also known as *blogs*). This type of CMD has been known under this name from 1997 (Herring et al., 2004: 1). Weblogs originated as personal web journals with links to the author's favourite websites. They have been described in the literature as "frequently modified web pages in which dated entries are listed in reverse chronological sequence" (Herring and Paolillo, 2006: 1). Recently, however, weblogs have developed into a rather interactive type of CMD, since it became technologically possible for readers to post their comments and questions in weblogs. Hence, many weblogs have turned from entirely monologues of their owners into dialogues between the weblog owner and the readers, sometimes becoming polylogues among the weblog readers geographically scattered all over the world. Thus, as many other types of Internet discourse, weblogs are not geographically or institutionally confined and are a product of cooperative work (Huffaker, 2004).

The number of weblogs on the Internet has increased dramatically recently, which inspires linguists to investigate language use in them. However, as linguistic study of weblogs is a rather new area of research, there have been relatively few linguistic studies devoted to the investigation of language use in weblogs comparing with other types of CMD. From a linguistic perspective, weblogs have been investigated by Herring, et al. (2004), Miller and Shepherd (2004), Herring and Paolillo (2006), Stuart (2006), Yus (2007).

Nilsson (2004) uses the term *blogspeak* to refer to a variety of language that is commonly used in weblogs, thus emphasizing specific linguistic characteristics of weblogs distinguishing them from other types of language on the Internet (called *Netspeak* by Crystal, 2001):

"Blogspeak, however, is somewhat different than Netspeak in that it combines writing and speech in a unique environment, one that supports both the written internal monologue and threads of conversation" (Nilsson, 2004).

Although communication in weblogs is realised through writing (keyboard typing), Nilsson finds the qualities of both written and spoken communication mode in them. On the one hand, weblogs reveal the features attributed to writing, as contributors have enough time for constructing and revising their texts. On the other hand, they often are as spontaneous as spoken interchanges. In addition, similarly to participants in spoken communication, bloggers make situational reference explicit using hyperlinks. This makes texts in weblogs interactive. Moreover, they combine the monologue and the dialogue, in many places resembling spoken communication. Nilsson also finds that weblogs contain many linguistic features characteristic to spoken language: sentence fragments, informal expressions and dialectal forms. Bloggers use specific to the electronic medium features to create an illusion of speech, e.g. emoticons, italics and unconventional punctuation.

Recently, there has been a discussion in the literature on whether the generic characteristics of weblogs make them a new conceived genre or a genre in the process of formation (Herring and Paolillo, 2006; Stuart, 2006; Yus, 2007). Having investigated linguistic and visual features typical to weblogs, e.g. organisation of the text on the computer screen, the number of columns, background colour, division into sections, hyperlinks etc., Yus draws a conclusion that weblogs are a hybrid genre possessing "a procedural quality" (ibid:119), i.e. they possess specific features helping readers to identify webpages as weblogs. In fact, Yus argues that weblogs are easily distinguished by readers from other types of web pages, even if they "do not seem to differ substantially from other web pages" (Yus, 2007: 118). Yus asserts that the process of formation of weblogs into a new genre is nearly complete. He reminds that early weblogs did not differ much from other web pages. In contrast, contemporary weblogs are easily distinguishable because "weblog templates have now generated a highly identifiable and specific layout which clearly stands out from other competing discourses" (ibid: 138). Similar view on weblogs is expressed in the work of Herring and Paolillo (2006).

However, regarding weblogs as a new genre does not seem reasonable. Firstly, the peculiarities of weblogs enumerated by Yus should be attributed to the specific characteristic of the programme that is usually used for creating weblogs. They are not functionally necessary, i.e. purposefully created by writers tailoring their texts for a specific audience with a specific purpose, which has been identified as the cause of generic differences between texts (Halliday, 1978; Bhatia, 1993). Secondly, weblogs cannot be seen as one genre, as there are

many different types of them. Oberlander and Nowson (2006), for example, distinguish the following four types of weblogs: *news* or *filters*, *commentary* or *knowledge logs* (*k-logs*), (personal) *journal*, and weblogs in academia, i.e. among scholars engaged in high education and research.

It seems reasonable, however, to regard weblogs as a type of CMD associated with a specific technology, i.e. computer programme used for creating and maintaining them. This programme enables readers to converse with the weblog owner by leaving their comments on the webpage. The process of functional language use that emerges in such situational context is the type of CMD known as weblog. Moreover, depending of the purpose and the specific audience of each particular weblog, they may be further divided into genres and sub-genres, e.g. business weblogs, political weblogs, academic weblogs created by university professors to promote their ideas etc.

Academic weblogs have been studied by Rittenbruch, Mansfield and Cole (2003), Efimova and Fiedler (2004). Stuart (2006) presents an overview of academic weblogs, providing a long list of generically different weblogs investigated by the researcher.

Gender issues in weblogs have been studied by Herring et al. (2004), Huffaker and Calvert (2005), Herring and Paolillo (2006), and Oberlander and Nowson (2006). The researchers report that both genders equally participate in authoring weblogs; however, Herring et al. have found that females more often and in greater volume write in personal weblogs. This makes weblogs a useful language teaching tool, especially for shy and not confident learners who are usually dominated by more active students in spoken communication.

1.1.4.3.2 Transactional CMAD Types

Transactional CMAD is a computer-mediated process of academic language use in which language serves a transactional function. Brown and Yule state that the transactional function of language is performed by expressing 'content', i.e. conveying factual information. According to Brown and Yule (1983: 2), "In primarily transactional language we assume that what the speaker (or writer) has primarily in mind is the efficient transference of information. Language used in such a situation is primarily 'message oriented." Thus, transactional CMAD is informational in nature and reader or listener oriented, i.e. proper understanding of the message by the information recipient is crucially important. In English academic tradition, it is the writer who is responsible for the interpretation of the text by the reader. Therefore, writers are supposed to take all possible measures to prevent misunderstanding or ambiguity.

English transactional academic language is rather context independent and characterised by precision and clarity of expression (ibid.).

Transactional CMAD is represented in the present study by academic hypertexts (written) and academic seminars (spoken discourse).

Written transactional CMAD

Academic hypertext

During two decades of the study of hypertext, many researchers have approached the task from a variety of different perspectives: linguistics and semiotics, applied linguistics, communication and cultural studies, pedagogy and computer science (Marchionini 1988; Nielsen, 1990; Bolter, 1991, 1998; Landow, 1992; Charney, 1994; Hackbarth, 1996; Joyce 1997; Lewis and Jansen, 1997; Burbules, 1998; Snyder, 1998; 2002; Huang, 2002; Nelson, 2006). The analysis of their attempts to define a hypertext reveals that, notwithstanding the disciplinary differences in their approaches and the use of terminology specific to each discipline, most of the scholars and scientists refer the term *hypertext* to a computer-mediated written text with some coloured or underlined words, phrases or visual elements connected (*hyperlinked*) to other texts or the parts of the same text. Hypertext is defined as an electronic system of "non-linear organized and accessed screens of text and static diagrams, pictures, and tables" (Hackbarth, 1996: 229), in which "the user is allowed to determine the activities of the system" (Lewis and Jansen, 1997: 5). For example, references to academic publications in the body of an academic article or in the bibliography section may be connected to the original articles published in the same or another academic journal. The reader may quickly switch from one text to another by activating the hyperlinked part of the text.

The early studies of the significance of hypertext for academic writing and publishing (e.g. Balestri, 1988; Moulthrop, 1991; Joyce, 1995; Snyder, 1998) already notice that it has a high potential of changing the nature of academic discourse and academic writing practices. Most of the researchers emphasise two the most important features of hypertext that make it qualitatively different from traditional printed text: its *non-linearity* and *reader control* over the reading sequence (e.g. Nielsen, 1990). They consider hypertext a technology that might offer multiple authorships, make no precise distinction between the author and the reader, provide different reading paths, and expand works with links to other works and media, offering some new ways of writing and interacting for authors and new ways for readers to

read. As far as in 1998, Beavers claimed that hypertext had already become a very well-established means of communication (Beavers, 1998).

Recent technological development has further broadened the range of discussed issues (Snider, 2002). For example, in contrast with earlier hypertext systems, modern software allows the writer to link any character, phrase or any part of text, visual element, image or table with any other one, irrespectively whether it is in the same text or in any other text or place on the Web, whether it has a single element on the page or the whole frame as its destination, whether the destination is a text, video or audio file. This results in an unlimited variety in which this facility may be used by academic writers. However, this also raises the discussions on hyperlink obsolescence in scholarly on-line journals (Ho, 2005), broken hyperlinks in educational texts (Markwell and Brooks, 2002; Greenhill and Fletcher, 2003) and on the approaches to *electronic referencing* and scholarly *citation of Internet sources* (see the overview in Lester, 1997; also Hyland, 2000).

Although the above mentioned questions refer to the domain of academic writing, they have not yet attracted as much attention from the side of researchers as, for instance, the research into the influence of electronic environment on writing and the process of acquisition of written academic language (e.g. in Broady, 2000; Herring, 2001; Warschauer and Kern, 2001; Wysocki, 2004; Wysocki et al., 2004; to name just a few). Such on-line academic journals as, for example, *Computers and Composition, Kairos,* etc. regularly publish articles on this issue; however, the object of the research discussed in the articles is the hypertext created by academic writers in their native tongue (e.g. in the discipline of Rhetoric and Composition). The needs of English for Academic Purposes (EAP) learners, for whom English is a foreign language, in at least some changes in academic writing programmes, corresponding to the demands set by new electronic means of communication, are still to be addressed. Computers are widely used in the teaching and learning; nevertheless, academic texts as a final product, in many cases, are still supposed to be presented in a traditional, paper form. Thus, one mode of communication is used to teach written communication via another, different mode; one medium – paper – is used instead of another – computer screen (Kress and van Leeuwen, 2001; Norris, 2004). This is due to a still widely shared belief that the electronic version of an academic text is exactly the same text sent to a computer screen without any significant changes. Computer-mediated academic discourse is supposed to remain rather conservative and not considerably changing under the influence of the new medium (Crystal, 2001).

However, the empirical data recently obtained in the study of thirty academic hypertexts (Cigankova, 2008) give the evidence that new conventions of academic writing are being

developed in the emerging digital academic culture as a result of intensive exploration of new facilities of the electronic medium by creative academics and increasing competition for the reader's attention in the contemporary academic world. For example, there are certain similarities in the use of linguistic and extralinguistic devices specific to the electronic medium in the texts of academic articles published in on-line academic journals. Moreover, the amount of the use of specific on-line cohesive devices, e.g. hyperlinked references, in academic articles makes qualitative difference in the way those texts are read. There is a tendency to introduce more visual elements, colour, audio and video data, and quite extensive text files to provide additional factual support to the claims made by the authors of the articles. The present study continues the exploration of academic hypertexts in terms of the writers' choice and discourse functions of linguistic features in this type of written CMAD.

Spoken Transactional CMAD

Academic lectures and seminars

This type of spoken transactional CMAD has become so increasingly popular in academia, especially in distance education, that the word *webinar* invented to denote it has been introduced by Merriam Webster Dictionary as the word of the year 2008. An on-line environment for oral interaction on-line at such computer-mediated seminars and workshops may be created by various types of computer software for synchronous audio-graphic and video conferencing (e.g. HorizonLine Wimba, Lyceum, Interwise etc.). It enables PowerPoint presentation, video films, audio files, photographs, figures and tables to be presented simultaneously with the lecturer's speech. Moreover, listeners may actively participate in the event by asking questions or giving comments in at least two ways. The first way is by typing their messages in the synchronous CMC area ('chat rooms'), for example, in WebCT organized on-line seminars. Another way is by using means of vocal communication provided by the software: each participant may have his or her own communication channel. In the latter case they need to request and receive permission to speak from the seminar organisers each time they want to make a comment or ask a question. The studies of on-line lectures and seminars have tackled the issues of their effectiveness for teaching the content in distance education (e.g. Hampel and Hauck 2003). The use of audio-graphic conferencing for language teaching has been researched by Hampel, 2003; Hassan, Hauger, Nye and Smith (2005) and Rosell-Aguilar (2005). No studies have been found that would investigate the linguistic characteristics of on-line lectures/seminars for international audience and compared them with the characteristics of other types of CMAD.

The author of the present dissertation hypothesises that the process of functional language use in academic context varies across the described above types of CMAD, and this variation is medium-specific. The following section is devoted to the concept of linguistic variation, the factors causing it, its types and approaches to its linguistic investigation.

1.2 LINGUISTIC VARIATION IN COMPUTER-MEDIATED ACADEMIC DISCOURSE

Linguistic variation has always been one of the main concerns of linguistic research, especially in such applied linguistic sub-fields, overlapping to some extent, as sociolinguistics and linguistic pragmatics. These disciplines have conceived the concept of linguistic variation and developed a variety of empirical research approaches that provide insights into the factors influencing language use in society while the theoretical mechanisms causing linguistic variation have been elaborated in systemic-functional linguistics.

1.2.1 Conceptual Definition of Linguistic Variation in CMAD

The term *variation* originated from mathematical statistics where it means a measure of the deviation of the value from a central tendency. In linguistics, the term *variation* was initially used in the descriptive approaches that "presume systematically ordered heterogeneity of natural languages" (Bussmann, 1996) to mean a deviation from a standard variety of language. Nowadays, the existence of such a variety of language is questioned in many languages. For example, in the case of English, many regional language varieties are contesting for the leading role: British English, American English etc. (e.g. 18 varieties of English are listed as language choice options in the latest version of Microsoft Word programme.). Therefore, the use of the term has changed recently. The uses of the term in different contemporary linguistic subfields are sometimes homonymic. For example, in language development studies, *lexical variation* is a synonym of *lexical diversity*, which Read (2000) distinguishes as one of four components of *lexical richness* that is measured by the type-token ratio (Malvern et al., 2004).

In sociolinguistics, *linguistic variation* is defined as "[t]he focus of sociolinguistics, charting how language varies, and matching variation in language to social contexts and social group membership" (Jackson, 2007: 91). Thus, there has been a focus shift to a view of linguistic variation as a natural linguistic phenomenon, not a deficiency caused by improper use of some ideal standard language variety. This definition, however, reveals that sociolinguistics is interested primarily in social factors. However, other situational factors, such as for example mode of communication (e.g. use of computer), are also important. Therefore, they are in the focus of the present study, which investigates linguistic variation in computer-mediated academic discourse.

Discourse is seen in the present research as a process of language use (see the definition of discourse in 1.1) which is realised and empirically analysed in its discourse/ linguistic realisations – texts (Brown and Yule, 1983). Linguistic *variation* in discourse (as a process) is, logically, also such *a process* in which language use is varied by language users to suit different situational contexts. It is *realised* in quantifiable changes in the frequency of (co)occurrence of linguistic features in different texts. These realisations of linguistic variation as a process are the differences in linguistic characteristics, i.e. different frequency of use of linguistic features that appear within texts, for example, frequency of personal pronouns, passive voice constructions etc. They are quantifiable and, thus, can be empirically analysed.

The occurrence of differences in the frequency of use of linguistic features among different text types is caused by different factors discussed in the following section.

1.2.2 Causes and Mechanisms of Linguistic Variation in Discourse

Linguistic variation in discourse depends on the context of language use. It is caused by factors that have different nature and operate in texts (as discourse realisations) at different levels: phonological, morphological, syntactic, lexical and discourse/pragmatic. The factors on which linguistic variation depends fall into two major groups: the factors operating within the text and extratextual factors (Biber and Conrad, 2001).

The first group of factors, linguistic, perform at the level of text as discourse realisation. At this level – the level of immediate context (*co-text*), the form of the linguistic features in the text depends on the functions they perform and on the form of other linguistic features in the text (Halliday and Matthiessen, 2004). For example, the use of a plural noun in English determines, in most cases, the plural form of the verb. The analysis of the context at this level is the investigation of lexical, grammatical and rhetorical features in the text.

The elaborated theoretical explanation of the mechanism of linguistic variation caused by linguistic factors has been modelled in-detail in *systemic-functional linguistics* (SFL). Language in SFL is seen as "a system for making meanings: a semantic system with other systems for encoding the meanings it produces" (ibid: xvii). Such a broad understanding of language has led to broader understanding of what can be regarded as linguistic resources for making meaning and the factors determining the choice of semiotic means, e.g. linguistic features in a text.

The factors acting in a communicative situation are discriminated as the factors associated with the *ideas* the text is aimed to convey, *interactants* (communicating humans)

and the *physical environment* of the communicative event/situation. Hence, the process of making meaning as a part of the process of language use in context (i.e. discourse) is a three-fold process – *semiotic*, *social* and *material*. Correspondingly, language functions in texts simultaneously as *message*, as *exchange* and as *representation* of the phenomena in the surrounding world. Hence, language used in the process of communication has three major functions – *metafunctions* in SFL. For example, the function with which language conveys the message to make meaning (i.e. semiosis) is an *ideational* meta-function. Meaning making as a social process denotes that language functions in the society as an exchange, hence another Hallidayan metafunction – *interpersonal.* In contrast with ideational function seen as 'language as reflection' (as message), this metafunction is 'language as action' (as exchange), both 'interactive and personal' (ibid: 30). This is because, apart from construction of reality, language in systemic-functional linguistics is considered as always "enacting our personal and social relationships with the other people around us" (ibid: 29). One more 'mode of meaning' – the third component – 'enabling or facilitating function' of language as representation is relevant to the construction of text and allows it to have texture in real situational context. The *textual* metafunction enables other metafunctions of language, which "depend on being able to build up sequences of discourse, organizing the discursive flow and creating cohesion and continuity as it moves along" (ibid: 30).

Table 1.1

Language Functions and Factors Causing Linguistic Variation

	Language metafunctions		
	ideational	**interpersonal**	**textual**
Phenomena in the material world	ideas	interactants	physical environment of a communicative situation
Process of language use	semiotic	social	material
Language functions as	message	exchange	representation
Factors	field	tenor	mode

At the level of discourse realisation (text), three factors influence language use – *field, tenor* and *mode* – each associated with the corresponding language metafunction. The correspondence of language metafunctions and the factors causing linguistic variation in discourse as viewed in Systemic Functional Linguistics has been summarised by the author of the present text in Table 1.1.

A combination of these three aspects causes linguistic variation at the intratextual level (i.e. register variation) in each particular situational context (Martin, 2001).

According to Halliday and Matthiessen (2004), the choice of linguistic means in a particular text depends on the function of the language in the text. Thus the functions of

language are the linguistic factors that cause linguistic variation. For example, the choice of linguistic features used by a language user depends on whether the text is meant to inform or to persuade the reader, to maintain social relationship or to express disagreement or complaint. Moreover, the choice of linguistic means to even a greater extent depends on which linguistic features have already been used in that text. To illustrate, the choice of formal style of wording for a complaint would presuppose the use of formal lexis and syntactic structures and avoidance of contractions, biased and slang words.

The second group of factors influencing language use, extralinguistic, are the factors operating at the level of extratextual communicative situation (Hymes, 1972). Phonological, morphological, syntactic, lexical and pragmatic aspects of language use vary with regards to extralinguistic factors. To study them means to analyse the context in general (Biber and Conrad, 2001), which is hardly possible without further division of them into smaller analysable groups. The effect of different aspects of society on language use by various social groups is studied by sociolinguistics and especially by its subfield – *variation linguistics* (also known as *variationist* or *variational sociolinguistics*). The main concern in sociolinguistics is the investigation of linguistic variation that is caused by the social variables in language use in social context, characterising language users in terms of their age, gender, ethnicity, religion, level of education, geographical location, socio-economic status, power relations, etc. (Druviete, 2008). A quantitative study of linguistic variation caused by extralinguistic factors was started by Labov (1966) who proposed that the regularities of linguistic variation are quantitatively determinable. Hence, statistical methods are used to investigate linguistic variation. A special sub-field of sociolinguistics - *variation linguistics* is concerned with the description of linguistic variation and "the problems of the origin and quantification of linguistic varieties in relation to extralinguistic factors" (Bussmann, 1996) using statistical research methods.

The present research seeks to reveal linguistic variation between CMAD types caused by the use of different computer software for communication, i.e. extratextual factors; however, intratextual (i.e. linguistic) factors are also investigated to reveal the specific characteristics of each CMAD type. The research concentrates on the investigation of co-occurrence – *complimentary distribution* (Biber, 1988) – of linguistic features in different CMAD text types caused by the specificity of the *material means* through which the type of discourse is realised, i.e. by the type of software – computer application – that is used for communication. Hence, the research methodology of variation linguistics seems to be applicable to CMAD as a research subject. Nevertheless, this has not been done yet, as variation linguists primarily focus on the influence of society on language use and the study

of social factors. Preston (1993), however, insists that variation linguistics may study also linguistic factors. Moreover, having reviewed a large number of empirical studies, he has found that linguistic factors are far more powerful influences on language use than social factors. Nevertheless, the research in variation sociolinguistics has focused on social factors, in particular, on language users – interactants. Such extralinguistic factors as the use of computer software to mediate the process of communication and the intratextual factors causing linguistic variation in computer-mediated texts have attracted much less attention of sociolinguists than the study of social factors of language use.

1.2.3 Types of Linguistic Variation in Computer-Mediated Academic Discourse

Linguistic variation in CMAD is the difference between the functional use of linguistic features in different CMAD types. Scholars studying academic discourse base their classifications of various types of it on different levels of discourse variation: register level (Biber, 1988), disciplinary discourse level (Hyland, 2000), generic level (Swales, 1990, Bhatia, 1993). Register is a more general level of variation in discourse than generic and disciplinary variation. It is distributed across disciplines and genres. Register variation is determined by specific internal properties of written and spoken texts that distinguish them from non-academic texts. Internal properties of academic texts vary in different situational contexts due to the difference in functional use of language. The categories of functional use of language that cause register variation, which are the realisations of three metafunctions of language (see section 1.2.2.) are determined by Halliday as *field*, *tenor* and *mode*.

The category *field* (the realisation of *ideational* metafunction of language) is associated in the present research with the field of knowledge – the academic discipline that is responsible for disciplinary variation in discourse. In the present study, academic discourse is represented by the variety restricted to the academic register in one discipline – education – to avoid disciplinary variation in the use of linguistic means.

The second category – *tenor* – is associated with *interpersonal* metafunction of language, i.e. social relationship between interlocutors and the purpose of language use. It is known as the cause of generic variation in discourse. Genre variation is minimised in the present study by delimiting the intended audience and the purpose of writing to only academic ones. That is, the participants in CMAD are academic professionals – the members of *European Association of Teachers of Academic Writing* – people for whom excellent use of

academic language is an attribute of their profession. Moreover, only the texts produced for academic purposes had been selected for the analysis in the present research.

The third category – *mode* – is the realisation of *textual* metafunction that is responsible for the variation across text types associated with the mode of communication. Since *field* and *tenor* variables have been isolated in the present research, the variation in the CMAD types is assumed to be the result of the difference in the *mode*, i.e. a material realisation of discourse.

Thus, linguistic variation in text types is attributed to the type of CMAD that they represent, which is associated with a unique combination of the type of discourse, the mode of CMC, and the type of software (computer programme) used.

1.3 ANALYTICAL REVIEW OF EMPIRICAL STUDIES

The present research has solving language-related problems concerning effective use of language in computer-mediated academic discourse community as its ultimate goal. That is why, it falls into the field of applied linguistics. Positioning the study of CMAD within applied linguistics' disciplinary context, however, is hardened by its complex nature and the multiplicity of contextual factors that may influence the use of language in electronic settings. Nevertheless, CMAD is classified in this study as a specialisation within a broader area of *computer-mediated discourse* (CMD). The study of CMD, in its turn, falls into the more general field in applied linguistics – *computer-mediated communication* (CMC), which deals, apart from language, with other semiotic modes, e.g. non-verbal communication (Barnes, 2003). The study of language use in CMC is in itself a large and quickly developing area in applied linguistics that welcomes theoretical insights not only from linguistics but also from other disciplines. Research into CMD, however, differs from that into CMC in that it concentrates primarily on the study of verbal language above the sentence level, regarded in the electronic context of use (i.e. discourse).

The description and analysis of CMAD types presented in this study from a linguistics' perspective is based on current paradigm in applied linguistics and applies contemporary research methodology. It aims to obtain the results in quantitative corpus linguistics research into the use of formal features of language use in CMAD. The choice of the research method for the present study was based on the analysis of the literature on empirical research into CMD in linguistics. The conducted review of the literature on the empirical research into the process of language use in electronic contexts reveals that the research approaches applied by scholars to study CMD exploited a range of quantitative and qualitative research methods.

A number of corpus-based studies of register variation in CMD have been undertaken at those early days of the field when computers were used for communication in only a few universities and when there was not much variety in CMD types (Baron, 1984, Collot and Belmore, 1993, Ferrara, Brunner and Whittemore, 1991, Yates, 1996). For example, Baron (ibid.) studied CMD from a linguistics point of view, first raising the issue of the influence of CMC as a factor in language change. She studied CMD as a new register that had an impact on traditional spoken and written genres. Baron was the first who suggested that there might be the differences in language structure between CMD and other forms of language.

Nevertheless, the wide discussion on the status of the language variation in CMC started only after the publication of Ferrara, Bruner, and Whittemore's *Interactive written discourse as an emergent register* in 1991. As Herring reports,

> The immediately following years saw the rise of a wave of CMD researchers, working independently on what has since emerged as a more or less coherent agenda: the empirical description of computer-mediated language and varieties of computer-mediated discourse.
>
> Herring (2001: 626)

Early researchers, however, oversimplified CMD, for instance Ferrara et al. (1991), who regarded "interactive written discourse" as a single genre (as cited in Herring, 2001: 613). Most of the messages of that period were exchanged in computer-mediated communication in academic settings, as the access to computers connected into what later developed into a global network was mainly in universities. This and the excitement of the researchers observing the changes emerging in language use in CMC may explain the enthusiasm with which the birth of 'a new variety of language' was proclaimed and why the changes were attributed entirely to technology – it was difficult to notice other (e.g. linguistic, intratextual) factors that may have also caused the changes.

Meanwhile, a research methodology for studying those changes has emerged (Biber, 1988). The assumption underlying the methodology is that linguistic features do not randomly co-occur in textual realisations of discourse. If persistent co-occurrence of some linguistic features is observed in a group of texts, it is reasonable to suppose that there is an underlying functional relationship between the features that makes them co-occur. Thus the patterns of co-occurrence mark underlying functional dimensions. On the basis of the co-occurrence of linguistic features in different text types Biber (1988) has developed a method of multidimensional factor analysis that groups the linguistic features in a text into a limited number of factors (seven in Biber, 1988) – functional dimensions – according to the functions they perform in the texts. He selected 59 linguistic features, but reduced the number of variables to a small set of factors to find out the co-occurring linguistic features. Biber has applied this methodology also in his other studies of academic discourse, e.g. to the analysis of spoken and written academic discourse in American universities (Biber et al., 2004).

Many other researchers have applied Biber's methodology to investigate linguistic characteristics of popular and academic texts (Conrad, 2001; Carkin, 2001; Gries, 2003). Biber and Kurjian (2007), for example, have applied the multidimensional analytic method to the analysis of text categorisation in Google searchers on the Web and suggested some changes and improvements in the taxonomy of text categories. The corpus in their study was constructed by a stratified sample of typical web pages on two Google categories "Home" and "Science", i.e. popular, non-academic and rather informal texts. The researchers have found salient linguistic and functional differences in text types across these two categories. However, texts that would represent computer-mediated academic discourse have not been included in their research.

Other researchers have also used Biber's method to study CMD. Having applied a multidimensional analysis research methodology to the study of the corpus of computer-mediated language, Collot and Belmore made a distinction between the messages composed at the moment of communication and pre-written, carefully thought over before being sent messages. The different ways in which the messages are produced are now known as *synchronous* and *asynchronous* modes of CMC correspondingly. The researchers argue that the situational constraints by which the 'electronic language' is characterised make it different from other varieties of English, the main difference being that "electronic language displays some of the linguistic features which have been associated with certain forms of written language and others which are more usually associated with spoken language" (Collot and Belmore, 1993: 48).

In fact, later studies have revealed that computer-mediated discourse is "sensitive to a variety of technical and situational factors" (Herring, 2001: 613), e.g. the recent studies in sociolinguistics and discourse analysis (Androutsopoulos, 2006; Herring and Paolillo, 2006). Herring (2001), for example, proposes a qualitative approach to computer-mediated discourse analysis (CMDA), which is based on a well-known in social sciences qualitative method of *content analysis* (Huckin, 2004). The method involves coding the linguistic features in a corpus of computer-mediated messages into social categories and then counting the frequency of specific features in each category. This method has recently been applied by Bretag (2006), who investigated 276 e-mails with the purpose to identify the discourse features revealing the relationship of teacher and students in CMC.

A complex interconnection of linguistic and extralinguistic factors that may cause variation in computer-mediated discourse make researchers look for integrated – quantitative and qualitative – research methods. The reason for synthesising qualitative and quantitative methods is based on the assumption that linguistic and extralinguistic factors should be studied in their interconnection. Neither the properties of the language system can be separated from 'the uses,' nor can the study of discourse as text be based only on non-linguistic conventions (as, for example, in critical discourse analysis). CMD, as language in use, should be studied in the situational context, which implies that the frequency of occurrence of formal linguistic features alone cannot ensure a satisfactory quality of the analysis. It is necessary to relate the analysis of discourse, in terms of Halliday (1985: xvii), "to the non-linguistic universe of its situational and cultural environment." For example, Baron **(1998: 150) states that her survey of previous attempts to investigate social and formally linguistic properties of CMD suggests** that "**it is time to attempt an integrated analysis." Counting discourse features for** content analysis is often combined with generic

and pragmatic qualitative methods of analysis (Herring, 2001; Bazerman and Prior, 2004). Nevertheless, only a few studies of CMAD have integrated qualitative or quantitative approaches (e.g. Belz and Vyatkina, 2005; Jeong, 2005; Chen, 2006; Vandergriff, 2006). A sound methodology for synthesising the approaches, however, has not yet been developed.

In linguistics, where the ultimate goal is to study language, not the society, as the research subject, in contrast with social sciences, quantitative methods of researching linguistic variation in discourse have prevailed. In general, contemporary quantitative approaches to the study of linguistic variation in discourse utilise corpus linguistics principles of corpus compiling and analysing (Meyer, 2002; McEnery et al., 2006), make use of modern computer software for language analysis (e.g. Nation and Coxhead, 2002) and apply statistical procedures to reveal the linguistic variation. Malvern et al. (2004) provide the justification for quantitative measurement of linguistic features in written language, explaining that there are "correspondences between meanings and forms" and "written language comprehension takes place through the analysis of linguistic form" (ibid.). Therefore, the author of the present paper considers that quantitative analysis of formal linguistic features can supply valuable information about the language used in CMAD. In addition, the author assumes that it is necessary to contextualise the instances of language by analysing their discourse functions (Halliday and Matthiessen, 2004) in the situations of their use.

Quantitative computerised methods of language analysis have been conducted by, among others, Laufer and Nation (1995), Rissanen et al. (1997), Jarvis (2002), Rissanen (2007). In Latvia, computerised analysis has been applied to the study of Latvian by J.Borzovs, G.Fricnovičs, A. Spektors (1997), Spektors (2001), Bārzdiņš (2006), Skadiņa (2008). A statistical analysis of language, however, has not been applied since the work of Kļaviņa (1980). A large-scale quantitative research into six types of CMAD had not been conducted before the present study, neither in Latvia, nor worldwide. Although there had been a considerable number of studies on the language of the Internet in general and academic register in particular types of computer-mediated discourse, linguistic variation across the text types in CMAD had not yet been investigated. Most of the previous studies into CMD had been, in terminology of Biber (1988), "microscopic", i.e. narrowly tailored to investigate the linguistic properties of specific CMD types or had had a very narrow scope of investigation of only a few features. Although they had supplied the empirical data on each particular type of computer-mediated discourse, they did not provide a comparison of linguistic characteristics across different types of CMAD. A "macroscopic" approach had been applied by Biber et al. (2004) to the study of language use in academia. However, that quantitative analysis had investigated traditional academic discourse, i.e. the use of language in typical academic

situations (lectures, seminars, tutorials, interaction with administrative staff) in universities in the United States. The computer-mediated varieties of academic language, judging from available academic sources, had not yet been sufficiently studied, especially from the applied linguistics' perspective.

Summary

The author has distinguished and theoretically conceptualised English computer-mediated academic discourse (CMAD) as a new specific type of English discourse that takes place in academic settings and is mediated by computer. The concepts of *academic discourse* and *computer-mediated discourse* have been investigated as constituent parts of the concept of CMAD. An original definition of CMAD has been elaborated as the result of thorough theoretical research. The term computer-mediated academic discourse has been compared and discriminated from other contesting terms, e.g. *electronic discourse, interactive networking, interactive written discourse, electronic language, electronic networked discourse.* The proposed definition has been used for further research into CMAD. On the basis of the conducted analysis of contextual factors influencing CMAD as a process, an original typological classification of CMAD has been proposed. It theoretically models how different combinations of factors result in different types of CMAD. The following six the most typical types of CMAD have been distinguished: academic e-mails, synchronous conferencing, on-line discussions, weblogs, hypertexts and computer-mediated seminars. Each type of CMAD is the result of a unique combination of transactional or interactional type of discourse, synchronous or asynchronous mode of interaction, spoken or written mode of discourse and the type of software used in communication (e.g. e-mail, discussion forum, weblog etc.). The author has provided a comprehensive linguistic description of the characteristic features of CMAD types summarising the knowledge accumulated in the studied literatures.

The notion of linguistic variation in discourse, its types and the factors influencing it have been presented. The linguistic variation in computer-mediated academic discourse has been defined as a process in which communicators vary the use of academic language to match the specific situational context in computer mediated communication. It is realized, as a product, in quantifiable differences in the frequency of occurrence of linguistic features in CMAD text types, for example, the frequency of personal pronouns, passive voice constructions, etc.

An overview of recent empirical research approaches to the study of linguistic variation in CMAD and the rationales for the present research are provided. The review of the literature has revealed that the problem had not yet been sufficiently studied in applied linguistics.

There had been no linguistic studies that would have compared different text types in CMAD in order to apply the knowledge to teaching and learning academic language at university. This was identified as the gap in knowledge that the present research aimed to cover.

The empirical research revealing the linguistic variation in CMAD is presented in Part2.

2. LINGUISTIC VARIATION IN COMPUTER-MEDIATED ACADEMIC DISCOURSE EMPIRICAL RESEARCH

The second part of the dissertation is devoted to the description of the research method and research procedure applied in the present study. Further, the results of the research, the discussion on the findings and drawn conclusions are presented.

2.1 RESEARCH METHODOLOGY AND DESIGN

As has been demonstrated in Part 1 of the present research, the author considers that academic language, being constantly shaped by the use in CMC, may have developed specific patterns of co-occurrence of linguistic features in different types of CMAD. They depend on a variety of combinations of CMC channel (e.g. e-mail, weblog), synchronicity (e.g. synchronous chats, asynchronous discussions) and the mode, from written to spoken through a variety of intermediate forms, characteristic to CMC language ('netspeak' in Crystal, 2001). To reveal the patterns in the use of linguistic features in texts as realisations of CMAD, it is necessary to identify the linguistic features in representative samples of each CMAD type and compute the frequency of their occurrence. Therefore, the author assumes that the measurement of register variation in CMAD would be best conducted by applying quantitative frequency analysis of linguistic features in CMAD texts. To reveal the differences in the patterns of register variation across CMAD types, it is necessary to apply statistical procedures for the analysis of variance that would allow the comparison of different types of CMAD.

2.1.1 Research Methods Applied in the Study

The decision on which quantitative research methodology to apply in the present empirical research was based on two important assumptions. Firstly, the general rule of thumb in science says that for the investigation of new, previously unstudied data, a well-known, validated and approbated research methodology must be applied while newly-developed research methods are to be first validated on well-studied data before they may be applied to the investigation of new data. The research object of the present study is comprised of previously not studied data representing six types of CMAD. Therefore, the decision was made to choose the methodology that had been validated by the time of the present research in previous corpus-based studies of language in use and was recognized and valued in academia (see an overview of the use of corpora to explore linguistic variation in Reppen, Fitzmaurice and Biber, 2002).

Hence, the main quantitative research method applied in this study was Biber's (1988) method of multidimensional analysis of variance of linguistic features. As the conducted literature review presented in 1.3 shows, it is a famous and widely recognised corpus-based method that is considered to be highly effective for revealing register variation across large number of texts in the corpus.

The most important notion in the method of Biber is the concept of textual *dimension*. According to Biber, it is not possible to analyse linguistic variation in discourse along one dichotomous dimension, e.g. speaking/ writing: a multidimensional approach is necessary. The researcher emphasises that linguistic features do not randomly occur in texts. Having used a multivariate analysis statistical method in his 1988 study, Biber identified which linguistic features typically co-occur in different types of texts. He revealed that the linguistic features that serve similar discourse functions tend to appear in similar text types. According to Biber, the strong patterns of co-occurrence of linguistic features in texts signal the existence of underlying functional dimensions. Different groups of co-occurring features constitute different dimensions. Thus, according to Biber (1988), the linguistic dimensions are the continua along which register variation occurs and the types of discourse differ from each other in the English language.

The present study applies Biber's multidimensional view of the variation in discourse. A five-dimensional model has been build, according to which CMAD is investigated, including the following dimensions identified by Biber (1988: 13):

1. *Involved/ Informational production,*
2. *Narrative/ Non-narrative concerns,*
3. *Explicit/ Situation-dependent reference,*
4. *Abstract/ Non-abstract information,*
5. *Overt expression of persuasion*.

This approach allows systematic description of different types of CMAD in the present study and the comparison between them and other types of English academic discourse.

2.1.2 Research Design

The research into variation in CMAD presented in this dissertation is **descriptive**, as it intends to investigate naturally occurring language in its context of use, i.e. discourse. The study is **corpus-based**, as its research object is a corpus of authentic texts processed with text-processing computer tools. The research is also **analytic**, as the focus is on the constituent parts (i.e. linguistic features) of computer-mediated academic texts as they are represented in

the compiled corpus. Finally, the study is **hypothesis-driven** in that it investigates "naturally occurring phenomena … with a preconceived hypothesis" (Seliger and Shohamy, 1989: 117).

The main research question addresses the linguistic variation in the use of linguistic features among the types of CMAD. The search for the answer to the main research question presupposes finding the answers to the following supporting research questions:

- What is the frequency of occurrence of each linguistic feature set as a dependent variable in this study?
- Which linguistic features occur more frequently in which CMAD types?
- What is the mean frequency of each CMAD type on each Biber's (1988) textual dimension?
- Is there a statistically significant difference in the frequency of (co)-occurrence of linguistic features between different CMAD text types?
- What is the difference in the mean frequency scores between CMAD types and other types of discourse?

The author proposes the following research hypothesis: the frequency of use of linguistic features varies in the texts as realisations of different types of CMAD. To put it in statistical terms, there is a significant statistical difference among CMAD types in a number of formal and functional properties of language along each textual dimension.

As language users make different linguistic choices in different types of CMAD, the frequency of use of linguistic features varies in the texts as realisations of different types of CMAD. This variation depends on the functions the linguistic features perform in different situational contexts created by each particular type of computer software, as it has been shown in Part 1. This means that each type of CMAD is characterised by the use of a number of specific linguistic features that are typical to it and a number of linguistic features that are unlikely to occur in it, i.e. the features in a "complimentary distribution" Biber, 1988: 101).

In the following part of the research, the author attempts to measure the linguistic variation in CMAD and apply statistical procedures for quantitative evaluation of the measurements in relation to the proposed hypothesis.

Applied research tasks

To answer the research questions and test the hypothesis, the following research tasks were set:

1. To provide an operational definition of CMAD, i.e. to identify the variables that can reveal possible linguistic variation in it.

2. To compile a corpus of texts representing the academic language produced in different types of CMAD.
3. To measure the frequency of occurrence (per 1 000 words) of 55 linguistic features as dependent variables.
4. To compare the frequency scores of different types of CMAD.
5. To identify the linguistic features that are the most and the least frequent in each type of CMAD.
6. To reduce the initial number of variables (Data Reduction statistical procedure) to a number of analysable functional dimensions of language variation in CMAD.
7. To compute the factor scores for each CMAD type on each dimension.
8. To compute whether there is a significant difference among six types of the CMAD, represented by the compiled corpora, in the number of occurrence of the linguistic features on each functional dimension.
9. To compare the results of the research with previous studies.
10. To identify the discourse functions of the most frequently occurring linguistic features and main characteristics of each type of CMAD by manual qualitative analysis of all the collected samples in the corpora.

2.1.3 Operational Definition of the Research Subject

For the operationalisation of linguistic variation in CMAD as the research subject, it is necessary to turn it to observable and measurable quantities – statistically countable and comparable linguistic features – that would enable data collection.

Analysis of discourse in this study, however, is not restricted to only abstract grammatical categories traditional for formal structural analysis. Modern corpus linguistics methods also allow quantifiable categories of qualitative analysis, e.g. specific lexis, speech act words, discourse markers, hedges to be included in statistical analysis of frequency of their use in different types of discourse. As the present study is mainly concerned with the investigation of linguistic properties of computer-mediated texts as discourse realizations, the study of the process of interaction, the behaviour, gender differences and social relationships of the interactants are beyond the scope of it. Instead, the main focus is placed on the variation in the use of linguistic features in text types produced in the process of communication depending on the use of different types of communication technology. Therefore, such an overwhelming phenomenon as discourse inevitably has to be delimited to analysable units – statistically countable and comparable linguistic features.

Thus, for the quantification of a possible variation in linguistic dimensions across six types of CMAD the frequency of occurrence of the most characteristic linguistic features (Biber 1988) was measured, such as an average length of words and sentences, contractions, type/token ratio, personal pronouns, demonstratives, modals, prepositions etc. (in total, 55 linguistic features). The author used computer processing of the texts in the corpora to identify the features relevant to each dimension and obtain the evidence of the variation among CMAD types along the dimensions that had been influenced by the electronic context of language use.

Dependent and independent variables were set and the developed procedure for quantitative evaluation of the measurements of linguistic variation in relation to the hypothesis. The fluctuations in academic register in text types as discourse realisations were viewed in the present study as depending on the type of CMAD associated with a certain type of technology for CMC. Thus, the type of CMAD was assigned as **independent variable** that causes variation in the use of linguistic features in text types. The frequency values of occurrence of linguistic features in texts were assigned as **dependent variables**.

Thus, the variation is CMAD has been operationalised (i.e. the variables were strictly defined into measurable factors) as a statistically significant difference in the frequency of occurrence of the following fifty-five lexicogrammatical linguistic features in CMAD text types (the list of linguistic features was adapted from Biber (1888), Biber et al. (1999) and Biber et al., 2004):

Pronouns

1. First person pronouns (PRO1) include *I, me, my, mine, myself, we, us, our, ours, ourselves* and all the contracted (short) forms.
2. Second person pronouns (PRO2) include *you, your, yours, yourself, yourselves* and all the contracted forms.
3. Third person pronouns (PRO3) include *he, his, him, himself, she, her, hers, herself* and all the contracted forms, but exclude the pronoun *it*.
4. Pronoun *it* (IT), as it is the most generalized pronoun that can substitute for nouns, phrases or whole clauses.
5. Demonstrative pronouns (DEMP) include *this, that, these, those,* and *that's*; they refer to a noun either inside or outside of the text.
6. Indefinite pronouns (INDP) include *anybody, anyone, anything, everybody, everyone, everything, nobody, none, nothing, nowhere, somebody, someone,* and *something.*

Reduced forms and dispreferred structures

7. Contractions (CONTR) include short forms, e.g. *can't, don't*, etc.

8. Complementizer *that* deletion (THATD), e.g. I *think [0] he went away.*
9. Final (stranded) prepositions (FPR), e.g. *the girl that I was thinking of.*
10. Split auxiliaries (SAUX), e.g. *they were apparently shown to...*

Questions

11. WH questions (WHQ) include only direct questions starting with *what, when, where, why* and *how.*

Lexical specific features

12. Type/token ratio (TTR) is considered to be an important of measure lexical richness, the type/token ratio was calculated by dividing the number of different words by the total number of words in the text sample and then multiplying it by 100. As type/token ratio is text-size dependent and decreases nonlinearly with text length (Baayen, 2001), the number of tokens was standardized.
13. Mean word length (MWL) - the number of characters in a word.
14. Mean syntactic length per sentence/utterance (MSL) – the number of words per sentence.

Nouns

15. Nominalizations (NOM) – the words ending in *–tion, -ment, -ness, -ity*
16. Number of nouns (N), excluding gerunds and nominalizations

Verbs

a) Verb tense

17. Past tense verbs (PTV)
18. Perfect aspect verbs (PAV)
19. Present tense verbs (PRTV)

b) Passives

20. Agentless passives (ALPASS) – a passive construction in which the doer is not mentioned
21. *by* passives (BYPASS) – a passive construction in which the doer is introduced with the preposition *by*

c) Modals

22. Possibility modals (PMOD), e.g. *can, may, might, could*
23. Necessity modals (NMOD), e.g. *ought, must, should*
24. Predictive modals (PRMOD), e.g. *will, would, shall*

d) Semantic categories

25. *be* as the main verb (BE) – the verb *to be* and all its forms as the main verb/predicate of a sentence/utterance.

26. Private verbs (PVERB) – express intellectual states or acts, and are used for the overt expression of private attitudes, thoughts, and emotions; the category includes verbs such as *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,* and *understand.*
27. Public verbs (PUBV) **–** the following speech act verbs and other communication verbs: *add, announce, advise, answer, argue, allege, ask, assert, assure, charge, claim, confide, confess, contend, convey, convince, declare, demand, deny, emphasize, explain, express, forewarn, grant, hear, hint, hold, imply, inform, insist, maintain, mention, mutter, notify, order, persuade, petition, phone, pray, proclaim, promise, propose, protest, reassure, recommend, remark, reply, report, respond, reveal, say, shout, state, stress, suggest, swear, sworn, teach, telephone, tell, urge, vow, warn, whisper, wire, write* etc.
28. Suasive verbs (SUV), e.g. *command* (somebody to do something), *demand (*from somebody to do something) or *insist* (on somebody doing something)

e) Non-finite forms

29. Infinitives (INF) **–** this category includes all the forms of the infinitive

Adjectives

30. Attributive adjectives (ATADJ)
31. Predicative adjectives (PRADJ)

Adverbs and adverbials

1. Semantic categories of adverbs

32. Place adverbials (PLADV) can show position, direction or distance. The following place adverbs appear in the compiled corpus: *across, alongside, around, aside, away, backward, behind, below, beside, far, forward, here, inside, locally, near, nowhere, outside, overseas*, *there,* and their hyphenated forms, e.g. *far-reaching*. Biber et al. have found that only *here* is frequent enough in academic prose, appearing at least 200 times per million, while being much more frequent (over 1000 times per million) in face-to-face conversation (Biber et al., 1999: 552).
33. Time adverbials (TADV). The following adverbials give information about time (position in time, frequency, duration or relationship): *afterwards, again, ago, currently, earlier, early, eventually, ever, formerly, immediately, initially, instantly, late, lately, later, momentarily, never, now, nowadays, once, originally, presently, previously, recently, shortly, simultaneously, so far, soon, still, subsequently, today, tomorrow, tonight, yesterday, yet* (Biber, 1988: 224, and Biber et al., 1999: 552). Biber et al. have also found

that only the following time adverbials appear at least 200 times per million in academic texts: *now, then, again, always, already, sometimes, often, usually* (ibid: 561). In the 1988 study, Biber excluded *last* and *next* from the list, as they may have other major functions in texts, e.g. showing logical relations. Following Biber, in order to make the results comparable with the previous research data, in the present study, *last* and *next* were coded as other adverbs (ADV, 34).

34. Other adverbs/adverbials (ADV) The following groups of adverbs fall into this category:
    - Manner adverbs: *together, significantly, well, along (with).* According to Biber et al. (1999), *together* is not characteristic to academic prose, but *significantly* and *well* appear in it at least 200 times per million (ibid: 561).
    - Degree adverbs other than amplifiers (AMP, 37): *slightly, very, really, too, quite, exactly, right, pretty, real, more, less, relatively, quite (Br. E. use), nearly, pretty, far from.* Degree adverbs are used to show that the degree or extent to which the characteristic of something holds in comparison with its usual state (ibid: 554).
    - Additive/restrictive adverbs help to add items at clausal or phrasal level (*also, too, else*) or restrict the value to only some items by emphasizing their importance: *especially, only, just, even, particularly*). Although almost all additive/restrictive adverbs are rather common in academic language, *only* is much more frequent appearing over1000 times per million (ibid: 562).
    - Stance adverbs, including downtoners but excluding hedges (see Hedges. No 36 in this list). Downtoners, in contrast to hedges, are characteristic to academic prose and indicate probability and, thus, are used by writers in academic texts to demonstrate the reliability of their research (Biber, 1988: 240). The following downtoners were included: *almost, barely, hardly, merely, mildly, nearly, only, partially, partly, practically, scarcely, and slightly.*
    - The following other epistemic stance adverbs are included into this category: *generally, indeed, probably, definitely, really, actually, apparently, reportedly, evidently, mainly, typically, approximately.*
35. Conjuncts (CONJ) (Biber, 1988) are linking adverbials that are used for explicit marking of logical relations in academic discourse: *additionally, alternatively, altogether, as a result, as a consequence, by comparison, by contrast, consequently, conversely, e.g. firstly, for example, for instance, furthermore, hence, however, i.e. incidentally, instead, in comparison, in contrast, in particular, in addition, in conclusion, in consequence, in sum, in summary, in any case, in other words, likewise, moreover, namely, nevertheless, nonetheless, notwithstanding, on the contrary, on the one/other hand, otherwise, overall,*

rather, secondly, similarly, so, then, therefore, thirdly, though, thus, etc. In general, conjuncts are more common in academic prose than in conversation, as "their frequency reflects the importance in academic prose of making the connections between ideas and explicitly showing the development of logical arguments" (Biber et al., 1999: 562).

36. Hedges (HED), in contrast to downtoners (see *Stance adverbs* in 34), are stance adverbs/adverbials that are "informal, less specific markers of probability or uncertainty" (Biber, 1988: 240). They may signal lack of commitment to what is said or written and are characteristic to informal involved rather than to informational academic discourse (Chafe and Danielewicz, 1987; Biber, 1988). This category includes *at about, something like, more or less, almost, maybe, sort of, kind of* (list taken from Biber, 1988: 240), *at least, o'clock, as well* (Biber et al., 1999: 542) and *roughly, perhaps* (ibid: 557), *so to speak* (ibid: 764).
37. Amplifiers (AMP), in contrast to hedges, are the words that indicate feelings of certainty or conviction of the speaker or writer. Such words as *absolutely, altogether, completely, enormously, entirely, extremely, fully, greatly, highly, intensely, perfectly, strongly, thoroughly, totally, utterly,* and *very* add power and force to a verb. Similarly to downtoners, they are used in academic discourse, but in contrast to them, amplifiers indicate certainty and, thus, reliability of proposition. Biber (1988) considers amplifies to be characteristic of involvement (Dimension 1).
38. General emphatics (GENEM) are words and expressions which mark the presence of certainty, e.g. *for sure, a lot, such a, real* (+ adj.)*, so* (+ adj.)*, just, really, most,* and *more* (forming comparative and superlative degree). However, in contrast to amplifiers, "emphatics simply mark the presence (versus absence) of certainty while amplifiers indicate the degree of certainty towards a proposition" (Biber, 1988: 241). Contrary to clausal conjuncts (see Linking adverbs in 35), emphatics are characteristic to the discourse that is informal, colloquial and involved with the topic and co-occur frequently with hedges (see Hedges in 36) in conversational genres (ibid.).
39. Discourse particles (DPART), e.g. sentence initial *well, now, anyway*, *anyhow,* and *anyways, so,* which are used to maintain conversational coherence. According to Biber, discourse particles are characteristic to conversational genres rather than to academic prose (1988: 241). Discourse markers that are particularly attributed to spoken dialogue are *please*, *(oh) well, good grief* (Biber et al., 1999: 140).

b) <u>Adverbial subordination</u>

40. Causative adverbial subordination (CSUB) – the use of *because, since, for* and *as* to subordinate a clause, and to present the idea of cause and effect.

41. Conditional adverbial subordinators (COND), e.g. *if, unless*
42. Other adverbial subordinators (ADVS), e.g. *since, while, whereas, so … then, either…or, neither…nor, as…as, so…as (not to), in order (to), as though, as if, as for (*Biber et al., 1999: 838).

Clauses

43. WH clauses (WHC) serve as complements to verbs. For example, *I believe what you tell me*.

2. Nominal post-modifying clauses

44. *that* relatives (RTHAT), e.g. *the dog that bit me; the dog that I saw*
45. WH relatives on object position (WHRCO), e.g. *the man who Sally likes*
46. WH relatives on subject position (WHRCS), e.g. *the man who likes popcorn*

3. Participial clauses

47. Past participial postnominal (reduced relative) clauses (PPC), e.g. *the solution produced by this process*
48. Present participle clauses (PRPCL)

Negation

49. Analytic (*not*) negation (ANEG) – using *not* and all of its contracted forms.
50. Synthetic negation (SYNEG) – negative words (e.g. *neither, nor, never)* and no- negation, e.g. *nobody, nowhere, nothing, no one*, *none* etc.

Other features

51. Gerunds (GER)
52. Prepositions (NPR) – the category does not include coordinators (see in 54 and 55)
53. Sentence relatives (SREL) are used to express the attitude of the speaker/writer; they refer to an entire clause. For example, *She never has time for any leisure activities*, *which I think is a pity*.
54. Phrasal coordination (PHC) – *and, but, or* connecting nouns, verbs, adjectives or adverbs or their phrases.
55. Clausal coordination (CLC) – *for, and, nor, but, or, yet, so* connecting independent clauses.

Three main reasons for selecting the listed-above linguistic features as dependent variables in the present study were the following:

1. They had been known as the features that help to attribute different text types to certain register dimensions.
2. They had been used by previous researchers, which was important for comparison of the results.

3. They had been known (Biber, ibid.) as adequately representing lexicogrammatical system of the English language.

Controlling moderator variables

The present study was focused on the internal linguistic properties of computer-mediated types of academic discourse, i.e. register variation in CMAD. As it has been explained in 1.2.3, the variation in discourse depending on the academic discipline was beyond its scope. Therefore, the field of knowledge as a variable responsible for variation in CMAD was isolated by restricting the disciplinary variation to only the texts representing the field of education.

The author also attempted to control the factors responsible for genre variation: the variables known as causing it, such as the differences in the purpose of writing and the intended audience, were reasonably isolated: all the CMAD texts chosen for the analysis in the present research had been written or orally presented to only academic audience with only educational purposes. The study was limited to the investigation of academic discourse, i.e. the use of language for teaching the subject matter at universities, as opposed to non-academic electronic discourses (e.g. personal communication, work place communication, etc.). Thus it is assumed that the effect of the factors causing generic variation had been minimized.

The following section provides the description of the compiled corpora and data-collection methods.

2.1.4 Description of the Compiled Corpora

The present research faced the challenge to create a corpus which would reflect a reasonably broad range of computer-mediated academic discourse types available via networked computers. However, the amount of CMAD posted on the WWW (as texts) and transmitted via the Internet (as messages) was at the time of the research and still is overwhelming and constantly growing, new discourse types and genres emerging, long-existing discourse types changing or merging. Therefore, it was not possible to represent all possible discourse types, genres, registers and styles. Thus, it was necessary to delimit the scope of this research to the types of CMAD currently predominant on the WWW. As the review of the literature on recent empirical research into computer-mediated discourse presented in Part 1 revealed, the six types of computer-mediated discourse had attracted the most attention of scholars and had been classified by previous researchers as generally conceived (while still developing) and predominant on the Internet in academic communication. Therefore, in the present research, CMAD was represented by the following

six different types: 1) oral presentations at on-line academic seminars, 2) academic e-mails, 3) synchronous academic conferencing (chat) sessions, 4) discussion forum messages, 5) academic weblogs and 6) academic hypertexts.

The following underlying assumptions guided planning and compilation of the corpus for this study:

1 The analysis was based on "actual instances of speech or writing" (Meyer, 2002: xiii) represented in the electronic form on the Internet.
2 Corpus linguistics methodology for corpus compilation was applied.
3 The corpus was compiled in the way that enabled identification and analysis of the specific linguistic features possessed by different types of CMAD.
4 The compiled corpus allowed to compare/contrast different types of CMAD.
5 Formal linguistic features characteristic to each type of CMAD in the corpus were regarded in the context of use and the function they perform.

Thus, the ultimate goal of designing the specialised corpora for this study was to enable both the systematic study of individual CMAD types and a comparison between them.

Compiled corpora

The whole compiled corpus contained computer-mediated messages and texts, both oral and written, exchanged by the members of European on-line communities of university teachers: *European Association for the Teaching of Academic Writing* (http://www.eataw.eu/listserv/), *Academici Knowledge Network* (http://www.academici.com) and the community of teachers using WebCT/ Blackboard technology (http://www.blackboard.com/Communities.aspx). The corpus contained the messages and texts that fall into the disciplinary domain of education. The data comprised approximately 1350 participants in computer-mediated communication representing a wide range of first language backgrounds (42 languages): Austrian, Danish, Dutch, English, Estonian, Finnish, French, German, Hungarian, Italian, Latvian, Lithuanian, Polish, Serbian etc. The percentage of native speakers of English was not higher than 0.3%. It was assumed that the participants possessed a native-like level of the English language proficiency, as they were teaching English academic writing at universities at the moment of research. Where available, the information was also collected on the participants' age group and gender.

The collection of data took place in 2002 -2008. Academic e-mails were collected in the process of computer-mediated communication by the author of the present research during that period. Many e-mails were forwarded by the author's colleagues who were willing to help with the data collection. The samples of academic discussion forums, weblog entries and

hypertexts, freely available on the WWW, were compiled and systematised. The data from synchronous CMC were obtained by automatic logging of all the participants' activities in the chat. The computer system automatically stored linguistic contributions of the participants in synchronous conferencing and the information about the senders of the messages and the time they accessed the system. Academic on-line seminars were automatically recorded, transcribed into text by the computer programme and stored in the WebCT seminar archive. The samples of compiled data are presented in Appendices 2-7. All possible measures were taken to prevent revealing the identity of the informants: names were changed and personal data were removed. Where possible, the permission to use the text of messages in the present research was obtained.

The whole corpus compiled for the present study contained a significant amount of linguistic data. The total number of words in the whole corpus was 987634. The data comprised six specialized corpora representing six types of CMAD. The division of the data in the corpus was made on the basis of the classification of CMAD developed for the present study (Fig. 1.1 on page 41). The data comprised the following types of texts:

1. The transcripts of five on-line seminars, i.e. oral presentations given by academic professionals to other university teachers and Ph.D. students and transmitted via the Internet. The texts of the transcripts were freely available on the Internet at the time of data collection for the resent research.
2. A collection of written/typed messages exchanged in synchronous computer-mediated communication ('chat' sessions), i.e. the messages exchanged among the participants of the mentioned above seminars (university teachers and Ph.D. students).
3. A collection of discussion forum messages, comprising the messages posted in academic discussion forums by language education professionals.
4. A collection of academic e-mails compiled in 2002 - 2008, consisting of academic e-mails sent to more than one person (i.e. not private) maintaining academic communication among university teachers, e.g. informational e-mails, requests, 'thank you' e-mails etc.
5. Messages posted in academic weblogs of applied linguists and teachers of academic writing.
6. Academic articles (hypertexts) published in on-line academic journals on the Internet (see the list of analysed articles in Appendix 17).

The structure of the corpus, the number of words in it, the number of messages/ texts in the corpus and the number of words per message/ text are presented in Table 2.1.

Table 2.1

Corpora compiled for the present study

Specialised Corpora	**Number of words in corpus**	**Number of messages/ texts**	**Mean number of words per message/ text**
Electronic seminars	50 005	5 seminars (transcripts)	10 001
Synchronous Conferencing (academic chat)	17154	10 (sessions)	1715
Academic e-mails	223 009	2687	83
Discussion postings	219 607	1523	144
Academic weblogs	290 709	1112	261
Academic hypertexts	183060	30 (articles)	6102
Corpus total	**938544**	**5387**	**174**

The corpora represented the language naturally occurring in authentic texts, i.e. the data were not specially elicited for the purposes of the present study. Although the texts exchanged in CMC were in the centre of the research, two types of CMAD – oral academic seminars and academic hypertexts – were included in the study with the purpose of comparison and revealing the linguistic variation on a broader scale of the continuum from oral speech, through computer-mediated messages, to edited and published hypertexts. Each message, text or transcript was labelled: the information on the date, type of communication, the topic, the number of authors/participants and the name of the discourse community in which the text had been produced was recorded and encoded in such a way that the computer software used in the research would not process the words in the label.

The whole corpus was divided into two unequal parts to be used in the qualitative and the quantitative analyses. The first part was used for the qualitative analysis. It was made 'semi-balanced', i.e. it consisted of six portions of different size representing different types of CMAD. There were two reasons for compiling different in size representative corpora of CMAD for the qualitative research. The first reason was the difference in availability that was delimited by time and access constraints, i.e. some types of CMAD were easier to obtain than other types. For example, academic hypertext and weblogs are normally freely available on the WWW for long periods of time while the access to academic chats is usually restricted to

only the students taking academic courses or the members of discussion forums, on the one hand, and a limited time of the session – they disappear as soon as the applet dialogue window is closed – on the other hand. This also posed the problem of ethical use of data and obtaining the permission for copying. The second reason was unequal distribution of medium-specific features in different types of CMAD and the intention of the author to compile the corpora that would contain enough material for qualitative analysis of all types of CMAD selected for this research. In some types of electronic discourse, e.g. synchronous chats and e-mails, medium-specific features appear more often than in other, e.g. in academic hypertexts. For this reason, the parts of the corpus representing academic weblogs, asynchronous discussion forums and published academic hypertexts were considerably bigger than those representing, for example, synchronous chats. It was assumed by the author that bigger samples are more informative for qualitative analysis of individual discourse types than smaller, though equal in size samples.

The second part of the corpus comprised the samples that were used for computerised quantitative research that aimed to test the hypothesis. These samples were delimited in size (1000 words) and strictly standardized. This provided a controlled body of data for verification of the statistical hypotheses. This part of the corpus was balanced, which enabled the comparison and the accuracy in the applied statistical procedures, since some measures, e.g. type/ token ratio, are text-size dependent. The total size of the corpus used in the quantitative study was limited to 60, 000 words. The corpus was divided into six equal parts of 10, 000 words, each representing a CMAD type, divided in their turn into 1000-word samples. [To compare: The Brown Corpus was divided into 2,000-word samples and was well-balanced (Meyer, 2002: xii)]. Thus, the corpus prepared for computer-processing in this study was made well-balanced and reasonably representative.

2.1.5 Computer software applied in the study

The research tools were selected for the present research from a range of available software for quantitative language analysis that had been developed for academic and scientific purposes and piloted in reliable university settings. The author applied a systematic approach to evaluation and selection of the programmes. Each programme was analysed on the following criteria: the number of linguistic features it measures, the length and number of text samples it is able to process. However, the most important criterion was whether the use of the software had been validated and the results had been published in refereed academic journals, i.e. the software is recognised by the academic community as reliable.

In the present research, the author gave the preference to RANGE and FREQUENCY programmes, which had been validated and successfully used in previously conducted studies and which results were reported in refereed academic publications. The programmes were designed by Nation and Coxhead (2002) and programmed by Alex Heatley at the School of Linguistics and Applied Language Studies, Victoria University, Wellington, New Zealand. RANGE is the programme that can be used to analyse up to 32 different texts at the same time. The programme provides a range or distribution frequency figures for each word in the texts. RANGE was used in this study to compare lexical richness (TTR) of the language in the corpora and the lists of the most frequent words. The programme FREQUENCY was used to make a frequency list of all the words in each specialised corpus in this study. It was used for finding the most frequent words in the messages posted in the samples of CMAD.

It was not possible, however, to analyse automatically all the linguistic features. The analysis of the frequency of occurrence and the types of such features as greetings, and partings in e-mails, abbreviations and emoticons in chats was conducted manually.

2.2 MEASURING VARIATION IN LINGUISTIC DIMENSIONS IN CMAD: THE RESEARCH PROCEDURE AND FINDINGS

The frequency of occurrence of each linguistic feature was measured to answer the first research question: *What is the frequency of occurrence of each linguistic feature set as a dependent variable in this study?* Further, the procedure of computing the descriptive statistics for the compiled corpus of CMAD is presented.

2.2.1 Descriptive Statistics

To make the results of the present study comparable with other similar studies, the author systematically applied the research instruments and followed the statistical procedures that were recommended in the description of the methodology by Biber (1988).

Normalisation

Normalisation was conducted for the comparability of the frequency counts in texts. Where it was possible, the samples used for computerised analysis were delimited by size to 1000 words. The frequency counts were computed per 1000-word samples, with the exception of word length, sentence length and type/token ratio (as recommended in Biber, 1988: 75). In other cases, the normalisation of frequency counts was made according to the following formula:

$$\frac{N * 1000}{L}$$

In the formula, *N* stands for the number of occurrences of the linguistic feature in the text and *L* for the total number of the words in the text. For each of the texts in the corpus, the frequency of occurrence for each linguistic feature was counted and normalised per 1000 words.

Characteristics of CMAD corpus as a whole

The descriptive statistics calculated for the whole CMAD corpus are presented in Table 2.1. The data in the table show the mean frequency of linguistic features per text (Mean), the minimum (Min) and the maximum (Max) values, the absolute number of occurrences (Sum) and the standard deviation (SD) for each dependent variable in the specialised corpus of CMAD texts.

As the statistics in Table 2.1 show, some linguistic features appear in the corpus much more frequently than other features.

Table 2.2

Descriptive Statistics for the Corpus as a Whole

	Code	Linguistic features	Mean	Min.	Max.	Sum	SD
1.	PRO1	First person pronouns	37.4	0.0	90.0	2243.0	18.5
2.	PRO2	Second person pronouns	15.4	0.0	32.0	924.0	9.7
3.	PRO3	Third person pronouns	13.4	0.0	29.0	802.0	6.8
4.	IT	Pronoun *it*	9.6	2.0	20.0	578.0	4.7
5.	DEMP	Demonstrative pronouns	7.3	0.0	18.0	436.0	4.9
6.	INDP	Indefinite pronouns	8.2	0.0	18.0	493.0	4.7
7.	CONTR	Contractions	13.1	0.0	28.0	786.0	7.0
8.	THATD	*that* deletion	3.7	0.0	17.0	222.0	4.0
9.	FPR	Stranded prepositions	1.2	0.0	6.0	74.0	1.5
10.	SAUX	Split auxiliaries	2.5	0.0	9.0	152.0	2.5
11.	WHQ	WH questions	3.1	0.0	16.0	184.0	4.2
12.	TTR	Type/token ratio	39.0	29.7	49.2	n/a	4.2
13.	MWL	Mean word length	4.8	4	6.2	n/a	0.5
14.	MSL	Mean syntactic length	17.9	6.4	32.7	n/a	7.0
15.	NOM	Nominalizations	23.0	7.0	57.0	1379.0	13.5
16.	N	Nouns	184.5	127	259	11069.	33.1
17.	PTV	Past tense verbs	12.6	0.0	44.0	757	12.2
18.	PAV	Perfect aspect verbs	1.8	0.0	15.0	465.0	4.0
19.	PRTV	Present tense verbs	32.0	10.0	77.0	1917.0	14.3
20.	ALPASS	Agentless passives	8.0	0.0	42.0	481.0	13.0
21.	BYPASS	*By* passives	1.9	0.0	9.0	111.0	2.1
22.	PMOD	Possibility modals	8.9	0.0	20.0	533.0	4.7
23.	NMOD	Necessity modals	1.5	0.0	7.0	89.0	1.6
24.	PRMOD	Prediction modals	8.0	0.0	25.0	479.0	7.5
25.	BE	*Be* as main verb	27.2	8.0	57.0	1632.0	13.4
26.	PVERB	Private verbs	6.6	0.0	16.0	395.0	4.3
27.	PUBV	Public verbs	3.52	0.0	16.0	211.0	3.6
28.	SUV	Suasive verbs	0.3	0.0	2.0	17.0	0.5
29.	INF	Infinitives	32.1	16.0	66.0	1929.0	12.9
30.	ATADJ	Attributive adjectives	50.2	17.0	109.0	3011.0	25.3
31.	PRADJ	Predicative adjectives	17.0	5.0	39.0	1018.0	6.8
32.	PLADV	Place adverbials	3.6	0.0	9.0	218.0	2.6
33.	TADV	Time adverbials	9.7	0.0	27.0	580.0	7.1
34.	ADV	Other adverbs	18.8	4.0	53.0	1126.0	12.0
35.	CONJ	Conjuncts	4.5	0.0	13.0	267.0	3.7
36.	HED	Hedges	2.1	0.0	9.0	125.0	2.3
37.	AMP	Amplifiers	4.2	0.0	14.0	252.0	3.7
38.	GENEM	General emphatics	8.0	0.0	19.0	478.0	6.4
39.	DPART	Discourse particles	5.0	0.0	17.0	297.0	4.4
40.	CSUB	Causative adverbial	1.3	0.0	5.0	76.0	1.4
41.	COND	Conditional adverbial	3.6	0.0	12.0	216.0	3.0
42.	ADVS	Other adverbial	3.0	0.0	11.0	180.0	2.3
43.	WHC	WH clauses	5.9	0.0	14.0	352.0	3.6
44.	RTHAT	*That* relatives	9.8	0.0	20.0	588.0	5.4
45.	WHRCO	WH relatives (object)	3.3	0.0	13.0	198.0	2.8
46.	WHRCS	WH relatives (subject)	1.4	0.0	6.0	86.0	1.6
47.	PPC	Past participial clauses	6.4	0.0	25.0	382.0	6.3
48.	PRPCL	Present participle clauses	5.1	0.0	18.0	304.0	4.4
49.	ANEG	Analytic negation	5.6	1.0	15.0	335.0	3.4
50.	SYNEG	Synthetic negation	2.4	0.0	9.0	146.0	2.3
51.	GER	Gerunds	13.3	4.0	27.0	798.0	4.8
52.	NPR	Number of prepositions	106.9	37.0	146.0	6411.0	22.2
53.	SREL	Sentence relatives	1.3	0.0	5.0	79.0	1.5
54.	PHC	Phrasal coordination	18.6	6.0	43.0	1115.0	8.8
55.	CLC	Clausal coordination	9.7	3.0	18.0	584.0	3.2

For example, nouns occur in the corpus 11069 times and prepositions – 6411 times. Other features are rather rare in the texts, for example suasive verbs appear in the whole corpus only 17 times.

Meaningful interpretation of these absolute figures, however, is restricted by limited comparability of the data in the table. Nouns and preposition are more frequent than many other linguistic features in the English language almost in any text, and the number of suasive verbs increases only in persuasive writing. Thus, the data in the table cannot give a complete and definite answer to the second research question – *Which linguistic features occur more frequently in which CMAD types?* Therefore, this question will be again addressed further in this section.

The data in the table, however, give valuable information about linguistic variation in the CMAD corpus as a whole. The data in table 2.1 already imply considerable variation in the use of linguistic features in the text samples in the corpus. A considerable difference, for example, can be seen between the minimum and the maximum values of many linguistic features, for example first-person pronouns, attributive adjectives, amplifiers. This implies that in some texts in the corpus the same linguistic feature may appear much more frequently than in other texts.

The CMAD corpus compiled for the present study has a number of distinctive characteristics in comparison with other corpora of academic English. This can be illustrated by the frequencies of occurrence of personal pronouns in the corpus.

Firstly, the difference has been found in the use of first and second-person pronouns between the present study and previous research. Biber et al. (1999), for example, have found that the most frequent pronoun in academic texts is *it*, followed by *we* and *they*. Other personal pronouns he reported to be rare in academic prose. However, as descriptive statistics presented in Table 2.1 and Table 2.2 on Page 82 show, the most frequent pronoun in the CMAD corpus compiled for the present study is the first-person pronoun *I* (total N=1082), followed by *you* (N=743). The third-person pronoun *it*, which has been found by Biber et al. (ibid.) to be the most frequent in academic prose, holds only the third position (N=578) in the CMAD corpus. It is followed by *we* (N=505) and they (N=270), which Biber found to be much more frequent in academic prose than *I* and *you*. Thus, the difference has been found in the distribution of personal pronouns in the compiled corpus representing CMAD in much high frequency of the first and the second-person pronouns in comparison with the data obtained by Biber (ibid.) for the corpus of written academic English. The higher frequency of the first and the second-person pronouns in CMAD corpus implies increased personal involvement and informality.

Secondly, the first-person pronoun *I* in the specialised CMAD corpora (Table 2.2) is the most frequent in synchronous academic conferences (chats) and the least frequent in academic hypertexts.

Table 2.3

Distribution of Personal Pronouns in CMAD Specialised Corpora

Pronouns	sem	chats	webl	e-mails	disc	htexts	Total
First-person pronouns (PRO 1)							
I	104	345	260	249	199	5	1162
me	6	65	39	27	14	0	151
my	5	61	59	48	29	0	202
mine	0	1	0	2	4	0	7
myself	1	2	3	1	6	0	13
we	206	53	43	80	87	36	505
us	16	8	19	15	23	1	82
our	43	8	5	28	32	0	116
ours	1	0	1	0	1	0	3
ourselves	0	0	0	1	1	0	2
Total 1st person pronouns	382	543	429	451	396	42	2243
Normalised per 1000 w	38	54	43	45	40	4	
Second-person pronouns (PRO 2)							
you	158	211	86	186	102	0	743
your	25	26	30	58	22	0	161
yours	2	1	0	5	0	0	8
yourself	1	2	4	3	2	0	12
yourselves	0	0	0	0	0	0	0
Total 2nd person pronouns	186	240	120	252	126	0	924
Normalised per 1000 w	19	24	12	25	13	0	
Third-person pronouns (PRO 3)							
they	75	26	63	34	70	2	270
them	16	16	13	19	28	8	100
their	32	4	36	13	45	42	172
theirs	0	0	0	0	1	0	1
themselves	6	6	7	0	3	1	23
she	3	17	35	2	3	0	60
her	0	13	19	0	6	1	39
herself	1	1	2	0	0	0	4
he	21	11	18	8	17	0	75
his	5	6	11	10	14	3	49
him	11	9	3	1	2	0	26
Total 3rd person pronouns	170	109	207	87	189	57	819
Normalised per 1000 w	17	11	21	87	19	6	
Total personal pronouns	**738**	**892**	**756**	**790**	**711**	**99**	3986
it	140	114	105	128	41	50	578

The explanation may lie in the fact that academic hypertexts still preserve many features of traditional academic prose while synchronous academic conferencing tends to resemble spoken interaction. This is also confirmed by a very high frequency of the second-person pronouns in synchronous academic conferences and their absence in academic hypertexts.

Thirdly, the pronoun *we* is more than two times more frequent in on-line seminars than in other types of CMAD. This may be explained by the intention of the presenters to use *we* inclusively to get the listeners more involved and to encourage their active participation.

The frequencies of all 55 linguistic features set as dependent variables for each text in each of the six specialised corpora (total number 3300) are presented in Appendices 8-13.

Standardisation of measurements

To answer the second research question and identify the most frequent linguistic features in CMAD types, further computing was necessary. The statistical procedure *standardisation* enabled easy comparison of the frequency of occurrence of linguistic features in different CMAD texts by translating the absolute frequency values to a single scale. It allowed the author to see how far the frequency of occurrence of individual features was from the mean value calculated for the corpus as a whole. All frequencies of the linguistic features in the corpus were standardised to mean 0.0 and standard deviation of 1.0. Variation of the frequency of linguistic features in each text from the mean in the whole corpus was measured in standard deviations, applying the following formula (McEnery et al., 2006: 303)**:**

$$k = \frac{F - \mu}{\sigma}$$

In the formula, k is the computed standard value, F stands for the frequency of the linguistic feature in the text, μ is the mean value and σ is standard deviation (SD).

For example:

The frequency values of k, μ and σ for on-line seminar (1) are the following:

F = 36 (the number of first person pronouns in the first on-line seminar text)

μ = 37.4 (the number of first person pronouns in the whole corpus)

σ = 18.5 (standard deviation calculated for the number of first person pronouns in the whole corpus.

The standardized value for the first-person pronouns in the first sample of on-line seminars was computed as follows:

$$k = \frac{36-37.4}{18.5} = -0.07475$$

The result means that the frequency of the first-person pronouns in this particular text is slightly lower than in the corpus as a whole.

In a similar way, all the standardised values (total number *N*= 3300) were computed for each text in the corpus. The mean standardised values of each dependent variable calculated for each CMAD type are presented in Table 2.3. The larger the values in the table, the more frequent the linguistic feature is in the corpus in comparison with other linguistic features, irrespectively of their frequency in the English language. Negative values show that the frequencies of the linguistic features are significantly lower than the mean frequency for the entire corpus.

Table 2.4

Variation of Mean Standardised Values of Dependent Variables in CMAD Corpus

Features	sem	chat	webl	email	disc	htext
First person pronouns	0,044132	0,914156	0,298114	0,416999	0,119786	-1,79319
Second person pronouns	0,33046	0,888112	-0,35111	1,012034	-0,28915	-1,59034
Third person pronouns	0,533554	-0,36223	1,076898	-0,6853	0,562924	-1,12585
Pronoun *it*	0,933883	0,37783	0,185351	0,677243	-1,18339	-0,99091
Demonstrative pronouns	1,005688	-0,54361	-0,48246	0,047566	1,148386	-1,17557
Indefinite pronouns	-0,1522	0,612331	-0,10972	0,527384	0,76099	-1,63878
Contractions	0,085561	0,613184	0,998206	0,142601	0,02852	-1,86807
that deletion	-0,64302	0,024732	1,681749	-0,49463	0,321511	-0,89034
Stranded prepositions	0,377801	0,511143	0,644485	0,111118	-0,82227	-0,82227
Split auxiliaries	-0,61823	-0,69887	0,02688	1,155819	0,792945	-0,65855
WH questions	-0,58703	1,935603	-0,15866	-0,53943	0,007933	-0,65842
Type/token ratio	-0,8122	0,113097	0,465203	-1,31209	1,200555	0,345439
Mean word length	-0,40875	-0,89924	-0,54252	0,148635	0,104045	1,597828
Mean syntactic length	0,036837	-1,50746	0,381291	-0,74679	0,277955	1,558173
Nominalizations	0,201289	-0,8953	-0,75452	-0,8138	0,579168	1,683167
Nouns	0,305408	-0,29836	-0,08704	-0,23195	-1,33685	1,648801
Past tense verbs	-0,83484	-0,49981	2,033303	0,014981	-0,0177	-0,69593
Perfect aspect verbs	-1,29932	0,433106	0,829089	0,482604	-1,05183	0,606349
Present tense verbs	-0,53554	1,704639	-0,26952	-1,00458	-0,54254	0,647553
Agentless passives	-0,18644	-0,57218	-0,53361	-0,56447	-0,30216	2,158854
By passives	-0,59358	-0,68855	0,118715	-0,1662	-0,02374	1,353351
Possibility modals	-0,29721	0,003581	-1,17808	0,648122	1,421571	-0,59799
Necessity modals	-0,17951	0,073916	-0,49629	-0,55965	0,644126	0,517412
Prediction modals	1,688389	-0,92718	-0,46248	0,121706	0,599677	-1,02011
Be as main verb	0,634102	1,603905	-0,5222	-0,81314	0,10444	-1,0071
Private verbs	-0,01936	-0,29822	0,956631	1,049583	-0,3447	-1,34393
Public verbs	1,398977	-0,85773	0,243101	-0,77517	0,600872	-0,61005
Infinitives	0,716345	-0,848	-0,42981	-0,35236	1,645657	-0,73183

Features	sem	chat	webl	email	disc	htext
Attributive adjectives	-0,85378	-0,57292	-0,21691	-0,23668	-0,22877	2,109063
Predicative adjectives	-0,36276	0,18138	1,372606	-0,46571	0,269619	-0,99514
Place adverbials	0,770955	0,888558	-0,05227	0,222139	-0,95389	-0,87549
Time adverbials	0,301019	1,599166	-0,20695	0,258689	-0,63026	-1,32166
Other adverbs	-0,75543	-0,41382	1,969128	-0,22219	0,186081	-0,76377
Conjuncts	-0,30831	-1,16623	1,005369	-0,76408	0,335123	0,898129
Hedges	0,581406	-0,56669	-0,52253	0,095674	1,287924	-0,87579
Amplifiers	-0,5939	1,106814	0,971837	-0,6209	0,242959	-1,10681
General emphatics	-1,10893	-0,48123	0,789849	0,931081	1,025235	-1,156
Discourse particles	-0,19432	0,834451	-0,743	1,383131	-0,1486	-1,13165
Causative adverbial subordinator	0,096033	-0,40814	1,032356	-0,76826	0,456157	-0,40814
Conditional adverbial subordinator	0,437724	-0,60608	-0,63975	0,437724	1,380515	-1,01013
Other adverbial subordinator	-0,81079	-0,55475	-0,59742	1,066826	-0,08535	0,98148
WH clauses	1,21701	-0,29281	0,338567	0,173859	0,173859	-1,61048
That relatives	0,699416	-1,47245	0,93869	-0,79144	0,496953	0,12884
WH relatives (object)	-0,10929	-0,54643	-0,255	1,529997	0,182142	-0,80143
WH relatives (subject)	1,638744	-0,7236	-0,34052	-0,21282	-0,78745	0,425648
Past participial clauses	-0,48489	-0,70625	0,052705	-0,65882	-0,21609	2,013345
Present participle clauses	-0,15262	-0,35866	-0,12973	-0,45023	-0,77074	1,861988
Analytic negation	0,004907	0,21101	-0,26008	0,03435	0,888204	-0,87839
Synthetic negation	-0,62257	1,679483	-0,10135	-0,27509	0,159261	-0,83974
Gerunds	0,398593	-1,11187	-0,08391	0,14685	1,237737	-0,5874
Number of prepositions	0,272167	-1,29335	-0,46111	0,564577	-0,41162	1,329343
Sentence relatives	-0,0111	-0,8104	0,921413	-0,8104	0,255331	0,455156
Phrasal coordination	-0,64928	-0,62643	-0,53504	-0,56931	0,801604	1,578454
Clausal coordination	0,429373	-0,60741	0,994889	0,08378	-0,60741	-0,29323

The data in Table 2.3 demonstrate that linguistic features are not distributed equally among the CMAD types. However, a too large number of dependent variables makes a comprehensive description of the variation among CMAD types rather difficult. For this reason, a multidimensional approach to register variation (Biber, 1988), described also in McEnery et al. (2006: 160-165), was used in order to answer the third research question in the nest section, which was *What is the mean frequency of each CMAD type on Biber (1988) textual dimensions?*

2.2.2 Statistical Analysis of Variation in Computer-Mediated Academic Discourse

Factor Analysis statistical procedure was applied to put the variables that serve some common function in the texts into groups and, thus, to reduce the number of variables to a manageable size.

Factor Analysis and Interpretation

Factor Analysis (data reduction) statistical procedure was conducted with the help of standard SPPS 6 statistical package to define the number of dimensions along which the variation in CMAD would be analysed in the present study. A full technical description of the procedure is given in Biber (1988, Chapter 5). In the present study, Factor Analysis, as a statistical instrument, was used to reduce the large number of linguistic features defined as dependent variables (55) to a small set of derived variables – *factors*. Each factor was a group of variables in the original data that could be generalised, i.e. a group of linguistic features with similar (negative or positive) size of correlation serving a similar function in the text. In other words, a factor (i.e. Biber's textual *dimension*) is comprised of the linguistic features that either frequently co-occur (*positive correlation*) or rarely co-occur (*negative correlation*) in the same type of texts, i.e. they are in 'complimentary distribution' (Biber, ibid.).

The Factor Analysis procedure was repeated five times, each time with a different number of extracted factors set (from 9 to 5) in order to make a decision on the optimal number of factors. The best result, in terms of Kaiser-Meyer-Olkin Measure of Sampling Adequacy, Chi-Square and the level of significance, was received for five extracted factors. Hence, the decision was made to apply a five-factor model in the present research. Table 2 presents Kaiser-Meyer-Olkin Measure of Sampling Adequacy, Chi-Square and the level of significance for five-factor model applied in the present study.

Table 2.5

KMO and Bartlett's Test results

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,602
Bartlett's Test of Sphericity Degrees of freedom Significance level	Approx. Chi-Square	3,778E3
	df	1275
	Sig.	,000

The extracted factors correspond to five major dimensions in the English language identified by Biber (ibid.):

1. *Involved versus Informational Production*
2. *Narrative versus Non-narrative Concerns*
3. *Elaborated versus Situation-Dependent Reference*
4. *Overt expression of persuasion*
5. *Abstract versus Non-abstract Style*

The linguistic variation in CMAD was investigated along these dimensions. The calculated standardised values were further used to compute the factor scores on each dimension. The linguistic features for each dimension had been identified by Biber in his 1988 study. They were used in the present study for the comparability of the findings with the results of previous research (Tables 2.5 – 2.9 adapted from Biber, 1988, and Biber et al., 2004, 2006).

Table 2.5 presents the linguistic features constituting Dimension 1 – *Involved/ Informational production.*

Table 2.6

Linguistic features constituting Dimension 1. *Involved/ Informational Production*

Features positively associated with involvement	**Features associated with informational focus**
1. PVERB private verbs	24. TTR type/token ratio
2. CONTR contractions	25. PLADV Place adverbs
3. PRTV present tense verbs	26. N nouns
4. PRO2 second person pronouns	27. MWL mean word length
5. ANEG analytic (*not*-)negation	28. NPR number of prepositions
6. DEMP demonstrative pronouns	29. ATADJ attributive adjectives
7. GENEM general emphatics	30. MSL mean syntactic length
8. PRO1 first person pronouns	
9. IT pronoun *it*	
10. BE *be* as main verb	
11. CSUB causative subordination	
12. DPART discourse particles	
13. INDP indefinite pronouns	
14. AMP amplifiers	
15. SREL sentence relatives	
16. WHQ 'Wh' questions	
17. PMOD possibility modals	
18. WHC 'Wh' clauses	
19. FPR final (stranded) prepositions	
20. THAD *that* deletion	
21. CLC (clause coordination)	
22. HED hedges	
23. GER gerund	

According to Biber (1988: 107), "Factor 1 represents a dimension marking high informational density and exact informational content versus affective, interactional, and

generalized content." On the one end of the continuum in this dimension, there are the linguistic features that indicate highly interactive communicative situation in which the language has been produced (positive features), affective and involved character of information production and generalised choice of lexis, for example, first and second-person pronouns, contractions, discourse particles, amplifiers and Wh-questions. In addition, the production of information has been characterised as generalised and fragmented. On the other end of the dimension continuum, there are the features that signify high informational density, precision and careful lexical choice (negative features), for example, high lexical richness (TTR), mean syntactic length, number of nouns and attributive adjectives.

Factor 2 was interpreted by Biber (ibid.) as a textual dimension that represents the continuum between narrative and non-narrative concerns of the language users (Table 2.6). High frequency of such linguistic features (positive features on this dimension) as past tense verbs, third person pronouns, perfect aspect verbs, public verbs and present participle clauses, according to Biber (Ibid: 109), "can be considered as distinguishing narrative discourse from other types of discourse". Present tense verbs negatively correlate with narration and, thus, their high frequency signal non-narrative concerns of the writer.

Table 2.7

Linguistic features constituting Dimension 2. ***Narrative/ Non-narrative Concerns***

Features positively associated with narrative concerns	**Features associated with non-narrative concerns**
1. PTV past tense verbs 2. PRO3 third person pronouns 3. PAV perfect aspect verbs 4. PUBV public verbs 5. SYNEG synthetic (*no*-)negation 6. PRPCL present participle clauses	7. PRTV present tense verbs 8. (ATADJ attributive adjectives)

The third dimension is constituted out of 'Wh' relative clauses on object position, 'Wh' relative clauses on subject position, phrasal coordination, nominalizations. These linguistic features have been defined by Biber as signalling text-internal, explicitly stated in the text reference versus text-external, depending on the situational context reference (endophoric versus exophoric reference in Halliday, ibid.). Time and place adverbials make reference explicit and situation-dependent; therefore, they are defined as negative features on this dimension (Table 2.7).

Table 2.8

Linguistic features constituting Dimension 3. ***Explicit versus Situation-Dependent Reference***

Features positively associated with explicit reference	**Features associated situation dependent reference**
1. WHRCO 'Wh' relative clauses on object position 2. WHRCS 'Wh' relative clauses on subject position 3. PHC phrasal coordination 4. NOM nominalizations	5. TADV time adverbials 6. (PLADV place adverbials) 7. (ADV other adverbs)

In a similar vein, Dimension 4 was conceived. A high frequency of prediction modals, suasive verbs, conditional subordination, necessity modals and split auxiliaries characterise texts as explicitly persuasive. Dimension 4 does not have negative features on the other side of the dimension continuum. Thus it was identified by the high concentration of only positive features in text types (Table 2.8).

Table 2.9

Linguistic features constituting Dimension 4 ***Overt Expression of Persuasion***

Features positively associated with overt expression of persuasion	**No negative features**
1. PRMOD Prediction modals 2. SUV Suasive verbs 3. COND Conditional subordination 4. NMOD Necessity modals 5. SAUX Split auxiliaries	

Finally, the features constituting Dimension 5 are the linguistic features that are positively associated with abstract information production – conjuncts, the passive voice, past participial clauses, predicative adjectives, infinitives – concentrate on the one end of Dimension 5 while high lexical richness, expressed by the type/token ratio, - on the other end (Table 2.9).

Table 2.10

Linguistic features constituting Dimension 5. ***Abstract Versus Non-abstract Style***

Features positively associated with abstract information production	**Features associated with non-abstract information production**
1. CONJ Conjuncts 2. ALPASS Agentless passives 3. PPC Past participial clauses 4. BYPASS By-passives 5. PRADJ Predicative adjectives 6. INF Infifitives	7. (TTR type/token ratio)

The factor score on each dimension for each text in the corpus was computed by calculating the sum of all the standardised values for the identified features. For example, the factor score on Dimension 1 for the text marked as Sem1 (on-line seminar) was computed by calculating the sum of all the features constituting Dimension 1:

> private verbs + contractions + present tense verbs + second person pronouns + analytic (*not*-)negation + demonstrative pronouns + general emphatics + first person pronouns + pronoun *it* + *be* as main verb + causative subordination + discourse particles + indefinite pronouns + amplifiers + sentence relatives + 'Wh' questions + possibility modals + 'Wh' clauses + final (stranded) prepositions + 'that' deletion + clause coordination + hedges + gerund + type/token ratio + Place adverbials + nouns + mean word length + number of prepositions + attributive adjectives + mean syntactic length.

The computed sum = 0.525391.

In a similar way, factor scores for each text in the corpus were computed. After that, the mean value for each factor in the group of texts was calculated for each textual dimension. For example, the mean value for Dimension 1 for all the texts representing on-line seminars was computed as being 2.032263 (~2). The results of the calculation of the mean values of factor scores for each textual dimension for each CMAD type are presented in Table 2.10.

Table 2.11

Textual Dimension Scores across CMAD types

	Dimension scores				
	Dim1	Dim 2	Dim 3	Dim 4	Dim 5
On-line seminars	2	-1,5	1,4	2.1	-1.5
Academic 'chats'	-1,1	1,7	-1,2	-2.7	-3
Academic e-mails	6,3	3,7	-2,1	-1.7	2
Discussion forums	0,9	-2,7	0,2	1.2	-3.6
Academic weblogs	4,4	-1,1	0,1	3.3	3.1
Academic hypertexts	-12,5	-0,2	1,6	-2.1	3

Notes: Dimension 1 (Dim 1) = Involved/ Informational production. Dimension 2 (Dim 2) = Narrative/ Non-narrative concerns. Dimension 3 (Dim 3) = Explicit/ Situation-dependent reference. Dimension 4 (Dim 4) = Overt expression of persuasion. Dimension 5 (Dim5) = abstract/ non-abstract style.

Statistical Procedure General Linear Models (ANOVA)

The next step in identifying the variation in CMAD, applying the methodology of Biber (1988), was calculating the statistical difference between CMAD types along five textual dimensions. The statistical procedure *General Linear Models* (ANOVA) was applied to test if there were significant differences among CMAD types in respect to each factor score.

Statistical Hypotheses

This study attempted to test the following statistical hypotheses:

Hull hypothesis

H0: There is no significant statistical difference (at significance level α=0.05) in the mean factor scores between the CMAD types along each dimension.

Alternative hypotheses

H1: There is a significant statistical difference (at significance level α=0.05) in the mean factor scores between the CMAD types along Dimension 1.

H2: There is a significant statistical difference (at significance level α=0.05) in the mean factor scores between the CMAD types along Dimension 2.

H3: There is a significant statistical difference (at significance level α=0.05) in the mean factor scores between the CMAD types along Dimension 3.

H4: There is a significant statistical difference (at significance level α=0.05) in the mean factor scores between the CMAD types along Dimension 4.

H5: There is a significant statistical difference (at significance level α=0.05) in the mean factor scores between the CMAD types along Dimension 5.

The author aimed to find sufficient sample evidence to reject the null hypothesis (H0), stating that there was no difference between CMAD types, in favour of the alternative hypotheses. Essential probability statistic called an *F* ratio (Fisher's *Six Sigma data set comparison*) was calculated with the help of SPSS statistical package.

Table 2.12

Results of the analysis of variance in CMAD

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Dim1	Between Groups	2229,995	5	445,999	28,144	,000
	Within Groups	855,752	54	15,847		
	Total	3085,747	59			
Dim2	Between Groups	272,647	5	54,529	46,137	,000
	Within Groups	63,822	54	1,182		
	Total	336,469	59			
Dim3	Between Groups	102,150	5	20,430	11,947	,000
	Within Groups	92,340	54	1,710		
	Total	194,491	59			
Dim4	Between Groups	473,497	5	94,699	34,906	,000
	Within Groups	146,501	54	2,713		
	Total	619,998	59			
Dim5	Between Groups	313,956	5	62,791	20,762	,000
	Within Groups	163,315	54	3,024		
	Total	477,271	59			

As the results presented in Table 2.11 show, the level of significance α for all cases was found to be less than 0.05; therefore, in no one of the cases the H0 hypothesis could be accepted. However, to reject the null hypothesis with higher degree of confidence, further statistical testing was necessary (triangulation). The author aimed to prove that at least one of the mean values was not the same as the other mean values in the group.

2.2.3 Triangulation of Data and Research Instruments

For triangulation, Scheffé's test was conducted to analyze the pairs of mean values to see if there were differences between them and reveal where exactly the differences lay. Multiple comparisons were conducted between the mean standardised frequency value for each one type of CMAD and the mean standardised frequency values for the other types along each dimension.

Dimension 1

The results of Scheffé's test for Dimension 1 are presented in Table 2.12 on Page 99.

1. *On-line seminars (sem)*

The mean standardised frequency value for the text samples representing CMAD of on-line seminars (*sem*) was compared with the values for other CMAD types (*chat, webl, email, disc, htext*). The significance level in all the cases except the CMAD type marked *htext* (academic hypertexts) was >0.05. Consequently, the mean standardised frequency value for on-line academic seminars significantly differed only from the value for academic hypertexts on Dimension 1.

2. *Synchronous conferences (chat)*

The mean standardised frequency value for the text samples representing CMAD of synchronous conferences *(chat)* was compared with the values for other CMAD types (*sem, webl, email, disc, htext*). The significance level <0.05 was found only for the values of *htext* and *webl*. Consequently, the mean standardised frequency value for synchronous conferences was significantly different from the values for two types of CMAD – academic weblogs and academic hypertexts on Dimension 1.

3. *Academic weblogs (webl)*

The mean standardised frequency value for the text samples representing CMAD of academic weblogs *(webl)* was compared with the values for other CMAD types (*sem, chat, email, disc, htext*). The significance level <0.05 was found only for the values of *htext* and *chat*. Consequently, the mean standardised frequency value for academic weblogs was

significantly different from the values for two types of CMAD – academic synchronous conferences and academic hypertexts on Dimension 1.

Table 2.13

Scheffé's Test Results for Dimension 1

Dimen-sions	(I) Type	(J) Type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Dim1	Sem	Chat	3,15853	1,78030	,678	-2,9907	9,3077
		Webl	-4,31544	1,78030	,333	-10,4646	1,8338
		email	1,14297	1,78030	,995	-5,0062	7,2922
		disc	-2,36273	1,78030	,879	-8,5119	3,7865
		htext	14,57024*	1,78030	,000	8,4210	20,7194
	Chat	Sem	-3,15853	1,78030	,678	-9,3077	2,9907
		Webl	-7,47397*	1,78030	,008	-13,6232	-1,3248
		email	-2,01556	1,78030	,935	-8,1648	4,1336
		disc	-5,52126	1,78030	,105	-11,6705	,6279
		htext	11,41171*	1,78030	,000	5,2625	17,5609
	Webl	Sem	4,31544	1,78030	,333	-1,8338	10,4646
		Chat	7,47397*	1,78030	,008	1,3248	13,6232
		email	5,45841	1,78030	,113	-,6908	11,6076
		disc	1,95271	1,78030	,943	-4,1965	8,1019
		htext	18,88568*	1,78030	,000	12,7365	25,0349
	email	Sem	-1,14297	1,78030	,995	-7,2922	5,0062
		Chat	2,01556	1,78030	,935	-4,1336	8,1648
		Webl	-5,45841	1,78030	,113	-11,6076	,6908
		disc	-3,50569	1,78030	,572	-9,6549	2,6435
		htext	13,42727*	1,78030	,000	7,2781	19,5765
	disc	Sem	2,36273	1,78030	,879	-3,7865	8,5119
		Chat	5,52126	1,78030	,105	-,6279	11,6705
		Webl	-1,95271	1,78030	,943	-8,1019	4,1965
		email	3,50569	1,78030	,572	-2,6435	9,6549
		htext	16,93296*	1,78030	,000	10,7838	23,0822
	htext	Sem	-14,57024*	1,78030	,000	-20,7194	-8,4210
		Chat	-11,41171*	1,78030	,000	-17,5609	-5,2625
		Webl	-18,88568*	1,78030	,000	-25,0349	-12,7365
		email	-13,42727*	1,78030	,000	-19,5765	-7,2781
		disc	-16,93296*	1,78030	,000	-23,0822	-10,7838

*The mean difference is significant at the 0.05 level.

4. *Academic e-mails (email)*

The mean standardised frequency value for the text samples representing CMAD of *academic e-mails (email)* was compared with the values for other CMAD types (*sem, webl, chat, disc, htext*). The significance level in all the cases except the CMAD type marked *htext* (academic hypertexts) was >0.05. Consequently, the mean standardised frequency value for academic e-mails significantly differed only from the value for academic hypertexts on Dimension 1.

5. *Academic discussion forums (disc)*

The mean standardised frequency value for the text samples representing CMAD of *academic discussion forums (disc)* was compared with the values for other CMAD types (*sem, webl, chat, email, htext*). The significance level in all the cases except the CMAD type marked *htext* (academic hypertexts) was >0.05. Consequently, the mean standardised frequency value for academic discussion forums significantly differed only from the value for academic hypertexts on Dimension 1.

6. *Academic hypertexts (htext)*

The mean standardised frequency value for the text samples representing CMAD of *academic hypertexts (htext)* was compared with the values for other CMAD types (*sem, webl, chat, email, disc*). The significance level in all the cases was <0.05. Consequently, the mean standardised frequency value for academic hypertexts significantly differed from the values for all other types of CMAD on Dimension 1.

To summarise, the mean frequency value for academic hypertext on Dimension 1 was found to be significantly different from the mean frequency values for other types of CMAD. In addition, a significant difference has been found between the mean frequency values for academic weblogs and synchronous academic conferences.

Dimension 2

The results of Scheffé's Test for Dimension 2 are presented in Table 2.13 on Page 101.

1. *On-line seminars (sem)*

The mean standardised frequency value for the text samples representing CMAD of on-line seminars (*sem*) was compared with the values for other CMAD types (*chat, webl, email, disc, htext*). The significance level <0.05 was found only for the values of *chat* and *webl*. Consequently, the mean standardised frequency value for on-line academic seminars significantly differed from only the value for academic weblogs and synchronous conferences on Dimension 2.

Table 2.14

Scheffé's Test Results for Dimension 2

Dimen-sions	(I) Type	(J) Type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Dim2	Sem	Chat	-3,25115*	,48619	,000	-4,9305	-1,5718
		Webl	-5,19415*	,48619	,000	-6,8735	-3,5148
		email	1,18043	,48619	,332	-,4989	2,8597
		disc	-,45260	,48619	,971	-2,1319	1,2267
		htext	-1,35669	,48619	,188	-3,0360	,3226
	Chat	Sem	3,25115*	,48619	,000	1,5718	4,9305
		Webl	-1,94300*	,48619	,013	-3,6223	-,2637
		email	4,43158*	,48619	,000	2,7523	6,1109
		disc	2,79855*	,48619	,000	1,1192	4,4779
		htext	1,89446*	,48619	,017	,2152	3,5738
	Webl	Sem	5,19415*	,48619	,000	3,5148	6,8735
		Chat	1,94300*	,48619	,013	,2637	3,6223
		email	6,37458*	,48619	,000	4,6953	8,0539
		disc	4,74156*	,48619	,000	3,0622	6,4209
		htext	3,83747*	,48619	,000	2,1582	5,5168
	email	Sem	-1,18043	,48619	,332	-2,8597	,4989
		Chat	-4,43158*	,48619	,000	-6,1109	-2,7523
		Webl	-6,37458*	,48619	,000	-8,0539	-4,6953
		disc	-1,63303	,48619	,062	-3,3123	,0463
		htext	-2,53712*	,48619	,000	-4,2164	-,8578
	disc	Sem	,45260	,48619	,971	-1,2267	2,1319
		Chat	-2,79855*	,48619	,000	-4,4779	-1,1192
		Webl	-4,74156*	,48619	,000	-6,4209	-3,0622
		email	1,63303	,48619	,062	-,0463	3,3123
		htext	-,90409	,48619	,632	-2,5834	,7752
	htext	Sem	1,35669	,48619	,188	-,3226	3,0360
		Chat	-1,89446*	,48619	,017	-3,5738	-,2152
		Webl	-3,83747*	,48619	,000	-5,5168	-2,1582
		email	2,53712*	,48619	,000	,8578	4,2164
		disc	,90409	,48619	,632	-,7752	2,5834

*The mean difference is significant at the 0.05 level.

2. *Synchronous conferences (chat)*

The mean standardised frequency value for the text samples representing CMAD of synchronous conferences *(chat)* was compared with the values for other CMAD types (*sem, webl, email, disc, htext*). The significance level in all the cases was <0.05. Consequently, the mean standardised frequency value for synchronous conferences significantly differed from the values for all other types of CMAD on Dimension 2.

3. Academic weblogs (webl)

The mean standardised frequency value for the text samples representing CMAD of academic weblogs *(webl)* was compared with the values for other CMAD types (*sem, chat, email, disc, htext*). The significance level in all the cases was <0.05. Consequently, the mean standardised frequency value for academic weblogs significantly differed from the values for all other types of CMAD on Dimension 2.

4. Academic e-mails (email)

The mean standardised frequency value for the text samples representing CMAD of *academic e-mails (email)* was compared with the values for other CMAD types (*sem, webl, chat, disc, htext*). The significance level <0.05 was found only for the values of *chat, webl* and *htext*. Consequently, the mean standardised frequency value for academic e-mails significantly differed from the values for academic synchronous conferences, academic weblogs and academic hypertexts on Dimension 2. No significant difference has been found on Dimension 2 between the mean frequency values for academic e-mails and academic seminars and discussions.

5. Academic discussion forums (disc)

The mean standardised frequency value for the text samples representing CMAD of *academic discussion forums (disc)* was compared with the values for other CMAD types (*sem, webl, chat, email, htext*). The significance level in all the cases except the CMAD types marked *chat* and *webl* was >0.05. Consequently, the mean standardised frequency value for academic discussion forums significantly differed only from the values for academic weblogs and synchronous conferences on Dimension 2.

6. Academic hypertexts (htext)

The mean standardised frequency value for the text samples representing CMAD of *academic hypertexts (htext)* was compared with the values for other CMAD types (*sem, webl, chat, email, disc*). The significance level <0.05 was found only for the values of *chat, webl* and *email*. Consequently, the mean standardised frequency value for academic hypertexts significantly differed from the values for academic synchronous conferences, academic weblogs and academic e-mails on Dimension 2. No significant difference has been found on

Dimension 2 between the mean frequency values for academic seminars and discussion forums.

To summarise, the mean frequency values for academic synchronous conferences and weblogs on Dimension 2 was found to be significantly different from the mean frequency values for other types of CMAD. At least two pairs of means have been found significantly different in each comparison. On-line academic discussion forums have been found similar to a considerable degree to academic seminars (Sig. 0.971) and rather similar to academic hypertexts (Sig. 0.632).

Dimension 3

The results of Scheffé's test for Dimension 3 are presented in Table 2.14 on Page 104.

1. *On-line seminars (sem)*

The mean standardised frequency value for the text samples representing CMAD of on-line seminars (*sem*) was compared with the values for other CMAD types (*chat, webl, email, disc, htext*). The significance level in all the cases except the CMAD types marked *chat* and *webl* was >0.05. Consequently, the mean standardised frequency value for on-line academic seminars significantly differed only from the value for academic synchronous conferences and academic weblogs on Dimension 3.

2. *Synchronous conferences (chat)*

The mean standardised frequency value for the text samples representing CMAD of synchronous conferences *(chat)* was compared with the values for other CMAD types (*sem, webl, email, disc, htext*). The significance level <0.05 was found only for the values of *sem* and *htext*. Consequently, the mean standardised frequency value for synchronous conferences was significantly different from the values for two types of CMAD – academic seminars and academic hypertexts on Dimension 3.

3. *Academic weblogs (webl)*

The mean standardised frequency value for the text samples representing CMAD of academic weblogs *(webl)* was compared with the values for other CMAD types (*sem, chat, email, disc, htext*). The significance level in all the cases except the CMAD type marked *chat* was <0.05. Consequently, the mean standardised frequency value for academic weblogs was significantly different from the values for all the types of CMAD except synchronous academic conferencing on Dimension 3.

Table 2.15

Scheffé's Test Results for Dimension 3

Dimen-sions	(I) Type	(J) Type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Dim3	Sem	Chat	2,57508*	,58481	,004	,5551	4,5950
		Webl	3,47452*	,58481	,000	1,4546	5,4945
		email	1,18973	,58481	,536	-,8302	3,2097
		disc	1,23728	,58481	,491	-,7827	3,2572
		htext	-,18169	,58481	1,000	-2,2016	1,8383
	Chat	Sem	-2,57508*	,58481	,004	-4,5950	-,5551
		Webl	,89943	,58481	,795	-1,1205	2,9194
		email	-1,38535	,58481	,360	-3,4053	,6346
		disc	-1,33780	,58481	,400	-3,3578	,6821
		htext	-2,75678*	,58481	,002	-4,7767	-,7368
	Webl	Sem	-3,47452*	,58481	,000	-5,4945	-1,4546
		Chat	-,89943	,58481	,795	-2,9194	1,1205
		email	-2,28478*	,58481	,017	-4,3047	-,2648
		disc	-2,23723*	,58481	,021	-4,2572	-,2173
		htext	-3,65621*	,58481	,000	-5,6762	-1,6363
	email	Sem	-1,18973	,58481	,536	-3,2097	,8302
		Chat	1,38535	,58481	,360	-,6346	3,4053
		Webl	2,28478*	,58481	,017	,2648	4,3047
		disc	,04755	,58481	1,000	-1,9724	2,0675
		htext	-1,37143	,58481	,371	-3,3914	,6485
	disc	Sem	-1,23728	,58481	,491	-3,2572	,7827
		Chat	1,33780	,58481	,400	-,6821	3,3578
		Webl	2,23723*	,58481	,021	,2173	4,2572
		email	-,04755	,58481	1,000	-2,0675	1,9724
		htext	-1,41897	,58481	,332	-3,4389	,6010
	htext	Sem	,18169	,58481	1,000	-1,8383	2,2016
		Chat	2,75678*	,58481	,002	,7368	4,7767
		Webl	3,65621*	,58481	,000	1,6363	5,6762
		email	1,37143	,58481	,371	-,6485	3,3914
		disc	1,41897	,58481	,332	-,6010	3,4389

*The mean difference is significant at the 0.05 level.

4. *Academic e-mails (email)*

The mean standardised frequency value for the text samples representing CMAD of *academic e-mails (email)* was compared with the values for other CMAD types (*sem, webl, chat, disc, htext*). The significance level in all the cases except the CMAD type marked *webl* was >0.05. Consequently, the mean standardised frequency value for academic e-mails significantly differed only from the value for academic weblogs on Dimension 3.

5. *Academic discussion forums (disc)*

The mean standardised frequency value for the text samples representing CMAD of *academic discussion forums (disc)* was compared with the values for other CMAD types (*sem, webl, chat, email, htext*). The significance level in all the cases except the CMAD type marked *webl* was >0.05. Consequently, the mean standardised frequency value for academic discussion forums significantly differed only from the value for academic weblogs on Dimension 3.

6. *Academic hypertexts (htext)*

The mean standardised frequency value for the text samples representing CMAD of *academic hypertexts (htext)* was compared with the values for other CMAD types (*sem, webl, chat, email, disc*). The significance level in all the cases except the CMAD types marked *chat* and *webl* was >0.05. Consequently, the mean standardised frequency value for academic hypertexts significantly differed only from the value for academic synchronous conferences and academic weblogs on Dimension 3.

To summarise, the mean frequency value for academic weblogs on Dimension 3 was found to be significantly different from the mean frequency values for other types of CMAD except synchronous conferences. A very high degree of similarity (Sig. 1.0) has been found between on-line academic seminars (spoken mode) and academic hypertexts (written mode). Rather high degree of similarity has been found between synchronous conferences and academic weblogs (Sig. 0.795) and between academic e-mails and discussion forums (Sig. 1.0) on this dimension.

Dimension 4

The results of Scheffé's test for Dimension 4 are presented in Table 2.15 on Page 106.

1. *On-line seminars (sem)*

The mean standardised frequency value for the text samples representing CMAD of on-line seminars (*sem*) was compared with the values for other CMAD types (*chat, webl, email, disc, htext*). The significance level <0.05 was found only for the values of *chat, webl* and *htext*.

Table 2.16

Scheffé's Test Results for Dimension 4

Dimen-sions	(I) Type	(J) Type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Dim5	Sem	Chat	4,82315*	,77773	,000	2,1368	7,5095
		Webl	3,85471*	,77773	,001	1,1684	6,5410
		email	,93653	,77773	,916	-1,7498	3,6228
		disc	-1,13420	,77773	,829	-3,8205	1,5521
		htext	4,26351*	,77773	,000	1,5772	6,9498
	Chat	Sem	-4,82315*	,77773	,000	-7,5095	-2,1368
		Webl	-,96844	,77773	,905	-3,6548	1,7179
		email	-3,88662*	,77773	,001	-6,5729	-1,2003
		disc	-5,95735*	,77773	,000	-8,6437	-3,2710
		htext	-,55964	,77773	,991	-3,2460	2,1267
	Webl	Sem	-3,85471*	,77773	,001	-6,5410	-1,1684
		Chat	,96844	,77773	,905	-1,7179	3,6548
		email	-2,91818*	,77773	,025	-5,6045	-,2319
		disc	-4,98891*	,77773	,000	-7,6752	-2,3026
		htext	,40880	,77773	,998	-2,2775	3,0951
	email	Sem	-,93653	,77773	,916	-3,6228	1,7498
		Chat	3,88662*	,77773	,001	1,2003	6,5729
		Webl	2,91818*	,77773	,025	,2319	5,6045
		disc	-2,07073	,77773	,233	-4,7570	,6156
		htext	3,32698*	,77773	,006	,6407	6,0133
	disc	Sem	1,13420	,77773	,829	-1,5521	3,8205
		Chat	5,95735*	,77773	,000	3,2710	8,6437
		Webl	4,98891*	,77773	,000	2,3026	7,6752
		email	2,07073	,77773	,233	-,6156	4,7570
		htext	5,39771*	,77773	,000	2,7114	8,0840
	htext	Sem	-4,26351*	,77773	,000	-6,9498	-1,5772
		Chat	,55964	,77773	,991	-2,1267	3,2460
		Webl	-,40880	,77773	,998	-3,0951	2,2775
		email	-3,32698*	,77773	,006	-6,0133	-,6407
		disc	-5,39771*	,77773	,000	-8,0840	-2,7114

*The mean difference is significant at the 0.05 level.

Consequently, the mean standardised frequency value for on-line seminars significantly differed from the values for synchronous academic conferences, academic weblogs and academic hypertexts on Dimension 4.

2. *Synchronous conferences (chat)*

The mean standardised frequency value for the text samples representing CMAD of synchronous conferences *(chat)* was compared with the values for other CMAD types (*sem, webl, email, disc, htext*). The significance level <0.05 was found only for the values of *sem, email* and *disc*. Consequently, the mean standardised frequency value for synchronous conferences significantly differed from the values for on-line seminars, academic e-mails and academic discussion forums on Dimension 4.

3. Academic weblogs (webl)

The mean standardised frequency value for the text samples representing CMAD of academic weblogs *(webl)* was compared with the values for other CMAD types (*sem, chat, email, disc, htext*). The significance level <0.05 was found only for the values of *sem, email* and *disc*. Consequently, the mean standardised frequency value for academic weblogs significantly differed from the values for on-line seminars, academic e-mails and academic discussion forums on Dimension 4.

4. Academic e-mails (email)

The mean standardised frequency value for the text samples representing CMAD of academic e-mails *(email)* was compared with the values for other CMAD types (*sem, webl, chat, disc, htext*). The significance level <0.05 was found only for the values of *chat, webl* and *htext*. Consequently, the mean standardised frequency value for academic e-mails significantly differed from the values for synchronous academic conferences, academic weblogs and academic hypertexts on Dimension 4.

5. Academic discussion forums (disc)

The mean standardised frequency value for the text samples representing CMAD of academic discussion forums *(disc)* was compared with the values for other CMAD types (*sem, webl, chat, email, htext*). The significance level <0.05 was found only for the values of *chat, webl* and *htext*. Consequently, the mean standardised frequency value for academic discussion forums significantly differed from the values for synchronous academic conferences, academic weblogs and academic hypertexts on Dimension 4.

6. Academic hypertexts (htext)

The mean standardised frequency value for the text samples representing CMAD of academic hypertexts *(htext)* was compared with the values for other CMAD types (*sem, webl, chat, email, disc*). The significance level <0.05 was found only for the values of *sem, email* and *disc*. Consequently, the mean standardised frequency value for academic hypertexts significantly differed from the values for on-line seminars, academic e-mails and academic discussion forums on Dimension 4.

To summarise the results for Dimension 4, at least three pairs of means have been found significantly different in each comparison signifying the existence of two distinct groups. The first group comprises three types of CMAD –discussion forums, on-line seminars and academic e-mails. The mean frequency values of each of these types of CMAD on Dimension 5 are significantly different from the values for each of the other three types of CMAD – academic weblogs, synchronous conferences and academic hypertexts. However, a high degree of similarity among the means has been found within the groups.

Dimension 5

The results of Scheffé's test for Dimension 5 are presented in Table 2.16 on page 109.

1. On-line seminars (sem)

The mean standardised frequency value for the text samples representing CMAD of on-line seminars (*sem*) was compared with the values for other CMAD types (*chat, webl, email, disc, htext*).

The significance level <0.05 was found only for the values of *webl, disc* and *htext*. Consequently, the mean standardised frequency value for on-line seminars significantly differed from the values for academic weblogs and academic discussion forums and academic hypertexts on Dimension 5.

2. *Synchronous conferences (chat)*

The mean standardised frequency value for the text samples representing CMAD of synchronous conferences *(chat)* was compared with the values for other CMAD types (*sem, webl, email, disc, htext*). The significance level <0.05 was found only for the values of *webl, disc* and *htext*. Consequently, the mean standardised frequency value for synchronous conferences significantly differed from the values for academic weblogs, academic discussion forums and academic hypertexts on Dimension 5.

3. Academic weblogs (webl)

The mean standardised frequency value for the text samples representing CMAD of academic weblogs *(webl)* was compared with the values for other CMAD types (*sem, chat, email, disc, htext*). The significance level <0.05 was found only for the values of *sem, chat* and *email*. Consequently, the mean standardised frequency value for academic weblogs significantly differed from the values for academic on-line seminars, academic synchronous conferences and academic e-mails on Dimension 5.

Table 2.17

Scheffé's Test Results for Dimension 5

Dimen-sions	(I) Type	(J) Type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Dim4	Sem	Chat	1,43353	,73661	,584	-1,1107	3,9778
		Webl	-3,54542*	,73661	,001	-6,0897	-1,0011
		email	2,07797	,73661	,178	-,4663	4,6222
		disc	-4,67199*	,73661	,000	-7,2163	-2,1277
		htext	-4,57575*	,73661	,000	-7,1200	-2,0315
	Chat	Sem	-1,43353	,73661	,584	-3,9778	1,1107
		Webl	-4,97896*	,73661	,000	-7,5232	-2,4347
		email	,64444	,73661	,978	-1,8998	3,1887
		disc	-6,10552*	,73661	,000	-8,6498	-3,5612
		htext	-6,00928*	,73661	,000	-8,5536	-3,4650
	Webl	Sem	3,54542*	,73661	,001	1,0011	6,0897
		Chat	4,97896*	,73661	,000	2,4347	7,5232
		email	5,62339*	,73661	,000	3,0791	8,1677
		disc	-1,12657	,73661	,799	-3,6708	1,4177
		htext	-1,03032	,73661	,853	-3,5746	1,5140
	email	Sem	-2,07797	,73661	,178	-4,6222	,4663
		Chat	-,64444	,73661	,978	-3,1887	1,8998
		Webl	-5,62339*	,73661	,000	-8,1677	-3,0791
		disc	-6,74996*	,73661	,000	-9,2942	-4,2057
		htext	-6,65371*	,73661	,000	-9,1980	-4,1094
	disc	Sem	4,67199*	,73661	,000	2,1277	7,2163
		Chat	6,10552*	,73661	,000	3,5612	8,6498
		Webl	1,12657	,73661	,799	-1,4177	3,6708
		email	6,74996*	,73661	,000	4,2057	9,2942
		htext	,09625	,73661	1,000	-2,4480	2,6405
	htext	Sem	4,57575*	,73661	,000	2,0315	7,1200
		Chat	6,00928*	,73661	,000	3,4650	8,5536
		Webl	1,03032	,73661	,853	-1,5140	3,5746
		email	6,65371*	,73661	,000	4,1094	9,1980
		disc	-,09625	,73661	1,000	-2,6405	2,4480

*The mean difference is significant at the 0.05 level.

4. Academic e-mails (email)

The mean standardised frequency value for the text samples representing CMAD of academic e-mails *(email)* was compared with the values for other CMAD types (*sem, webl,*

chat, disc, htext). The significance level <0.05 was found only for the values of *webl, disc* and *htext*. Consequently, the mean standardised frequency value for academic e-mails significantly differed from the values for academic weblogs, academic discussion forums and academic hypertexts on Dimension 5.

5. Academic discussion forums (disc)

The mean standardised frequency value for the text samples representing CMAD of academic discussion forums *(disc)* was compared with the values for other CMAD types (*sem, webl, chat, email, htext*). The significance level <0.05 was found only for the values of *sem, chat* and *email*. Consequently, the mean standardised frequency value for academic discussion forums significantly differed from the values for academic on-line seminars, academic synchronous conferences and academic e-mails on Dimension 5.

6. Academic hypertexts (htext)

The mean standardised frequency value for the text samples representing CMAD of academic hypertexts *(htext)* was compared with the values for other CMAD types (*sem, webl, chat, email, disc*). The significance level <0.05 was found only for the values of *sem, chat* and *email*. Consequently, the mean standardised frequency value for academic hypertexts significantly differed from the values for academic on-line seminars, academic synchronous conferences and academic e-mails on Dimension 5.

To summarise, at least three pairs of means have been found significantly different in each comparison signifying the existence of two distinct groups. The first group comprises three types of CMAD – synchronous academic conferences, on-line seminars and academic e-mails. The mean frequency values of these types of CMAD on Dimension 5 are significantly different from the other three types of CMAD – academic weblogs, discussion forums and academic hypertexts. Additionally, a very high degree of similarity has been found between academic discussions and hypertexts on this dimension.

To summarise, the mean frequency value for academic hypertext on Dimension 1 was found to be significantly different from the mean frequency values for other types of CMAD. In addition, a significant difference has been found between the mean frequency values for academic weblogs and synchronous academic conferences.

To conclude, the results of Scheffé's test showed that there was a significant statistical difference at the level of significance <0.05 between at least one pair of the mean values in each dimension. For this reason, H0 hypothesis was rejected in favour of the alternative statistical hypotheses in all the cases. Thus the applied method with a 95% level of confidence provided evidence that the type of CMAD was a possible reason for the variance in the frequency of co-occurrence of linguistic features in the specialised CMAD corpora.

2.3 DISCUSSION AND INTERPRETATION OF THE RESULTS

2.3.1 Scheffé's Test Results Interpretation

The results of the multidimensional analysis of variance of linguistic features revealed a significant statistical difference in the frequency of the use of lexicogrammatical features among different CMAD text types. For the interpretation of the results, it was important to find out where exactly the differences lay. For this reason, the frequency polygons for the mean values for all five dimensions were further analysed.

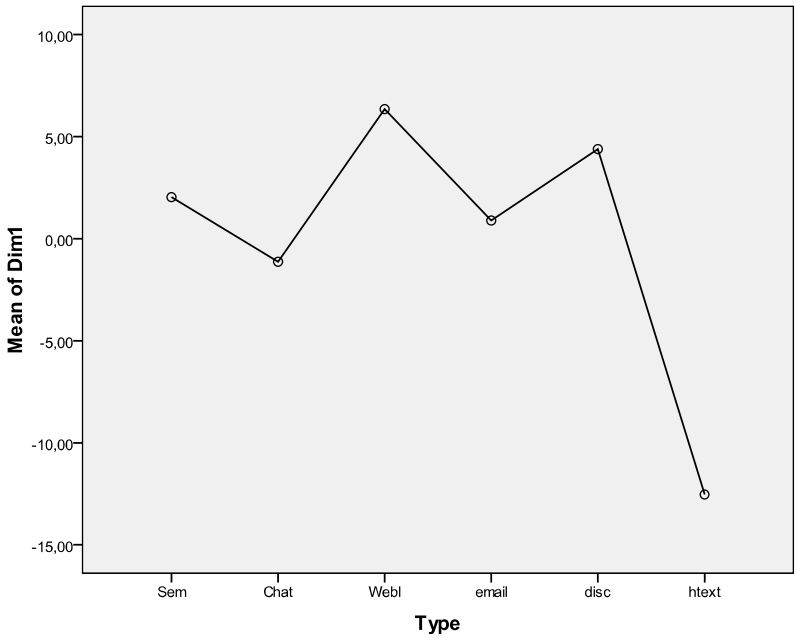

Figure 2.1 Mean difference values on Dimension 1 for six types of CMAD.

The frequency polygon in Figure 2.1 demonstrates that the mean difference values of five out of six types of CMAD hold high position on Dimension 1 – *Involved/ Informational production.*

The lower the position of the CMAD type in the frequency polygon, the less similarity it has with other CMAD types. The lowest position in this dimension is held by academic hypertexts.

Other CMAD types have the degree of similarity above zero, except synchronous conferences, which are close to zero. This confirms the previously made inference that the CMAD type marked academic hypertext significantly differs in the frequency of specific linguistic features from other five types of CMAD on Dimension 1, having the highest frequency of negative linguistic features on this dimension. The mean difference figures of

other types of CMAD hold the position above or close to zero, implying a very small or no difference between them.

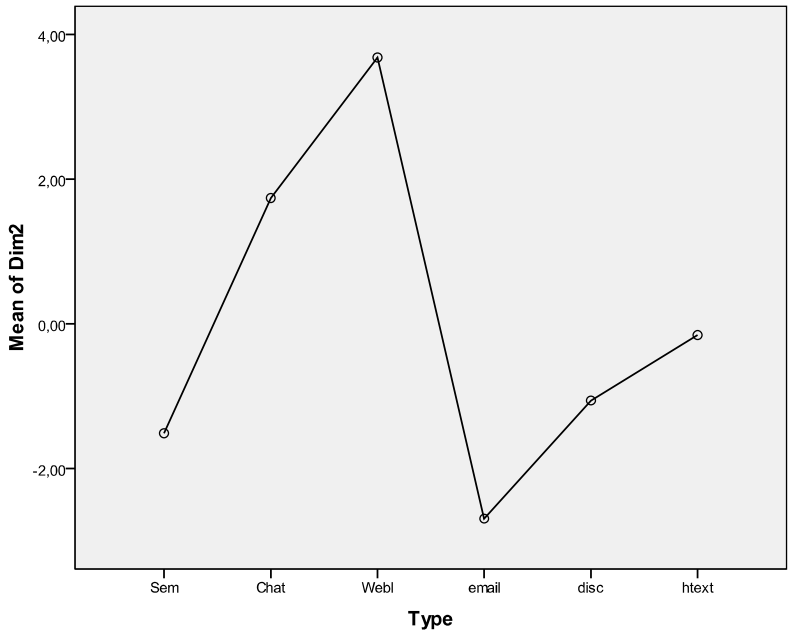

Figure 2.2 Frequency polygon presenting the mean difference values for six types of CMAD on Dimension 2.

The frequency polygon in Figure 2.2 demonstrates the mean difference values of CMAD types on Dimension 2 – *Narrative/ Non-narrative concerns.*

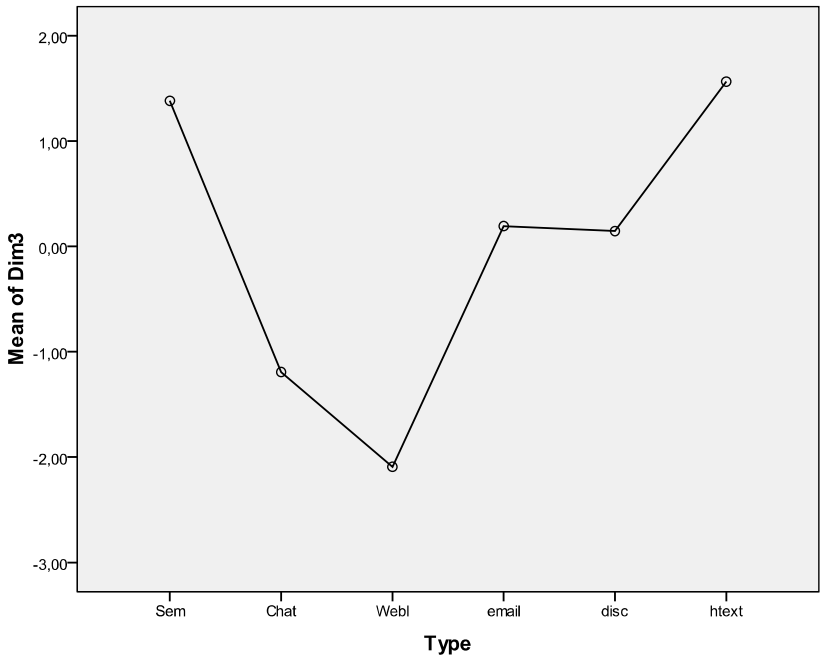

Figure 2.3 Frequency polygon presenting the mean difference values for six types of CMAD on Dimension 3.

Academic weblogs show the biggest mean difference from other types of texts on this dimension, followed by synchronous conferences. E-mails show the smallest mean difference from other types of CMAD. On-line academic discussion forums hold almost the same position as academic seminars and are close to academic hypertexts. This confirms the previously made inference that these types of CMAD have similar linguistic characteristics on this dimension.

The frequency polygon in Figure 2.3 on Page 112 demonstrates the mean difference values of CMAD types on Dimension 3 – *Explicit versus Situation-Dependent Reference.* Academic weblogs show the biggest mean difference from other types of texts on this dimension, followed by synchronous conferences. On-line academic seminars (spoken mode) and academic hypertexts (written mode) hold a high position on this dimension, which signals the high degree of similarity between them. The difference values of academic e-mails and discussion forums are almost identical, which implies high degree of similarity in linguistic characteristics between these two types of CMAD on this dimension.

The frequency polygon in Figure 2.4 demonstrates the position of CMAD types on Dimension 4 – *Overt expression of persuasion.*

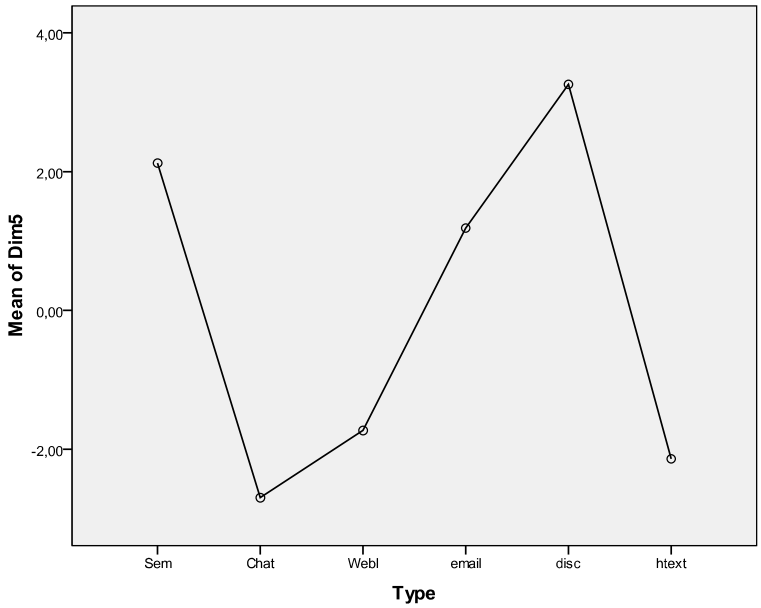

Figure 2.4 Frequency polygon presenting the mean difference values for six types of CMAD on Dimension 4.

On-line seminars, e-mails and discussion forums show similar mean difference figures on this dimension, implying that they contain the linguistic features signalling explicitly expressed persuasion. In contrast, synchronous conferences, weblogs and academic hypertexts do not demonstrate high frequency of such linguistic features. This confirms the previously

made inference that there are two distinct groups of CMAD types in respect of the explicitness of expression of persuasion.

The frequency polygon in Figure 2.5 demonstrates the mean difference values of CMAD types on Dimension 5 – *Abstract versus Non-abstract Style.*

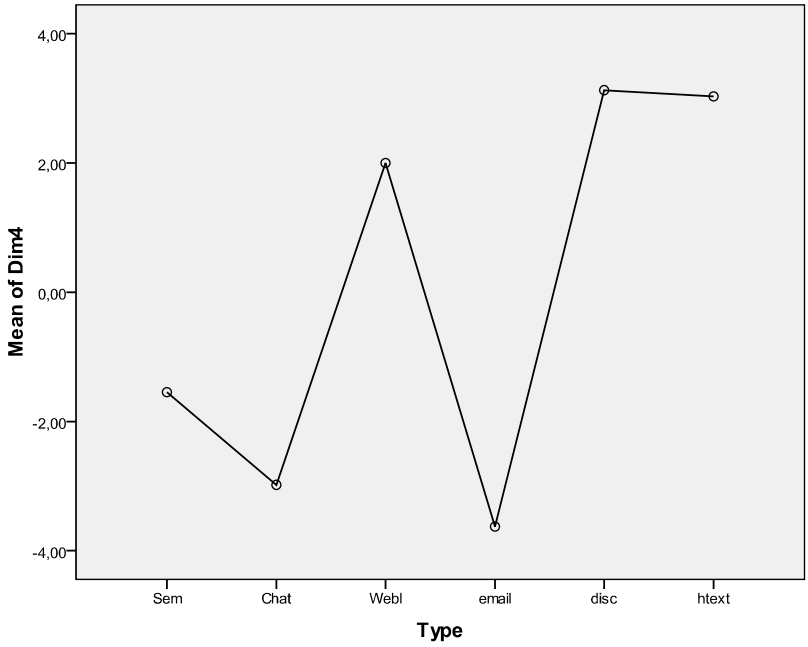

Figure 2.5 Frequency polygon presenting the mean difference values for six types of CMAD on Dimension 5.

In respect of the abstractness of the information in the texts, the CMAD types fall into two groups: academic weblogs, discussion forums and hypertexts demonstrate high degree of abstractness while on-line seminars, synchronous conferences and e-mails convey non-abstract information. The two groups considerably differ from each other in respect of abstractness of the information they convey. However, academic discussions and hypertexts show the highest degree of similarity, as they both render the most abstract information. In contrast, e-mails and synchronous conferences are rather similar in that they both convey non-abstract information.

The results of Scheffé's test show that there is a significant statistical difference between at least one pair of the mean values on each dimension. This means that the studied types of CMAD are rather similar on one dimension but different on another dimension. For example, academic hypertexts considerably differ from all the other types of CMAD on Dimension 1 (*Involved/ Informational production*), but are similar to discussion forums on Dimensions 2 (*Narrative/ Non-narrative concerns*) and Dimension 5 (*Abstract/ Non-abstract information*).

In addition, they are similar to on-line seminars on Dimensions 2 (*Narrative/ Non-narrative concerns*) and Dimension 3 (*Explicit/ Situation-dependent reference*) and to weblogs on Dimensions 5 (*Abstract/ Non-abstract information*) and Dimension 4 (*Overt expression of persuasion*). While academic weblogs differ from all other types of CMAD on Dimensions 2 (*Narrative/ Non-narrative concerns*) and Dimension 3 (*Explicit/ Situation-dependent reference*), they are similar to all but one (hypertexts) types of CMAD on Dimension 1 (*Involved/ Informational production*) and to hypertexts on Dimension 4 (*Overt expression of persuasion*) and Dimensions 5 (*Abstract/ Non-abstract information*).

Thus, the findings obtained in the research provide the evidence of significant multidimensional variation in the frequency of appearance of linguistic features in different types of CMAD.

2.3.2 Comparison of the Results with the Previous Research

The last research question to be answered in the present research was what the difference in the mean frequency scores between CMAD types in the specialised corpora in the present research and other types of discourse studied by previous researchers was. This question presupposed the comparison of CMAD types with other types of discourse.

The comparison of the results obtained in the present study with the results obtained by Biber in his large-scale study of genres and registers in English in 1988 is presented in Figures 2.6, 2.7, 2.11, 2.12 and 2.13.

The findings obtained in the research provide evidence of significant variation in the frequency of appearance of linguistic features in different types of CMAD, which may be interpreted as a generic difference between them. The mean scores of frequency and the configuration of the occurrence of linguistic features give the evidence of the position of each CMAD type in the dimensions among other types of discourse and genres identified by Biber (ibid). This makes a comparison among them and other types of discourse possible.

Positive and negative features that co-occur in different CMAD types are in a complimentary distribution, i.e. when positive features occur in the text, negative features are unlikely to occur in the same text. Figure 2.6 presents the position of CMAD types defined in the present study among the position of other types of English discourse on the Dimension 1 continuum of mean frequency scores.

The types of CMAD with high occurrence of positive features tend to locate closer to one of the ends of the dimension continuum, while those in which negative features prevail appear closer to the other end. The positive features on Dimension 1 are associated with

several major functional domains, including interactiveness and personal involvement. For example, discourse particles, *that*-deletions, contractions, first and second-person pronouns, 'Wh' questions are characteristic to academic weblogs and e-mails and, to a lesser extent, to academic seminars.

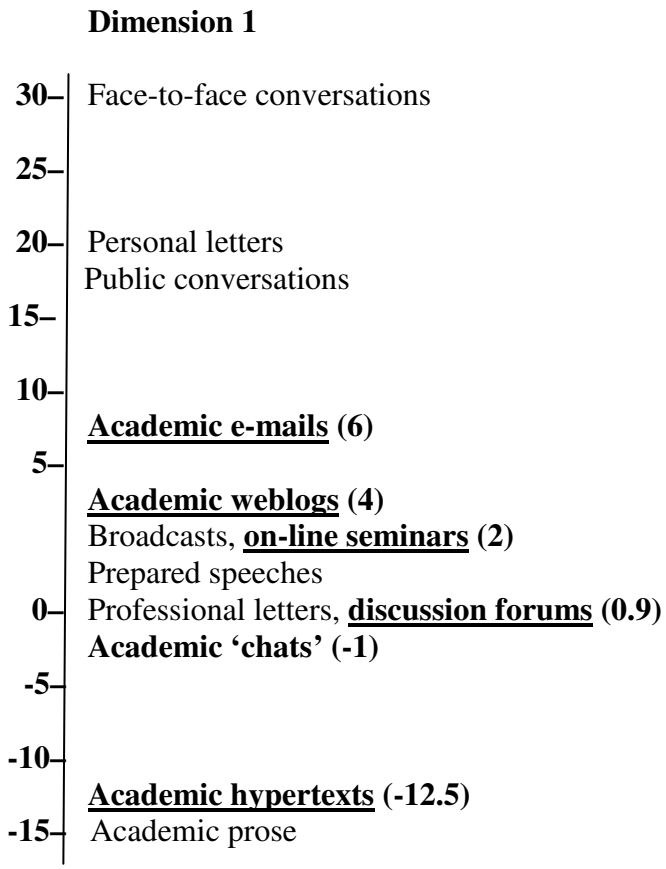

Figure 2.6 Mean scores of Dimension 1 – *Involved/ Informational production* computed in the present study for six types of CMAD (in bold), compared with six other English registers studied by Biber (1988).

In contrast, the negative features are associated mostly with informational density and complex noun phrase structures (frequent nouns and nominalizations, prepositional phrases, adjectives, and relative clauses) together with passive constructions. The negative features, e.g. long words, high type/ token ratio, and attributive adjectives frequently occur in academic hypertexts.

Figure 2.6 demonstrates that five out of six types of CMAD hold high position on Dimension 1 – *Involved versus informational production*, which is characterised by "marking high informational density and exact informational content versus affective, interactional, and generalized content" (Biber, 1988: 107). This implies high frequency of co-occurrence of the linguistic features that constitute Dimension 1 in the specialised corpora representing

academic weblogs, discussion forums, e-mails and synchronous conferences, for example, first and second-person pronouns, hedges, contractions, discourse particles, amplifiers, Wh-questions, and very low frequency or absence of the features that signal high informational density, precision and careful lexical choice (negative features), for example, high lexical richness (TTR), mean syntactic length, number of nouns and attributive adjectives. The lowest position in this dimension is held by academic hypertexts, which can be interpreted as caused by the highly informational character of these texts. In this respect, their characteristics are similar to the qualities of academic prose identified by Biber in his study, as they occupy almost the same position on the dimension continuum. This can be explained by the fact that academic hypertexts, as a new genre, have evolved from a traditional academic article and still have many of its characteristics.

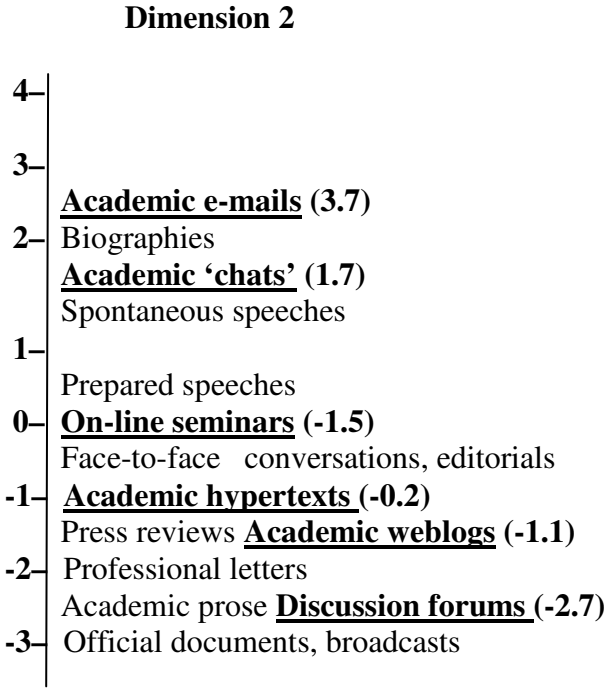

Figure 2.7 Mean scores of Dimension 2 – *Narrative versus Non-narrative concerns* computed in the present study for six types of CMAD (in bold), compared with six other English registers studied by Biber (1988).

Surprisingly, the most interactive type of CMAD – synchronous conferences (chats) – holds a significantly lower position on Dimension 1 than other interactive CMAD types and even than such a transactional type of CMAD as on-line seminars. The explanation may lie in the fact that academic chats, which accompany on-line seminars and lectures, are used by the participants to convey the information in a very condensed form: to ask their questions to the

presenter or to answer the presenter's questions; hence, there is a high level of information density in them. In addition, a big number of geographically distributed participants (N=100) at the seminar, informing other participants about their location, weather in their location and their affiliation (conference 'small talk'), resulted in a high frequency of place adverbials, other adverbs, nouns and other linguistic features that signal informational focus of the text (see a sample of synchronous conference text in the Appendix 7).

Figure 2.7 on Page 117 demonstrates the position of CMAD types on Dimension 2 – *Narrative/ Non-narrative concerns.*

Academic e-mails hold the highest position on this dimension. This signals of a rather narrative character of the messages, in which the writers (international academic language educators) share their stories about their teaching situations or the history of their research concerns with their colleagues (see a sample of academic e-mail in the Appendix 4).

Academic e-mails hold the highest position in Dimension 1 and Dimension 2, as positive features for these dimensions prevail in them. The most characteristic feature of Dimension 1 is the high frequency of personal pronouns in texts.

Figure 2.8 illustrates the distribution of personal pronouns across six CMAD types.

The use of first-person pronouns, for example, varies from 4 in hypertext to 54 in academic 'chat' samples per 1000 words. The biggest number of the third-person pronouns has been found in the specialised corpus of academic e-mail in the present study. This fact explained the high position of this type of CMAD on Dimension 2 continuum, as the third-person pronouns are the characteristic linguistic features (positive feature) on this dimension.

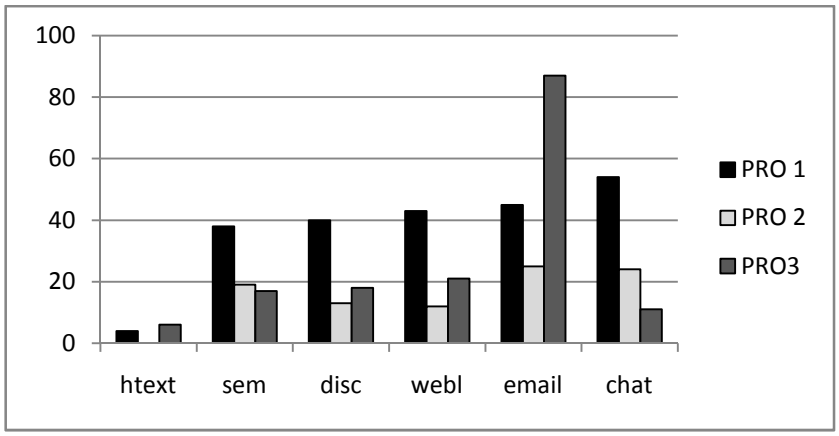

Figure 2.8 Distribution of personal pronouns across CMAD types per 1000 words.
Note: chat = academic 'chat', sem = on-line seminars, htext = academic hypertexts.

Further illustration is shown in Figure 2.9, which presents the distribution of occurrence of contracted forms, complimentiser *that* deletion and mean syntactic length per sentence in different types of CMAD.

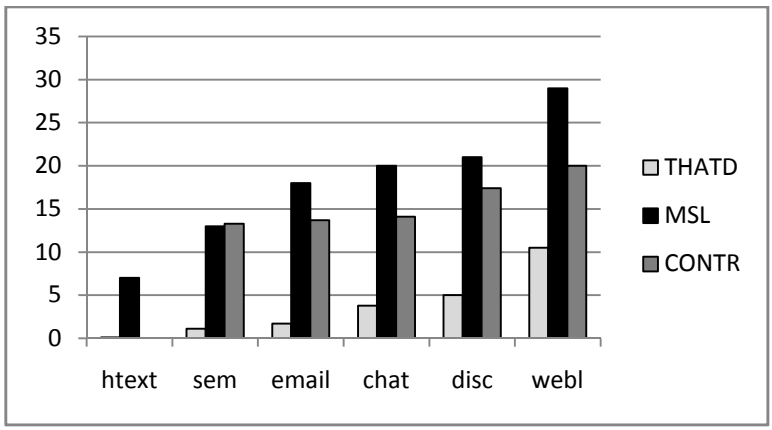

Figure 2.9 Number of contractions, occurrences of *that* deletion, and mean syntactic length per sentence in CMAD types per 1000 words.

No contractions have been found in the hypertext sample; on the contrary, the weblog sample contained 20 contractions, closely followed by 15 contractions in the sample of synchronous academic conferencing (*chat)*. The number of occurrences of *that* deletion changes from 0 in academic hypertexts to 10.5 in weblogs.

The explanation may lie in the fact that the software used for CMC in chats demand high speed of writing. These features reflect that the discourse type is interactional, informal and the focus of communication in this type of CMAD is on personal involvement.

However, academic hypertexts and discussion forum messages contain more negative features in this dimension: they contain more long words, more words per sentence, prepositions and passive constructions. They are usually well-thought while written without time constraints. While mistakes and careless choice of words are usually tolerated in chats, the errors are carefully eliminated and words are formal and accurately selected in academic hypertexts.

In addition, as illustrated in Figure 2.9, an average length of sentences increases from 7 words per sentence in academic chats to 29 words in academic hypertexts. The number of nominalisations grows from 10.9 in chats to 45.7 in academic hypertexts (Figure 2.10).

The emphasis in these texts is on the precision and clarity in information rendering, i.e. they have the emphasis on information transmission. Therefore, the type of discourse in them is predominantly transactional.

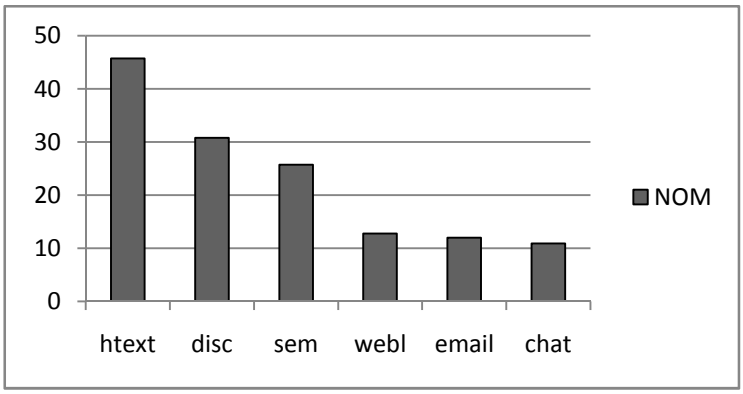

Figure 2.10 Number of nominalisations in CMAD types per 1000 words.

Figure 2.11 presents the position of CMAD types on Dimension 3 – *Explicit versus Situation-Dependent Reference*.

All the CMAD types but two (academic e-mails and synchronous conferences) demonstrate the emphasis on explicit reference, i.e. the antecedent is a noun that is present in the text. This may be explained by the fact that in computer-mediated communication the participants do not share the same context, apart from the computer software, and need to be more explicit for successful communicating of the meaning.

Both transaction types of CMAD, however, occupy a much lower position on this dimension than traditional academic prose in the study of Biber. The explanation may lie in the fact that, being transmitted by the Internet as a visually rich medium, these types of discourse utilise more facilities for visualisation offered by technology than traditional academic texts. Language users make explicit references to the objects on the screen, for example, "*click here*", "*look at this diagram*", "*now, let's go to the next slide.*"

Synchronous academic conferences and e-mails demonstrate the biggest number of situation-dependent references; however, in this respect, they resemble fiction much more than real face-to-face conversations. The participants have to be more explicit in these types of CMAD because they do not share the same physical context.

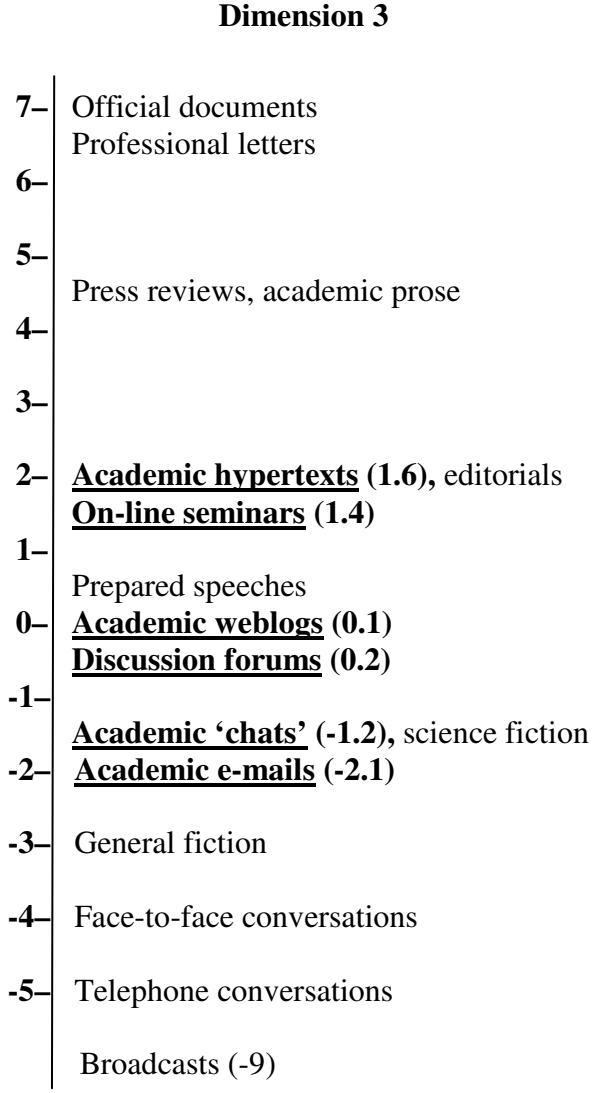

Figure 2.11 Mean scores of Dimension 3, *Explicit versus Situation-Dependent Reference* computed in the present study for CMAD types, compared with six other English registers studied by Biber (1988)

The position of CMAD types along Dimension 4 (*Overt expression of persuasion*) is plotted in Figure 2.12 on Page 122. Academic weblogs, discussions and on-line seminars show significantly higher scores on this dimension than academic e-mails, hypertexts and chats. This means that positive linguistic features that are important on this dimension are more frequent in them than in the other three types of CMAD, e.g. prediction and necessity modals, conditional subordination and suasive verbs. In this respect, they occupy the position between profession and personal letters on this dimension. In contrast, synchronous conferences, academic e-mails and hypertexts demonstrate much less overt expression of persuasion than face-to-face conversation and academic prose in the study of Biber.

Dimension 4

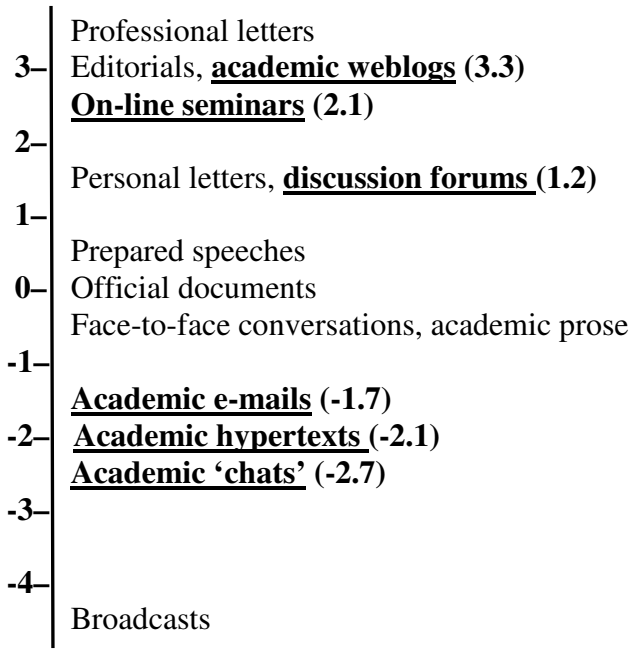

Figure 2.12 Mean scores of Dimension 4, *Overt expression of persuasion,* computed in the present study for CMAD types, compared with six other English registers studied by Biber (1988).

Dimension 5

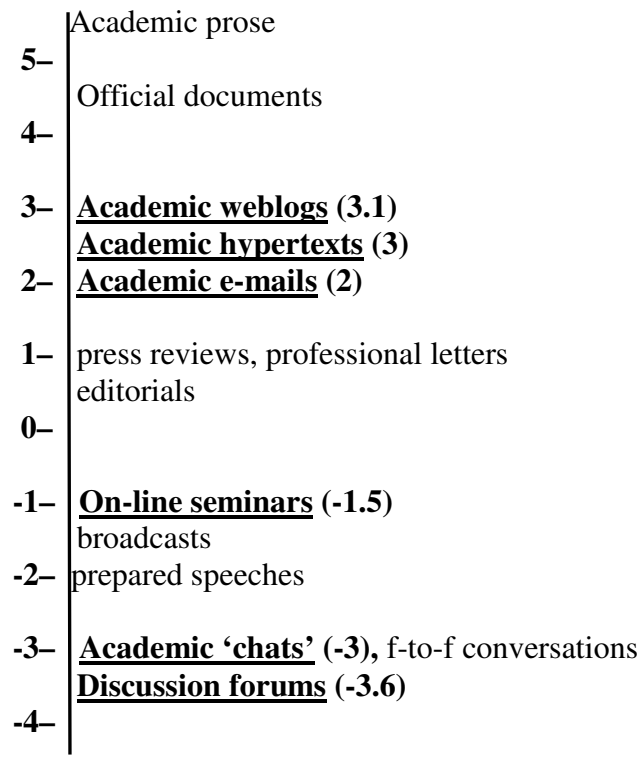

Figure 2.13 Mean scores of Dimension 5, *Abstract versus Non-abstract Style*, computed in the present study for CMAD types, compared with six other English registers studied by Biber (1988).

Figure 2.13 on Page 122 demonstrates the position of CMAD types on Dimension 5 – *Abstract versus Non-abstract Style*. In concern of the abstractness of the information conveyed, academic weblogs, hypertexts and e-mails hold a rather high position on this dimension. Although they resemble official documents in this respect, their position is much lower than that occupied by academic prose and official documents in the study of Biber. These three types of CMAD form a distinct group, demonstrating a high degree of similarity among each other.

They demonstrate a rather high frequency of the linguistic features that are important for this dimension, e.g. conjuncts, the passive voice constructions, past participial clauses, predicative adjectives and infinitives. On the opposite end of the continuum of Dimension 4, there is another group of CMAD types: on-line seminars, synchronous conferences and discussion forums. In respect of the abstractness of the conveyed information, these types of CMAD significantly differ from the previously described group. On-line seminars resemble broadcasts and prepared speeches, not surprisingly, and synchronous academic conferences and asynchronous discussions are similar to face-to-face conversations. In these types of CMAD, the participants tend to be less abstract, e.g. they rather explicitly name the objet or a person (the doer) than use the passive voice. In addition, they prefer finite verbal forms to non-finite verbal forms and construction. This may be explained by the fact that synchronous conferences and asynchronous discussions are the types of interactional discourse that resemble face-to face conversation while on-line seminar is a spoken type of transactional discourse resembling spoken communication in conveying a rather non-abstract information.

The results of the comparison of the findings obtained in the present study with the results of the previous research show similarities and differences of CMAD types with other types of English discourse on five functional dimensions.

1. On the continuum of Dimension 1 (*Involved/ Informational production*) on-line seminars hold a similar position as prepared speeches and academic e-mails while academic weblogs are rather similar to romantic fiction. Academic discussion forums resemble professional letters, and academic hypertexts are rather similar to traditional academic prose. Synchronous conferencing (‘chats) hold the position between broadcasts and general fiction on this dimension.
2. On Dimension 2 (*Narrative/ Non-narrative concerns*), synchronous conferences resemble spontaneous speeches, but academic hypertexts are rather similar to prepared speeches. Academic weblogs resemble interviews and face-to-face conversations; on-line seminars have common features with press reviews, and discussion forums are similar to academic prose on this dimension.

3. On Dimension 3 (*Explicit/ Situation-dependent reference*), academic hypertexts are similar to biographies; on-line seminars resemble editorials; discussion forums and weblogs resemble prepared speech; synchronous conferencing resemble science fiction, and academic e-mails – personal letters.
4. On Dimension 4 (*Overt expression of persuasion)*, academic weblogs are similar to professional letters while on-line seminars resemble fiction and religious texts. Academic discussion forums are similar to personal letters; academic e-mails, hypertexts and synchronous conferences hold the position between adventure fiction and press reviews.
5. On Dimension 5 (*Abstract/ Non-abstract information)*, on-line seminars are similar to broadcasts. Synchronous conferences resemble face-to-face conversations while discussion forums – telephone conversations. Academic weblogs, hypertexts and e-mails hold on this continuum the position between religious texts, press reviews and official documents.

To summarise, none of the investigated CMAD types has been found similar to any other type of English discourse on all five dimensions. Each type of CMAD resembles some types of English discourse on one dimension but different types on other dimensions. This confirms the inferences that the variation among CMAD types is multidimensional and that each type of CMAD has specific linguistic characteristics distinguishing it form other types of English discourse and other types of CMAD.

2.4 SPECIFIC CHARACTERISTICS OF THE TYPES OF COMPUTER-MEDIATED ACADEMIC DISCOURSE IDENTIFIED IN THE PRESENT STUDY

The following section presents a short description of the characteristic linguistic features of six investigated CMAD types. This description is not intended to be fully covering all the characteristic features of each CMAD type ever distinguished by the researchers studying them. Those features have already been depicted in Part 1 of the present study. In this section, only the findings that have been obtained in the course of the present research are described.

Synchronous academic conferences (chats)

Academic conferences ('chats') are defined as synchronous interactional text-based computer computer-mediated academic discourse. They have been found in the present study to possess the following specific characteristics:

1. Informational rather than involved type of information production
2. Non-narrative type of discourse
3. Non-explicit reference
4. Persuasion is not explicitly expressed
5. A rather non-abstract style

The most frequent linguistic features in synchronous conferences are Wh-questions, present tense and perfect aspect verbs. Other frequent features are time and place adverbials, amplifiers, first- and second-person pronouns, discourse particles, contractions, indefinite pronouns, final prepositions. In comparison with the other types of CMAD investigated in the present study, synchronous conferences contain the shortest sentences and words. Although being usual in academic writing, conjunctions, relative clauses, nominalisations, gerunds and infinitives, passive voice and prediction modals are rare in synchronous academic conferences (Appendix 16).

The specific characteristics of synchronous academic conferences can be illustrated by the following examples. Consider the example from an on-line seminar (Text 1). The participants of the seminar, accessing it from globally distributed locations, can simultaneously listen to the lecturer presenting in English, see the charts and illustrations on the screen, and ask questions to the presenter, using a facility similar to that commonly known as 'chat'. However, at the beginning of the seminar the listeners use the facility mostly for informal communication among themselves.

Text 1

1. Ivan says, "saludos"
2. Ivan says, "esto del puzzle esta muy interesante"
3. Ivan says, "where are they available?"
4. maria says, "Carmeta: you're from Barry U?"
5. lynda says, "Hello everyone. Apologies for arriving late. Better late than never is what I am shooting for everything these days it seems. "
6. Carmeta says, "si muy interesante"
7. Chandra says, "Do we also get to use these los?"
8. Bertha says, "it's like a lot of our traditional classes, eh?"
9. martinez says, "lol. yo sabia que los latinos estaban por algun lugar aqui"
10. Carmeta says, "yes I am in the School of Ed at Barry"
11. Ivan says, "aqui estamos en PR"
12. Carmeta says, "I work as the instructional designer, distance edu"
13. Team says, "Why can't the user setup the puzzle without Flash?"
14. martinez says, "wow.. que bueno... "

Only Carmeta and Maria (in lines 4, 10 and 12) seem to maintain a kind of a dialogue, exchanging questions and answers. The rest of the participants exist in the context of their geographical location and cannot easily overcome the constraints of their language and socio-cultural background. The communication process and the language that is used for communication resembles a set of disconnected utterances rather than a conversation.

In the example below (Text 2), by the end of the same on-line seminar the participants have learned to work in the new environment to a certain extent, and the communication becomes more effective. The utterances, however, look as disconnected as before: there are many unanswered questions (e.g. lines 15, 16, 19, 23) and some answers without questions (e.g. lines 20, 21, 22, 26, 28). This happens because the communication takes place in many modes simultaneously: the presenter orally answers the questions asked by the participants in written form, and the participants type the answers to the questions they hear, or they comment on the visuals they can see on their screens.

Text 2

15. Carmeta says, "I think learning objects (reusable) a great idea. Are there issues about copyright and adaptability? "
16. s_falcone says, "are the learning objects shown here for sale or lease?"
17. Kevin_and_Marc says, "Well, any programming I have tried is extremely difficult. Flash scripting is intuitive."
18. Pablex says, "not flash is very easy"
19. Clare says, "How do you get the little smilies into your message?"
20. Lana says, "No there is not. Their residency means that you have to be enrolled in at least three consecutive semesters earning at leat 12 credits"
21. Harry says, "No, upstate New York"
22. sbg says, "no, class synchronous size"
23. shirley says, "I am unable to access the WebCT games page even though I have registered and logged in?"
24. Bertha says, "thank you for an interesting presentation--I must go now"

25. Brenda says, "I keep losing the audio"
26. sbg says, "yes, thanks"
27. Carmeta says, "I think this was very informative and very smooth too. Good job!"
28. maria says, "I see. Sounds great. Do you know who the person responsible for the program is? Any contact name there? Also, is there a webpage for the program? Thanks for all your info Lana"

As it is illustrated by examples, the density of information per utterance is very high, which defines informational rather than involved type of information production in synchronous conferences. The explanation may lie in the fact that very high speed of interaction demands for a fast response that is to be quickly written and read by both communicators before it disappears from the dialogue window. Other characteristics of synchronous conferences make them rather similar to face-to-face spoken conversations, though with very specific features attributed to text-based, as opposed to spoken, communication mode: unconventional spelling, abbreviations etc. (described in 1.1.4.3).

Academic e-mails

The type of discourse of academic e-mails is defined as asynchronous interactional text-based computer-mediated academic discourse. On-line academic emails have been found in the present study possessing the following specific characteristics:

1. Involved information production
2. Rather narrative type of discourse
3. Non-explicit in giving reference
4. Not overtly persuasive
5. A rather abstract style

The information production in academic e-mails has been found the most involved of all the studied types of CMAD. In addition, situational reference is expressed the least explicitly in comparison with other CMAD types.

The most frequent linguistic features in academic e-mails are Wh relative clauses on object position, discourse particles, suasive verbs, adverbials, private verbs, first- and second-person pronouns, general emphatics, pronoun *it*, possibility modals, indefinite pronouns, perfect aspect verbs, conditionals, time adverbials. As not frequent in academic e-mails have been found present tense verbs, nominalisations, *be* as the main verb, sentence relatives, *that* relatives, public verbs, clause subordination, conjunctions, third-person pronouns, past participial clauses, amplifiers, phrasal coordination and passives (Appendix 16). In addition, academic e-mails are characterised by low type/token ratio and rather short sentence length. This implies that academic e-mails, in contrast to traditional academic letters, usually are not carefully drafted and edited.

Notable peculiarity of academic e-mails is a specific form of address, in which a rather formal address word *dear* is followed by the first name of the addressee instead of the surname, as it is usually done in traditional academic letters. Text 3 provides an example of such greeting, in addition to a rather informal ending (*cheers*).

As the example of a typical academic e-mail (Text 3) illustrates, this type of CMAD is in transition between traditional, paper-format academic letters and informal personal letters. However, the border between still appropriate and unacceptable informality of the writing style is vague in academic e-mails, which makes the choice of linguistic means of expression in them rather difficult.

Text 3

Dear Mary,
Ah, now I get it. Thanks. Would this imply that one kind of class activity could be trawling publications for nice turns of phrase to use later? I can see how students might enjoy doing this.
Cheers

John

Academic e-mails are usually written more carefully than informal personal e-mails; they rarely contain spelling or grammar mistakes, unconventional spelling or peculiar use of punctuation. On the other hand, the analysis of the collected samples reveal frequent informal greeting and parting sections in academic e-mails and a rather frequent use of emotion ☺ to render a smile.

Discussions forums

Discussion forums are multi-user asynchronous, interactional, text-based computer-mediated academic discourse. On-line academic discussion forums have been found in the present study to possess the following specific characteristics:

1. Rather involved information production
2. Non-narrative type of discourse
3. Neither explicit nor non-explicit in giving reference
4. Rather overtly persuasive
5. A very abstract style

As the most frequent in discussion forums have been identified the following linguistic features: infinitives, possibility modals, conjunctions, hedges, gerunds, demonstratives, general emphatics, analytic negation, phrasal coordination, suasive verbs, indefinite pronouns, necessity modals, public verbs, prediction modals. These features and a high type/token ratio

characterise them as similar to academic prose. The rarest linguistic features in discussion forums are nouns, pronoun *it*, perfect aspect verbs, place adverbials, final prepositions, Wh-relative clauses on subject position, present participle clauses, time adverbials, clausal coordination, present tense verbs, private verbs, passives, second-person pronouns, attributive adjectives (Appendix 16).

Text 4 provides an example of a typical message posted in an academic discussion forum. It illustrates that this type of CMAD possesses the characteristics described above.

Text 4

> Dear All
> This is a very interesting topic you have here. At the Leiden English Department (Netherlands), I've been involved in designing and teaching a Blackboard course on how to write Masters theses at our department. The problem you mention in your abstract is exactly the one we encountered when we were designing the course. There were a number of books and articles that were somewhat useful to us, as were a number of existing on-line writing courses, mainly OWLs in the U.S., and other related on-line courses. But none of these hit the spot so to speak, because they were either too abstract or they were geared to local conditions which differed considerably from ours. At Leiden we ended up having to develop a lot of material of our own. I'd certainly be interested in attending your symposium and making a short contribution about our Leiden experience.
>
> Cheers

The academic discussion forum message in the example, on the one hand, resembles a spoken announcement made at a public meeting: it contains first-person pronouns, short forms (*I've, I'd*), informal expressions (*hit the spot, so to speak*). On the other hand, the author uses formal academic vocabulary, the pronoun *one* and specific formal expressions, e.g. *involved in designing, making a short contribution).* This implies that even professional language educators still have a problem of style choice in computer-mediated academic discussion forums. This type of CMAD is still in the process of development.

Academic Weblogs

Academic weblogs are characterised as asynchronous, rather interactional (i.e. transitional between interactional and transactional) written computer computer-mediated academic discourse. In the present study, they have been found to possess the following specific characteristics:

1. Involved information production
2. Rather non-narrative type of discourse

3. Neither explicit nor non-explicit reference
4. The most overtly persuasive of all CMAD types
5. A very abstract style

The most frequent linguistic features in academic weblogs are past tense verbs, adverbs, that-deletion, predicative adjectives, third-person pronouns, clausal subordination, conjunctions, contractions, amplifiers, present tense verbs, that relatives, sentence relatives, perfect aspect verbs, general emphatics. Rare features are possibility modals, nominalisations, discourse particles, conditionals, adverbials, phrasal coordination, passives, hedges, necessity modals, demonstratives, prediction modals and infinitives (Appendix 16).

The example (Text 5) illustrates that the text of an academic weblog consists of the messages written by different people. The entries written in italics are posted in the weblog by a reader. The weblog owner expresses his or her ideas and comments on the readers' responses. Therefore, the text in a weblog has some characteristics of interactive dialogue, e.g. such features as second-person pronouns, and comments.

Text 5

Well, you partly answered your own question. A university is supposed to generate new and interesting ideas. Just as a firefighter needs access to a lot of water, a professor needs access to a lot of ideas -- including ones that some people may find offensive, threatening, or ridiculous.

Among the ideas that have been called offensive, threatening, or ridiculous is the idea that faculty members who've proven themselves after years of effort shouldn't be fired for being eccentric or voicing outlandish opinions (so long as they are still performing the duties for which they were hired).

*Aaaaah...so it's *only* professors who need access to new ideas, whereas a "normal person" not only doesn't need new ideas, but also doesn't need to be able to express themselves in their personal life.*

*I'm obviously playing devil's advocate, but I genuinely find it disturbing to repeatedly hear that employers have the **right** to fire you because they don't like you expressing yourself outside of work.*

I can certainly support an employer's legal right without approving of every case in which an employer exercises that right, just as I support a citizen's constitutional right to free speech without approving of every instance in which the citizen uses that right.

Weblogs, thus, resemble an informal academic discussion among a university professor and students. As the example illustrates (Text 6), they contain many informal features imitating spoken conversation (*well*, *who've*, *shouldn't, aaaaah..., it's, doesn't, don't, I'm*) while the discussed topics are serious and important. In addition, both the professor and the student (whose text is presented in italics in the weblog) use rare words (e.g. *ridiculous,*

eccentric, outlandish), academic vocabulary (*to generate, including, voicing, opinions, performing, genuinely, obviously, repeatedly, approving, instance*) and formal expressions *(is supposed to, to express themselves, whereas, exercises that right*). This combination of formal, academic and informal, conversations features in writing mediated by computer is a characteristic feature of academic weblogs a new type of CMAD.

Academic hypertexts

The type of discourse of academic hypertexts is defined as transactional written computer-mediated academic discourse. Academic hypertexts have been found in the present study possessing the following specific characteristics:

1. Highly informational
2. Neither narrative nor non-narrative type of discourse
3. Very explicit in giving reference
4. No overt expression of persuasion
5. A very abstract style

In fact, academic hypertexts have been found the most informational in the type of information production of all the other types of CMAD. They are also the most abstract, similarly to academic weblogs. The following linguistic features have been identified as the most frequent in on-line seminars: passive voice constrictions, attributive adjectives, past and present participle clauses, nominalisations, nouns, phrasal coordination, adverbials, conjunctions, present tense and perfect aspect verbs. Mean syntactic length and mean word length are found to be the largest among all the studied CMAD types.

Rarely occurring linguistic features in academic hypertexts are contractions, first, second and third person pronouns, indefinite pronouns, Wh-questions, private verbs, time adverbials, demonstratives, general emphatics, discourse particles, amplifiers, prediction modals, conditionals and *be* as a main verb.

That academic hypertexts have these characteristics can be illustrated by the following example (Text 6):

Text 6

Although game play can be used as an educational experience, we are more interested in game design than game play, as this is a relatively new endeavor. The focus of the research described in this paper is to use interactive storywriting as a new vehicle for creative expression. In a traditional story, the world is created with words, using descriptive prose, and the story is told with words, through narrative prose. In an interactive story, the world is "painted" with a computer-aided design tool and the story is told dynamically as the PC navigates through the world. BioWare provides the Aurora toolset along with the NWN game system, so that amateur writers (including students) can paint the background for their stories. There are three potential benefits of interactive storywriting: first, students can improve the skills necessary to effectively use an increasingly important communications medium; second,

they will learn important logical thinking skills, similar to computer programming, but in an environment that does not have the stigma of computer programming; third, this new communications medium provides an alternative mechanism for creative expression that may allow students to improve their expressive skills.

Academic hypertexts are evolving from the genre of traditional academic articles and have many features resembling them. As they are usually written with great care, with several drafts followed by editing, academic hypertexts usually do not contain spelling and grammar mistakes. The authors spend much time on careful structuring and reviewing the text, which is, as a result, has high informational density and lexical richness.

On-line seminars

The type of discourse in on-line seminars in characterised as synchronous transactional spoken computer-mediated academic discourse.

On-line academic seminars have been found in the present study to possess the following specific characteristics:

1. Informational rather than involved information production
2. Non-narrative type of discourse
3. Moderate degree of explicitness in giving reference
4. Overt expression of persuasion
5. Rather non-abstract style

The most frequent linguistic features in on-line seminars are prediction modals, public verbs, Wh- and that relatives, Wh-clauses, demonstratives, pronoun *it*, suasive verbs, place adverbials, third-person pronouns, hedges, conditionals. The linguistic features that are rare in on-line seminars are the following: perfect aspect verbs, general emphatics, attributive adjectives, past tense verbs, adverbs and adverbials, that deletion, suasive verbs, amplifiers, passives, wh-questions, present tense verbs (Appendix 16). This combination of frequent and rare linguistic features characterise this type of CMAD as similar to the traditional oral academic discourse, i.e. university lectures and seminars, described in 1.1.2 in the present study. However, the on-line seminars have also specific characteristics that are connected with the use of computer as communication medium. The main difference is that situational reference is expressed less explicitly at on-line seminars than at face-to-face seminars as illustrated by (e.g. *the slide that follows, the one that is on the screen right now, the column that's on the right)*, although at times explicit reference is gives by the use of place and time adverbials (*today, right now*) and demonstratives *(this, these).* That seminars have these characteristics can be illustrated by the following extracts from automatically recorded seminar presentations (Text 7):

Text 7

THIS PRESENTATION WILL FOCUS ON LAST YEAR'S WINNERS, AND NEXT YEAR -- OR THIS YEAR'S WINNERS WILL ACTUALLY BE SHOWCASED WITHIN THE NEXT COUPLE OF WEEKS.
SO **WE** WOULD LOVE TO SEE **YOU** THERE.
AND **I** AM GOING TO TURN IT OVER TO DAVID.
WHAT **WE** ARE GOING TO DO TODAY IS TALK A LITTLE BIT ABOUT WHAT THESE COURSES HAVE BEEN AND THE ONES **WE** HAVE SELECTED AS BEING EXEMPLARY.
AS **YOU** CAN SEE IN THE SLIDE THAT FOLLOWS, THE ONE THAT IS ON THE SCREEN RIGHT NOW, **THEY** HAVE COME FROM A VERY WIDE RANGE OF DISCIPLINES.
WE HAVE A MIX OF -- AND BOTH IN THE NOMINATIONS AND IN THE COURSES **WE** SELECTED OVER THE FOUR YEARS.
COURSES THAT ARE FULLY ON-LINE AND OTHERS THAT ARE BLENDED LEARNING ENVIRONMENTS.
AND **THEY**'VE COME FROM EVERYWHERE, COLLEGES, UNIVERSITIES, PROFESSIONAL DEVELOPMENT INSTITUTES OR ORGANIZATIONS, EVEN HIGH SCHOOLS.AND FROM THE UNITED STATES, AUSTRALIA, CANADA, UNITED KINGDOM.
THAT'S WHERE THE SELECTED COURSES COME FROM.
TAKE A LOOK AT SOME STATS.
THE COURSES THAT **WE** -- NOW THESE ARE COURSES THAT HAVE WON OVER THE LAST FOUR YEARS, INCLUDING THIS YEAR'S COURSES.
SO **YOU** LOOK AT THAT, AND YOU SAY, WELL, OKAY, THERE'S BEEN THREE FROM AUSTRALIA AND NINE FROM CANADA, ONE FROM THE UK AND 35 FROM THE U.S. AND DIFFERENT STATES.
WHAT'S INTERESTING FOR **ME**, THOUGH, ASIDE FROM THAT BREAKDOWN IS THE COLUMN THAT'S ON THE RIGHT.
THE VARIOUS DISCIPLINES THAT HAVE BEEN CHOSEN WHILE REPRESENTED BY THE SELECTED COURSES, **I** SHOULD SAY.
EDUCATION, HEALTH, DISTANCE-BASED COURSES.
THESE ARE VERY, VERY GENERAL CATEGORIES WE HAVE CREATED TO SUMMARIZE A VIEW OF WHERE THEY ARE COMING FROM.

That the presentation is given for academic purposes can be illustrated by the use of academic vocabulary (*presentation, courses, exemplary, disciplines, nominations, professional, general categories, to summarize, development*), special expressions (*this presentation will focus, that follows, be showcased, have selected as being exemplary, a very wide range of*) and a well-planned structure of the text. However, it contains many linguistic features that are typical for spoken communication, for example, contractions (*they've, that's, what's*), informal abbreviations (*stats*), discourse markers (*well, okay*) etc. On-line seminar speech is also characterized by a rather high number of personal person pronouns as illustrated by Text 7 and Text 8.

Text 8

JUST BECAUSE SOMETIMES THAT HELPS TO UNDERSTAND HOW **YOU** COMPARE WITH OTHER INSTITUTIONS WHEN **YOU** THINK ABOUT HOW **THEY** DEPLOY A PRODUCT OR HOW **THEY** MIGRATE OR WHAT **THEIR** STRATEGY MIGHT BE FOR MIGRATING A PRODUCT.

Text 9 on Page 130 is a further illustration of the specific features on on-line seminars. Very specific lexis (*activate, close captioning, type forward slash, message interface, chat box*) is typical for computer discourse. In on-line academic seminars, this lexis is a part of discourse interwoven into the presentation on the subject matter.

Text 9

FOR TODAY'S EVENT.
TO ACTIVATE CLOSED CAPTIONING,IT'S VERY STRAIGHTFORWARD.
JUST SIMPLY TYPE FORWARD SLASH, CC INTO THE MESSAGE
INTERFACE IN THE CHAT BOX. AND TURN IT OFF BY DOING
EXACTLY THE SAME THING.

Thus, on-line seminars are a new developing type of computer-mediated academic discourse that demonstrates the characteristic features of a traditional academic seminar and new specific features that result from the use of computer medium.

To summarise, in the course of the present research, the empirical data on linguistic variation among various types of CMAD have been collected and analysed. In the present study, the author has measured linguistic variation in CMAD and applied statistical research instruments for quantitative evaluation of the measurements in relation to the proposed hypothesis. The frequencies of occurrence of linguistic features have been measured and presented in tables and figures. The most frequent characteristic features for each type of CMAD have been identified. The statistical procedures have been applied to reveal the multidimensional variation across text types in CMAD. Thus, the author has obtained objective statistical data on the differences in linguistic characteristics among CMAD types on five functional dimensions. The triangulation of the data and research instruments has been conducted by applying corresponding statistical procedures and tests. Thus, a sufficient level of the internal research validity has been achieved. The differences along five textual dimensions in the frequency of occurrence of linguistic features among the six specialised corpora investigated in the present research reveal the multidimensional linguistic variation in the English language use across the types of CMAD investigated in the present study.

CONCLUSIONS AND FURTHER RESEARCH PERSPECTIVES

The dissertation has presented the undertaken theoretical and empirical research in which linguistic variation across text types in computer-mediated academic discourse was revealed and studied. On the basis of the conducted research work the author draws the following conclusions.

1. A new specific type of English discourse that takes place in academic settings and is mediated by computer – *computer-mediated academic discourse* (CMAD) – has been distinguished and conceptualised. This particular type of English discourse had not been defined and described in applied linguistics before the present study. Therefore, an original definition of CMAD was elaborated as the result of thorough theoretical research. CMAD is defined as a computer-mediated process of functional use of language as a means of communication in academic context. It is realised in semantically connected and meaningful to the communicating users of language verbal instances of spoken or written language longer than a sentence. The proposed definition provides a foundation for further theoretical and empirical research.
2. The conducted theoretical and empirical investigation has demonstrated that CMAD has specific linguistic characteristics distinguishing it from other types of English discourse. The compiled for the present study 1, 000, 000-word corpus of CMAD comprised not previously studied authentic texts that were collected in the international on-line academic community of education professionals in 2003-2007. The undertaken empirical analysis has shown that the corpus, as a whole, has specific statistical characteristics that distinguish it from the corpora used in the previous linguistic research in the field. Considerable changes have been revealed in the use of fifty-five types of linguistic features in the corpus, e.g. pronouns, short forms, discourse markers etc. The research shows that CMAD differs both from the corpus of General English studied by Biber in 1988 and from academic discourse in the corpora that were used in the studies of the academic language use in university in traditional, i.e. not mediated by computer, communication (Biber, 2003; Biber et al., 2004). The found differences are attributed to the use of computer medium for spoken and written academic communication, which is important to consider in computer-assisted language learning and teaching at university.
3. On the basis of conducted analysis of contextual factors influencing CMAD as a process, an original typological classification of CMAD has been proposed. It theoretically models how different combinations of factors result in different types of CMAD. The following six most typical types of CMAD have been distinguished: academic e-mails, synchronous

conferencing, on-line discussions, weblogs, hypertexts and computer-mediated seminars. As has been proved in the dissertation, each type of CMAD is the result of a unique combination of transactional or interactional type of discourse, synchronous or asynchronous mode of interaction, spoken or written mode of discourse and the type of software used in communication (e.g. e-mail, discussion forum, weblog etc.). Significant linguistic variation has been found among the six specialised sub-corpora representing the types of CMAD. Thus, the description and analysis of the linguistic characteristics of each type of CMAD provided in the dissertation may be used for further applied linguistic research and practical language teaching and learning.

4. In the course of the undertaken research, the linguistic variation in CMAD has been discovered and empirically investigated. Its causes, mechanisms and types have been analysed in the dissertation. The linguistic variation in computer-mediated academic discourse is defined as a process in which communicators vary the use of academic language to match the specific situational context in computer mediated communication. It is realized, as a product, in quantifiable differences in the frequency of occurrence of linguistic features in CMAD text types, for example, the frequency of personal pronouns, passive voice constructions, etc. The frequency of occurrence of fifty-five linguistic features in each text in the corpus of CMAD was measured and analysed in the present study for their further comparison and interpretation.
5. In result of applying Biber's (1988) quantitative multidimensional statistical analysis to investigation of linguistic variation in CMAD, strong patterns of co-occurrence of linguistic features have been discovered in the compiled CMAD corpus. They are regarded as different underlying functional dimensions along which the variation in CMAD occurs and the types of CMAD differ from each other. On the basis of the conducted Factor Analysis statistical procedure, the author proposes a five-dimensional model of linguistic variation in CMAD. Applied to the corpus of CMAD, compiled for the present study, the model allowed the author to reveal statistically significant differences among six types of CMAD (academic e-mails, synchronous conferencing, on-line discussions, weblogs, hypertexts and computer-mediated seminars) along the following five functional dimensions determined by Biber (1988): Dimension 1 (*Involved/ Informational production),* Dimension 2 (*Narrative/ Non-narrative concerns),* Dimension 3 (*Explicit/ Situation-dependent reference),* Dimension 4 (*Overt expression of persuasion),* Dimension 5 (*Abstract/ Non-abstract information).*
6. Contrary to the traditional view, which regards academic discourse mediated by computer not significantly varying, the applied statistical method with a 95% level of confidence

provided the evidence that the type of CMAD is a possible reason for the variation in the frequency of co-occurrence of linguistic features in the text samples. The conducted triangulation of data and instruments gave similar results on linguistic variation in CMAD. The results of Scheffé's test demonstrate that there is a significant statistical difference between at least one pair of the mean values on each dimension. This means that the studied types of CMAD are rather similar on one dimension but different on another dimension. For example, while academic weblogs differ from all other types of CMAD on Dimensions 2 *(Narrative/ Non-narrative concerns*) and Dimension *3 (Explicit/ Situation-dependent reference*), they are similar to all but one (hypertexts) types of CMAD on Dimension *1 (Involved/ Informational production*) and to hypertexts on Dimension *4 (Overt expression of persuasion*) and Dimensions 5 *(Abstract/ Non-abstract information*). Thus, the findings obtained in the research provide the evidence of significant multidimensional variation in the frequency of appearance of linguistic features in different types of CMAD.

7. The results of the comparison of the findings obtained in the present study with the results of the previous research (Biber, 1988) show that none of the investigated CMAD types has the same linguistic characteristics on all five dimensions as any other type of English discourse. Each type of CMAD resembles some types of English discourse on one dimension but different types on other dimensions. These findings support the conclusion that the variation among CMAD types is multidimensional and that each type of CMAD has specific linguistic characteristics distinguishing it form other types of English discourse and other types of CMAD. Thus, the comparison of the results confirms that there are significant differences among the types of CMAD and between CMAD and other types of English discourse. To demonstrate the differences, the author offers a description of the distinguished linguistic characteristics of each CMAD type in the dissertation.

The obtained results and drawn conclusions allow the author to evaluate the results of the present research as positive. All the goals and objectives have been reached: English computer-mediated academic discourse has been studied theoretically and empirically and the objective statistical data on the variation in CMAD has been obtained. The research hypothesis has been proved: a significant variation in linguistic characteristics across text types in English computer-mediated academic discourse, reflecting the differences in the functional use of linguistic features in its different types, has been found. Each type of CMAD was described by a number of specific linguistic features that are characteristic to it and a number of linguistic features that are rare in it. The classification and description of CMAD

types is provided in the dissertation for further investigating and teaching the grammar and lexis of each type of CMAD and the differences between them. Thus, the findings obtained in the present study have theoretical and practical value as a methodological basis for further linguistic research and practical language teaching.

The completed research may be further extended in breadth and depth, as well as may serve as a basis for further research into neighbouring areas. Further research on a large specialised machine-readable corpus of CMAD is necessary to secure further generalisation on the findings of the present study. Although the present study has the external validity limited to only the types of CMAD represented in the corpus, the theoretical model of linguistic variation in CMAD allows to predict the linguistic characteristics of the types of computer-mediated discourse that may appear in future. Thus, it may be used as a methodological basis for further research. Future studies might seek to identify other existing types of CMAD that have not been studied, e.g. video- and audio-conferencing with the use of such programs as, for example, *Scype*, as well as other types of professional discourse for applied linguistics' purposes. The types of CMAD that are experiencing significant changes (e.g. merging genres) and new types emerging along with the development of technology need investigation and description.

The completed quantitative study was limited to only verbal linguistic features because of the difficulty of encoding and automatic processing of non-verbal and visual features specific for CMAD. Further research is needed to cover this unstudied research area.

The study applied the dimensions (and the linguistic features in them) that were determined by Biber (1988) for general corpus of the English language while specialised corpora complied for the present study represented computer-mediated academic language. This may imply the existence of some specific to CMAD textual dimensions that have not been researched. To reveal them compiling and computer-processing a much larger machine-readable corpus of CMAD texts is necessary.

Finally, the most important area of further research is the application of the findings to language teaching and learning. The linguistic characteristics of CMAD types determined in the present research, i.e. the lexicogrammatical features that have been identified as specific to each type of CMAD, may be used as specifications for the development and assessment of discourse competence for academic computer-mediated communication and creating the materials for technologically-equipped language teaching and learning. Further studies are necessary to investigate the sub-genres of each CMAD type (Fig. 2.14 on page 136) for educational purposes.

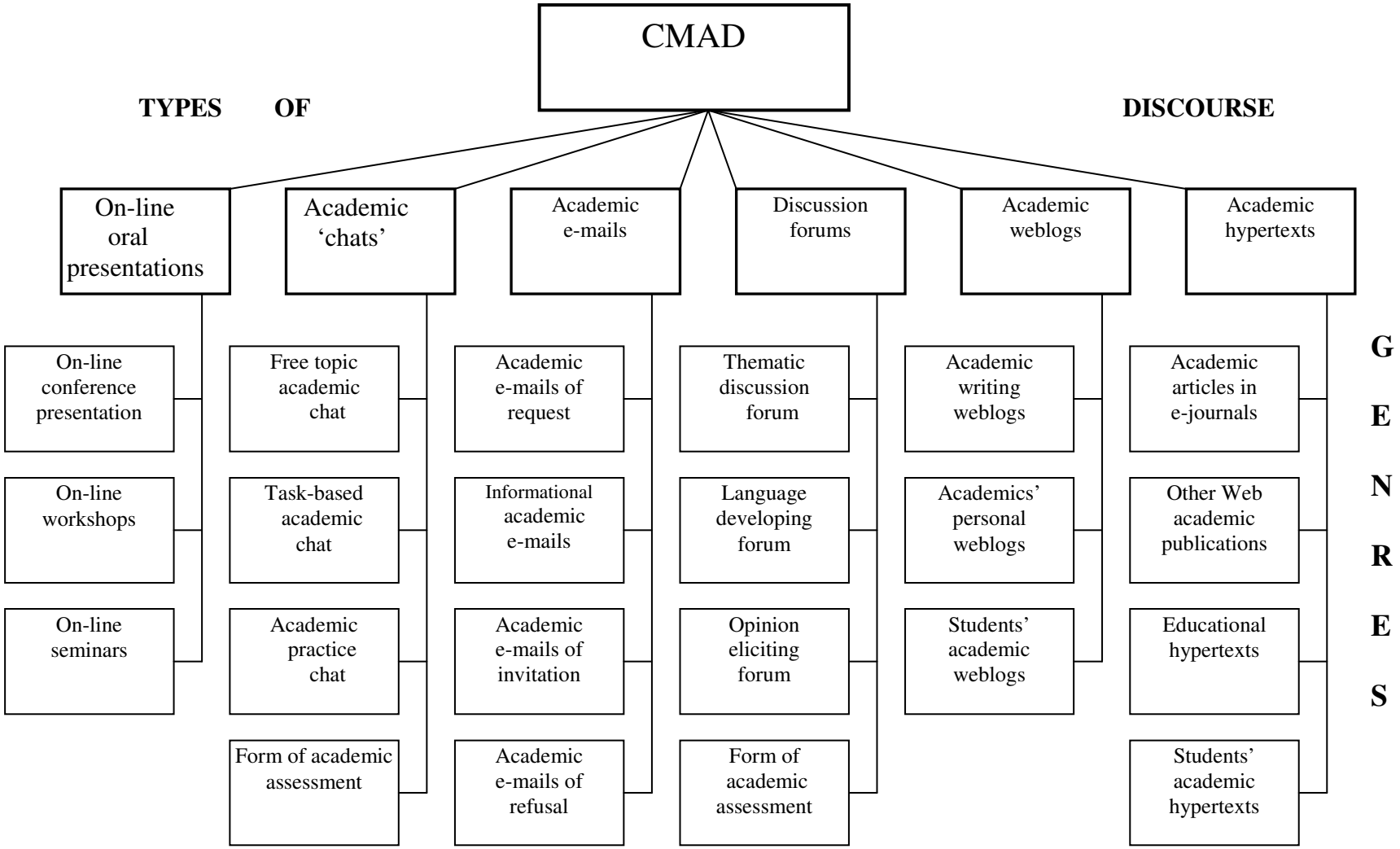

Figure 2.14 Taxonomy of CMAD types and genres recommended for further research.

BIBLIOGRAPHY

1. Abrams, Z. I. (2003) The effect of synchronous and asynchronous CMC on oral performance in German. *The Modern Language Journal, 87*(2), 157–167.
2. Aitchinson, J., and Lewis, D.M. (Eds.) (2003) *New Media Language* London: Routledge.
3. Ancarno, C. (2005) The style of academic e-mails and conventional letters: contrastive analysis of four conversation routines. *Ibérica,* Vol. 15: 103-122.
4. Androutsopoulos, J. (2006) Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics*, Vol. 10 (4): 419.
5. Bakhtin, M.M. (1986/1999) The Problem of Speech Genres. In Jaworski, A., Coupland, N. (1999) *The Discourse Reader* London, New York: Routledge, pp. 121-132.
6. Balestri, D. P. (1988) Softcopy and Hard: Word processing and Writing Process. *Academic Computer*, Vol. 2(5): 14-17, 41-45.
7. Bārzdiņš, Guntis. (2006) Latviešu valoda semantiskajā tīmeklī / Latvijas Zinātņu Akadēmijas Vēstis. A daļa, Sociālās un humanitārās zinātnes. - 60. sēj., Nr.6 (2006), 26.-42.lpp.
8. Barnes, S. (2003) *Computer-Mediated Communication: Human-to-Human Communication across the Internet,* Allyn and Bacon Publishers.
9. Baron, N. S. (1984) Computer-mediated communication as a force in Language change. *Visible Language,* Vol. 8 (2): 118-141.
10. Baron, N. S. (1998) Letters by phone or speech by other means; the linguistics of e-mail. *Language and Communication* Vol. 18: 133-170, Pergamon Press.
11. Baron, N. S. (2000) *Alphabet to Email: how written English evolved and where it's heading.* London: Routledge.
12. Baron, N. S. (2003) Why e-mail looks like speech: proofreading, pedagogy, and public face. In Aitchinson, J. and D. Dawis (eds.) *New Media Language* 2003: 102-113, London: Routledge.
13. Baron, N. S. (2005) The future of written culture: envisioning language in the new millennium. *Iberica,* Vol. 9: 7-31.
14. Bazerman, C. and P. Prior (Eds.) (2004) *What Writing does and How it Does It: An Introduction to Analysing Texts and Textual Practices.* London: Lawrence Erlbaum Associates Publishers.
15. Beavers, J. (1998) Essay on Hypertext and Academic Writing. [On-line] *Electronic Publications and Virtual Exhibitions in American Studies*. 1998. Retrieved on 23 April, 2007 from http://www.otal.umd.edu/~deruyver/amst418p/jbeavers/a7.htm
16. Belz, J., Vyatkina, N. (2005) Learner Corpus Analysis and the Development of L2 Pragmatic Competence in Networked Inter-cultural Language Study: The Case of German Modal Particles. *Canadian Modern Language Review* Vol. 62 (1): 17-48.
17. Belz, J., and Reinhardt, J. (2004) Aspects of advanced foreign language proficiency: Internet-mediated German language play. *International Journal of Applied Linguistics,* 14(3): 324-362, Blackwell.
18. Belz, J. (2003) Linguistic perspectives on the development of intercultural competence in telecollaboration. *Language Learning and Technology,* Vol. 7 (2): 68-117.

19. Berkenkotter, C., and Huckin, T. N. (1995) *Genre knowledge in disciplinary communication: Cognition, Culture, Power.* Hillsdale, N.J.: Erlbaum.

20. Beuchot, A., Bullen, M (2005) Interaction and interpersonality in on-line discussion forums. *Distance Education,* Vol. 26 (1): 67-87.

21. Bhatia, V.K. (2004) *Worlds of Written Discourse.* Continuum.

22. Bhatia, V. K. (1993) *Analyzing genre: Language use in professional settings*. New York: Longman.

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

24. Biber, D. (2003) Variation among University Spoken and Written Registers: A Multidimensional Analysis. In Meyer, C.F. (Ed.) *Corpus Analysis: Language Structure and Language Use*. Rodopi.

25. Biber, D., and Kurjian. J. (2007) Towards a taxonomy of web registers and text types: A multi-dimensional analysis. In M. Hundt, N. Nesselhauf, and C. Biewer (Eds.), *Corpus Linguistics and the Web*, pp. 109-132. Amsterdam: Rodopi.

26. Biber, D. (2004) Lexical bundles in academic speech and writing. In Lewandowska-Tomaszczyk B. (Ed.) *Practical Applications in Language and Computers* (PALC 2003): 165-178. Frankfurt on Main: Peter Lang.

27. Biber et al. (2004) Representing Language Use in the University: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus. Princeton, New Jersey: Educational Testing Service.

28. Biber, D., and Conrad, S. (2001) Register variation: A corpus approach. In Schiffrin, D., Tannen, D. and H. Hamilton (eds.), *The handbook of discourse analysis*, 175-96. Oxford: Blackwell.

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

30. Biber, D. (1988) *Variation across speech and writing.* Cambridge University Press.

31. Boardman, M. (2005) *The Language of Websites.* Routledge.

32. Bolter, J. D. (1998) Hypertext and the Question of Visual Literacy. In Reinking, D., Mckenna, M.C., Labbo, L.D., Kieffer, R.D. (Eds.) *Handbook of Literacy and Technology: Transformations in Post-Typographic World.* Mahwah, NJ: Lawrence Erlbaum.

33. Bolter, J. D. (1991) *Writing Space: The Computer, Hypertext, and the History of Writing.* Hillsdale, NJ: Lawrence Erlbaum Associates.

34. J.Borzovs, G.Fricnovičs, A.Spektors (1997) The Latvian Language and IT&T Information Technologies and Telecommunications in the Baltic States : Intern.Conf.&Exhib. ; Abstracts of papers - Riga, 1997: 35-39.

35. Brēde, M. (2004) Discourse markers as contextualization cues. *Sastatāmā un lietišķā valodniecība.* Rīga: LU Press, 16-28.

36. Bretag, T. (2006) Using email to explore new subject positions for teachers and additional language learners, in Cadman, K. And O'Regan, K. (Eds.) *Tales out of School: Identity and English Language Learning*, University of Adelaide.

37. Broady, E. (2000) Second Language Writing in a Computer Environment: A different kind of writing? A different kind of teaching? In Broady, E. (Ed.) *Second Language Writing in a Computer Environment.* London: CILT, 1-25.

38. Brown, G. and Yule, G. (1983) *Discourse analysis*. Cambridge: Cambridge University Press.

39. Brown, P. and Levinson, S. (1978) *Universals in language usage: Politeness phenomena.*

40. Burbules, N.C. (1998) Rhetorics of the Web: Hyperreading and Critical Literacy. In I. Snyder (Ed.) Page to Screen: Taking Literacy Into the Electronic Era. New South Wales: Allen and Unwin.

41. Bussmann, H. (1996) *Routledge Dictionary of Language and Linguistics*. London: Routledge.

42. Bušs, O. (2003) *Personvārdi, vietvārdi un citi vārdi: izpētes pakāpieni*. Rīga: LU.

43. Carr-Chellman, A., and Duchastel, P. (2000) The ideal online course. *British Journal of Educational Technology,* 31(3): 229-241.

44. Carter, R. (2004) *Language and Creativity*. London: Routledge.

45. Carter, R. (2001) Ten criteria for spoken grammar. Plenary presentation at AESLA XX, Leon, Spain, as cited in Swales, J. (2004) *Research Genres*. Cambridge: Cambridge University Press.

46. Carter, R. (1997) *Investigating English Discourse: Language, literacy and literature* London, New York: Routledge.

47. Cazden, C. (1988) Classroom discourse: The language of teaching and learning. Portsmouth, NH: Heinemann.

48. Celce-Murcia, M. and Olshtain, E. (2005) Discourse-Based Approaches: A New Framework for Second Language Teaching and Learning. In Hinkel, E. (Ed.) *Handbook of Research in Second Language Teaching and Learning* London: Lawrence Erlbaum, 729-741.

49. Chafe, W. (2001) The Analysis of Discourse Flow. In Schiffrin, D., Tannen, D., Hamilton, H.E. (Eds.) The Handbook of Discourse Analysis Blackwell Publishing, 673-687.

50. Chafe, W. and Danielewicz, J. (1987) *Properties of Spoken and Written Language* Accessed on 28 August, 2007, http://www.nwp.org/cs/public/print/resource/591

51. Charney, D. (1994) The Effect of Hypertext on Processes of Reading and Writing. In Selfe, C. and Hilligass, S. (Eds.) *Literacy and Computers: The Complications of Teaching and Learning with Technology.* New York: The Modern Language Association.

52. Chen, C.E. (2006) E-mail literacy: from writing to peers to writing to authority figures. *Language Learning and Technology,* Vol. 10 (2): 35-55.

53. Chih-Hua, K. (1999) The Use of Personal Pronouns: Role Relationships in Scientific Journal Articles, *English for Specific Purposes,* 18 (2): 121-138.

54. Chun, D. (1998) Using computer networking to facilitate the acquisition of interactive competence. *System,* 22 (1): 17-31.

55. Collot, M., Belmore, N. (1993) Electronic Language: A New Variety of English. In Aarts et al. (Ed.) *English Language Corpora: Design, Analysis and Exploitation,* Amsterdam: Rodopi, 41-55.

56. Conrad, S. (2001) Variation among disciplinary texts: A comparison of textbooks and journal articles in biology and history. In Conrad, S. and D. Biber (Eds.) *Multi-dimensional Studies of Register Variation in English,* Harlow: Pearson Education,94-107.

57. Coupland and Jaworski (2001) *The Discourse Reader,* London, New York: Routledge.

58. Coxhead, A. (2000) A new Academic Word List. *TESOL Quarterly* 34 (2): 213-238.

59. Christie, F., and Martin, J. R. (1997). *Genre and institutions: Social processes in the workplace and school.* London: Cassell.

60. Crystal, D. (2001) *Language and the Internet.* Cambridge University Press.

61. Csomay, E. (2007) Vocabulary-based discourse units in university class sessions. In Biber, D., Connor, U. & T. Upton. *Discourse on the Move.* Amsterdam and Philadelphia: John Benjamins, 213-238.

62. Csomay, E. (2006) Academic talk in American university classrooms: crossing the boundaries of oral-literate discourse? *Journal of English for Academic Purposes,* Vol. 5 (2): 117-135.

63. Csomay, E. (2004) Linguistic variation within university classroom talk: A corpus-based perspective. *Linguistics and Education,* Vol. 15 (3): 243-274.

64. Cummins, J. (2000) Academic Language Learning, Transformative Pedagogy, and Information Technology: Towards a Critical Balance. In *TESOL Quarterly,* Vol. 34 (3): 537-548.

65. Davis, B. and Brewer, J. (1997) *Electronic Discourse: Linguistic Individuals in Virtual Space.* State University of New York Press.

66. Davis, B., and Thiede, R. (2001) Writing into change: Style shifting in asynchronous electronic discourse. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching: Concepts and practice*. Cambridge: Cambridge University Press, 87-120.

67. Devitt, A. J. (1993) Generalizing genre: new conceptions of an old concept. *College Composition and Communication,* Vol. 44: 573-586.

68. van Dijk, T. A. (1997) *Discourse as structure and process.* Vol. 1, *Discourse as Social Interaction.* Vol. 2. London: Sage.

69. Dowling, C. (1998). Academic discourse in the age of the Internet. In McBeath, C. and R. Atkinson (Eds.) *Planning for Progress, Partnership and Profit*. Proceedings EdTech'98. Perth: Australian Society for Educational Technology. Accessed on 23rd August, 2007, http://www.aset.org.au/confs/edtech98/pubs/articles/dowling.html

70. Druviete, I. (2008) *Sociolingvistikas vēsture un perspektīvas pasaulē un Latvijā*. Riga: LU Language Institute, Vol. 17.

71. Dudley-Evans, A. (1994) Genre analysis: an approach to text analysis for ESP. In Coulthard, M. (Ed.) *Advances in Written Text Analysis*. London: Routledge. 219-228

72. Edminster, J., Moxley, J. (2002) Graduate education and the evolving genre of electronic theses and dissertations. *Computers and Composition,* Vol. 19: 89-104.

73. Efimova, L., & Fiedler, S. (2004) Learning webs: learning in weblog networks. in P. Kommers, P. Isaias, and M.b.Nunes (Eds.), *Proceedings of the IADIS International Conference Web Based Communities.* Lisbon, Portugal,IADIS Press, 490–494.

74. Eggins, S. (2004) *Introduction to Systemic Functional Linguistics.* Continuum.

75. Eggins, S. and Martin, J.R. (1997) Genres and Registers of Discourse. In van Dijk, T. A. (1997). *Discourse as structure and process: Discourse Studies: A Multidisciplinary Introduction.* Vol. 1: 230-256.

76. Elbow, P., (1998) Reflections on Academic Discourse. In Zamel,V., and Spack, R. *Negotiating Academic Literacies.*

77. Emmerson, P. (2004) *Email English*. Oxford: Macmillan.

78. Fairclough, N. (1993) Critical discourse analysis and the marketization of public discourse: The universities. *Discourse and Society, 4*(2): 133-168.

79. Fairclough, N. (2000) *Language and power* (2nd ed.). New York: Longman.

80. Ferrara, K., Brunner, H., Whittemore, G. (1991) Interactive written discourse as an emerging register. *Written Communication*, Vol. 8 (1): 8-34.

81. Ferris, P. (1997) Writing in Syberspace. *Computer-mediated Communication* Vol. 4(6) Accessed on 23rd August, 2007. http://www.december.com/cmc/mag/1997/jun/ferris.html

82. Firth, J. (1957) A synopsis of linguistic theory 1930-1955. In *Studies in Linguistic* Analysis. Oxford: Philological Society, 1-32. Reprinted in F. Palmer (Ed.) (1968) Selected Papers of J.R. Firth. Harlow: Longman.

83. Flowerdew, J. (2001) *Academic Discourse.* London: Longman.

84. Flowerdew, J. (Ed.), (1994) *Academic listening: Research perspectives*. Cambridge: Cambridge University Press.

85. Foucault, M. (1972/2000). *The essential works of Foucault (Volume 3, Power).* New York: The New Press.

86. Fox, R. (2003) Digital Culture and Educational Practice. *Web Proceedings of 2nd International Conference on Knowledge and Discourse*, Hong Kong.

87. Gee, J.P. (1996) *Social Linguistics and Literacies: Ideology in Discourses.* Routledge.

88. Giménez-Moreno, R. (2006) A new approach to register variation: the missing link. *Ibėrica* Vol. 12: 89-109.

89. Gimenez, J.C. (2000) Business e-mail communication: some emerging tendencies in register. *English for Specific Purposes* Vol. 19: 237-251.

90. Givón, T. (1995) *Functionalism and Grammar.* Amsterdam/Philadelphia: John Benjamins Publishing.

91. Gonzalez, D. (2003) Teaching and learning though chat: A taxonomy of educational chat for EFL/ESL. *Teaching English with Technology*, vol. 3 (4). Accessed on 23rd August, 2007. http://www.iatefl.org.pl/call/callnl.htm

92. Grabe, W., and Kaplan, R. (1997) On the writing of science and the science of writing: hedging in science texts and elsewhere. In Markkanen, r. and Schröder, H. (Eds.) *Hedging in Discourse.* Berlin: Walter de Gruyter, 151-167.

93. Grabe, W. and R. B. Kaplan (1996) *Theory and Practice of Writing*. London: Longman

94. Greenhill, A. and Fletcher, G. (2003) Eds. *Electronic References & Scholarly Citations of Internet Sources.* [on-line] Retrieved on 20 March, 2006 from: http://www.spaceless.com/WWWVL

95. Gries, S.T. (2003) Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement. New York, London: Continuum.

96. Gutiérrez, K. (1995) Unpackaging academic discourse. In *Discourse Processes*, 19 (1): 21-37.

97. Hackbarth, S. L. (1996) *The educational technology handbook: A comprehensive guide: Process and products for learning* Englewood Cliffs, NJ: Educational Technology Publications.

98. Hale, C. (1996) *Wired Style: Principles of English Usage in the Digital Age.* San Francisco: HardWired.

99. Halliday, M.A.K. (1973) *Explorations in the Functions of Language*, London: Edward Arnold.

100. Halliday, M. A. K., and Hasan, R. (1976) *Cohesion in English*. London: Longman.

101. Halliday, M.A.K. (1978) *Language as Social Semiotic: the sociological interpretation of language and meaning.* London: Edward Arnold.

102. Halliday, M.A.K. (1985) *Spoken and Written Language*, Victoria : Deakin University.

103. Halliday, M.A.K. (1988) On the language of physical science. In Ghadessey, M. (Ed.) Registers of Written English. London: Pinter, 162- 178.

104. Halliday, M.A.K. and Hassan, R. (1990) *Language Context and Text.* London: Oxford University Press.

105. Halliday, M.A.K., and Martin, J.R. (1993) *Writing Science: Literacy and Discourse Power*. London: Falmer Press.

106. Halliday, M.A.K. and C.M.I. Matthiessen (2004) *An Introduction to Functional Grammar*. London: Hodder Arnold.

107. Hampel, R. (2003) Theoretical perspectives and new practices in audio-graphic conferencing for language learning. *ReCall*, Vol. 15 (1): 21-36.

108. Hampel, R. and Hauck, M. (2003) Using Lyceum, an audio-graphic conferencing system, to talk at a distance. In Goodfellow D., Fenner A.-B., Garrido C. and Tella S. (Eds.) *The educational use of ICT in teacher education and distance language learning*. Graz: European Centre for Modern Languages of the Council of Europe. Available at: http://www.ecml.at/distance_learning/files/tier2_1.htm

109. Hassan, X., Hauger, D., Nye, G., Smith, P. (2005) The use and effectiveness of synchronous audiographic conferencing in modern language teaching and learning (online language tuition): a systematic review of available research. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

110. Harris, J. (1989) The idea of community in the study of writing. College Composition and Communication, 40, 11-22.

111. Harris, Z. S. (1952a) Discourse analysis. *Language* 28, No.1-30 Reprinted in Katz, Jerrold J. and Jerry A. Fodor, eds. (1964) *Readings in the Philosophy of Language*. Englewood Cliffs, NJ: Prentice-Hall.

112. Harris, Z. S. (1952b) Discourse Analysis. *Language,* Vol. 28 (1): 1-30.

113. Hawisher, G. and Selfe, S. (1998) (Eds.) *Passions pedagogies and 21st century technologies*. Logan: Utah State University Press.

114. Herring, S. C. and Paolillo, J. C. (2006) Gender and genre variation in weblogs *Journal of Sociolinguistics.* Vol. 10 (4): 439–459.

115. Herring, S. C., Scheidt, L. A., Bonus, S., and Wright, E. (2004) Bridging the Gap: A Genre Analysis of Weblogs. In Proceedings of the Thirty-seventh Hawaii International Conference on System Sciences (HICSS-37) (Ed.), Los Alamitos: IEEE Press.

116. Herring, S. (2001) Computer-mediated discourse. In Schiffrin, D., Tannen, D., Hamilton, H. E. *The Handbook of Discourse Analysis* Blackwell Publishing.

117. Herring, S.C. (1998) Variability and Change in E-mail Style. The English version of Le style du courrier électronique: variabilité et changement. *Terminogramme* 84-85 (March): 9-16. http://ella.slis.indiana.edu/~herring/term.html

118. Herring, S. (1996) Posting in a Different Voice: gender and ethics in computer-mediated communication, in C. Ess (Ed.) *Philosophical Approaches to Computer-mediated Communication*. Albany: SUNY Press.

119. Ho, J. (2005) Hyperlink obsolescence in scholarly online journals. *Journal of Computer-Mediated Communication, 10* (3), article 15. http://jcmc.indiana.edu/vol10/issue3/ho.html

120. Hoey, M. (2001) *Textual Interaction* London, New York: Routledge

121. Holmes, J. (1990) Hedges and boosters in women's and men's speech. *Language and Communication*, Vol. 10 (3): 185-205.

122. Huang, H. (2002) *Criteria for Evaluating Web-based Hypertext* [On-line] Retrieved on 30 March, 2006 from http://www.eric.ed.gov/ERICDocs/data/ericdocs2/content_storage_01/0000000b/80/27/fb/a5.pdf

123. Huckin, T. (2004) Content Analysis. In Bazerman, C. and P. Prior (Eds.) *What Writing does and How it Does It: An Introduction to Analysing Texts and Textual Practices.* London: Lawrence Erlbaum Associates Publishers.

124. Huffaker, D., and Calvert, S. (2005) Gender, Identity and Language Use in Teenage Blogs. *Journal of Computer-Mediated Communication*, 10(2). http://jcmc.indiana.edu/vol10/issue2/huffaker.html

125. Huffaker, D. (2004) The Educated Blogger: Using Weblogs to Promote Literacy in the Classroom. *First Monday* (9)6, 2004, http://www.davehuffaker.com/blog/?page_id=4.

126. Hunston, S. (1995) A corpus study of some English verbs of attribution. *Functions of Language*, Vol. 2 (2): 133 ff.

127. Hyland, K. (2000) *Disciplinary Discourses*. Pearson Education.

128. Hyland, K. (1998) Hedging in science research articles. Amsterdam: John Benjamins Publishing.

129. Hyland, K. (1994) Hedging in academic writing and EAP textbooks. *English for Specific Purposes,* Vol. 13 (3): 239-256.

130. Hymes, D. (1972) Models of interaction of language and social life. In J. Gumperz, & D. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (p. 35-71). New York: Holt, Rinehart, and Winston.

131. Iļjinska, L. (2007) *English for science and technology* (2nd ed.) Riga : RTU Publ. House.

132. Inman, J. (2004) *Computers and Writing: The Cyborg Era* Lawrence Erlbaum

133. Jackson, H. (2007) *Key Terms in Linguistics*, Continuum.

134. Jordan R. (1999) *Academic Writing Course,* Longman.

135. Jordan, R. *English for Academic Purposes: A Guide and Resource Book for Teachers*, Cambridge: Cambridge University Press.

136. Joyce, M. (1995) *Of Two Minds: Hypertext, Pedagogy and Poetics*. Ann Arbor: University of Michigan Press.

137. Karapetjana, I. (2009) *Linguo-functional research competence: studies in academic discourse.* Riga: RTU Publishing House.

138. Karapetjana, I. (2007) *Language in bachelor papers as a result of the development of linguo-functional research competence,* Riga: LU Press.

139. Kern, R. (1995) Restructuring classroom interaction with networked computers: effects on quantity and characteristics of language production. *Modern Language Journal,* 79: 457-476.

140. Kļaviņa S. *Statistika valodniecībā.* (1980) Riga, LU Press.

141. Ko, K-K. (1996) Structural characteristics of computer-mediated language: A comparative analysis of inter-change discourse. Electronic Journal of Communication, 6(3). Retrieved February 19, 2004, from http://www.cios.org/EJCPUBLIC/006/3/006315.HTML

142. Kramiņa (2000) *Linguo-didactic theories underlying multi-purpose language acquisition*. Riga: LU Press.

143. Kramiņa, I. (2009) Research Guidance: Explored and Explained. Riga: LU Press.

144. Kress, G. (2003) *Literacy in the New Media Age.* London and New York: Routledge.

145. Kress, G., van Leeuwen T. (2001) *Multimodal Discourse. The Modes and Media of Contemporary Communication.* London: Oxford University Press.

146. Labov, W. (1966) Hypercorrection as a factor in linguistic change. In Bright, W. *Sociolinguistics*. Hague: Mouton.

147. Laufer B. and Nation P. (1995) Vocabulary size and use: Lexical Richness in L2 Written Production. *Applied linguistics,* Vol. 16 (3), Oxford University Press.

148. Landow, G. P. (1992) *Hypertext: The convergence of contemporary literary theory and technology.* Baltimore: Johns Hopkins University Press.

149. Lester, J. (1997) *Citing Cyberspace.* New York: Addison, Wesley, Longman.

150. Lewis, W. and Jansen, D. (1997) *Characteristics of Hypermedia Presentations.* Las Vegas, NV: Grand Carousel.

151. Liaw, M. (1998) Using electronic mail for English as a foreign language instruction. *System*, Vol. 26: 335-351.

152. Light, P. and Light, V. (1999) Analysing asynchronous learning interactions: computer mediated communication in a conventional undergraduate setting, in: K. Littleton and P. Light (Eds) *Learning with Computers: analysing productive interactions*, London: Routledgepp, 162-178.

153. Luuka, M.-R. (1998) Language Usage, Conversation, and Electronic Mail. In: *Puolin ja toisin: Suomalais vorolaista kielentutkimusta* (On Both Sides: Finnish-Estonian Research on Language. AFinLA Yearbook 1998.) FL 025 589: 22.

154. Ma, R. (1996) Computer-mediated conversations as a new dimension of intercultural communication between East Asian and North American college students. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*. Philadelphia: John Benjamins, 173-185.

155. Mair, C., and Hundt, M. (1999) 'Agile' and 'Uptight' Genres: The Corpus-Based Approach to Language Change in Progress. *International Journal of Corpus Linguistics,* 4: 221-242.

156. Malinowski, B. (1923) The Problem of Meaning in Primitive Languages. In Ogden, C. and Richards, I. (1989) *The Meaning of Meanings.* Mariner Books, 296-337.

157. Malvern, D., Richards, B., Chipere, N. and P. Durán (2004) *Lexical Diversity and Language Development: Quantification and Assessment.* Palgrave Macmillan.

158. Marchionini, G. (1988) Hypermedia and Learning: Freedom and Chaos. *Educational Technology*, Vol. 28 (11): 8-12.

159. Markwell, J. and Brooks, D. (2002) Broken links: the ephemeral nature of educational WWW hyperlinks. *Journal of Science Education and Technology*, 11 (2): 105-108.

160. Martin, J.R. (2001) Cohesion and Texture. In Schiffrin, D., Tannen, D., Hamilton, H.E. (Eds.) The Handbook of Discourse Analysis Blackwell Publishing, 2001: 35-53.

161. Martin, B. and Ringham, F. (2000) *Dictionary of Semiotics*. New York: Cassell.

162. Maurinen, A. (2002) “One thing I’d like to clarify...”. Observations of academic speaking. *Helsinki English Studies.* The Electronic Journal of the Department of English at the University of Helsinki, Vol. 2.

163. Meyer, C. (2002) *English Corpus Linguistics: An Introduction.* Cambridge: Cambridge University Press.

164. McEnery, T., Xiao, R. and Y. Tono (2006) *Corpus-Based Language Studies: An Advanced Resource Book.* London, New York: Routledge.

165. Miller C. and Shepherd, D. (2004) Blogging as Social Action: A Genre Analysis of the Weblog. In L. Gurak, S. Antonijevic, L. Johnson, C. Ratliff, and J. Reyman (Eds.), *Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs***.** University of Minnesota. Accessed on 23.04.06 http://blog.lib.umn.edu/blogosphere/

166. Miller, J. (2002) Discourse Analysis. *Good Practice Guide* (on-line) Accessed on 23.04.06 http://www.lang.ltsn.ac.uk/resources/goodpractice.aspx?resourceid=132

167. Moore, J. L., and Marra, R. M. (2005) A Comparative Analysis of Online Discussion Participation Protocols. *Journal of Research on Technology in Education* Vol. 38 (2): 191-212.

168. Moran, C., and Hawisher, G.E. (1998) The rhetorics and languages of electronic mail. In Snyder, I. *Page to Screen: taking literacy into the electronic era.* London: Routledge.

169. Morand, D. A., and Ocker, R. J. (2003) Politeness theory and computer-mediated communication: A sociolinguistic approach to analyzing relational messages. *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS-36).* Los Alamitos: IEEE Press.

170. Motteram, G. (2000) Communicating with computers. In Brett, P. and Motteram, G. *A Special Interest in Computers: Learning and Teaching with Information and Communication Technologies (ICT).* IATEFL

171. Moulthrop, S. (1991) Beyond the Electronic Book: A Critique of Hypertext Rhetoric. In *Proceedings of ACM Hypertext '91*, San Antonio, TX, 291-298.

172. Murray, D.E. (2000) Protean Communication: The Language of Computer-Mediated Communication. In *TESOL Quarterly,* Vol. 34 (3): 397-421.

173. Murray, D.E. (2005) Technologies for Second Language Literacy. *Annual Review of Applied Linguistics.* Vol. 25: 188–201.

174. Myers, G. (1989) The pragmatics of politeness in scientific articles. Applied Linguistics Vol. 10 (1): 1- 35.

175. Načisčione (2001) *Phraseological units in discourse: Towards applied stylistics.* Riga: Latvian Academy of Culture.

176. Nation, P. (2001) *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

177. Nation, P. and Coxhead, A. (2002) Range and Frequency: Computer software and manual. Retrieved from http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx

178. Naya, Y. (1999) Effectiveness of e-mail exchange to low incentive EFL students. Paper presented at the 12th *World Congress of Applied Linguistics*, Tokyo.

179. Nelson, M.E. (2006) Mode, meaning, and synaesthesia in multimedia l2 writing. *Language Learning and Technology,* Vol. 10 (2): 56-78.

180. New, W. and Green, K. (2001) En-gendering democracy: a study of on-line academic discourse. *Pedagogy, Culture and Society,* Vol. 9 (2): 187- 220.

181. Nielsen, J. (1990) *Multimedia and Hypertext: The Internet and Beyond.* Cambridge, MA: AP Professional.

182. Nilsson, S. (2004) A brief overview of linguistic aspects of the blogophere, English Department, Umeå University. http://ilyagram.org/media/fetch/blogspeak.pdf

183. Nguyen, H. T., Kellogg, G. (2005) Emergent Identities in On-Line Discussions for Second Language Learning. *Canadian Modern Language Review.* Vol.62 (1): 111-136.

184. Norris, S. (2004) *Analyzing Multimodal Interaction.* New York: Routledge.

185. Oberlander, J. and Nowson, S. (2006) Whose thumb is it anyway? Classifying author personality from weblog text. In Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions, Sydney, Australia, 627-634. http://acl.ldc.upenn.edu/P/P06/P06-2081.pdf

186. Ochs, E., (1990) Indexicality and Socialisation. In Stigler, J. et al. (eds.) *Cultural psychology: Essays on Comparative Human Development,* Cambridge, UK: Cambridge University Press\

187. Ozola (2007) Linguistic and Pragmatic Preconditions to Processing Spoken Discourse in a Foreign Language. Unpublished doctoral dissertation. Riga, University of Latvia.

188. Paltridge, B. (1996) Genre, text type, and, and the language classroom. *ELT Journal,* 50 (3): 237-243.

189. Pellettieri, J. (2001) Negotiation in cyberspace: The role of *chatting* in the development of grammatical competence. In Warschauer, M. Kern, R. *Network-based Language Teaching*. Cambridge University Press, 2001: 59-86.

190. Pelowski, S., Frissell, L., Cabral, K., Yu (2005) So Far But Yet So Close: Student Chat Room Immediacy, Learning, and Performance in an Online Course. *Journal of Interactive Learning Research* Vol.16 (4): 395-407.

191. Phillips, N. and C. Hardy (2002) *Discourse Analysis: Investigating Processes of Social Construction*. Qualitative Research Methods series Vol. 50, London: Sage

192. Phillipson, R. (1992) *Linguistic Imperialism.* Oxford: OUP.

193. Pincas, A. (1998) Successful online course design: Virtual frameworks for discourse construction. *Educational Technology and Society*, 1 (1), available from http://ifets.gmd.de/periodical/

194. Poos, D. and Simpson, R. (2002) Cross-disciplinary comparisons of hedging: Some findings from the Michigan Corpus of Academic Spoken English. In Reppen, R., Fitzmaurice, S. and D. Biber (Eds.), *Using Corpora to Explore Linguistic Variation*. Amsterdam: John Benjamins, 3-23.

195. Posteguillo, S. (2002) Netlinguistics and English for Internet Purposes, *IBÉRICA* 4: 21-38 (Universitat Jaume I, Castelló) http://www.aelfe.org/documents/text4-Posteguillo.pdf

196. Preston, D. (1993) Variation linguistics and SLA. *Second Language Research*, Vol. 9 (2): 153-172.

197. Raščevska, M., Kristapsone, S. (2000) *Statistika psiholoģijas pētījumos.* Rīga: Izglītības soļi.

198. Reppen, R., Fitzmaurice, S. and D. Biber (2002) (Eds.). *Using corpora to explore linguistic variation*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

199. Read, J. (2000) Assessing Vocabulary. Cambridge: Cambridge University Press.

200. Rissanen, M. (2007) Studying linguistic variation and change with the help of computers. *Finish Academy of Science and Letters Yearbook 2007*, pp. 49-53.

201. Rissanen, M., Kytö, M., Heikkonen, K. Eds. (1997) *English in Transition: Corpus-based Studies in Linguistic Variation and Genre Styles.* Berlin, New York: Mouton de Gruyter.

202. Rittenbruch, M., Mansfield, T. and Cole L. (2003) Making sense of "Syndicated Collaboration" Available from http://en.scientificcommons.org/606086

203. Rosell-Aguilar, F. (2005) Task Design for Audiographic Conferencing: Promoting Beginner Oral Interaction in Distance Language Learning. *Computer Assisted Language Learning*, Vol. 18 (5): 417-442.

204. Sacks, H., Schegloff, E., and G. Jefferson (1974) A simplest systematics for the organization of turn-taking for conversation. *Language*, Vol. 50: 696-735.

205. Sanderland, J. (2002) New communication practices and the psychological and affective function of e-mail on a 'distance', partly-taught PhD programme. *Studies in Higher Education,* Vol. 27/2: 233 - 246.

206. Savignon, S. J., & Roithmeier, W. (2004) Computer-mediated communication: Texts and contexts. *CALICO Journal, 21*(2): 265–289.

207. Schiffrin, D., Tannen, D. and H. E. Hamilton, eds. (2001) *The Handbook of Discourse Analysis*. Oxford: Blackwell Publishers.

208. Schiffrin, D. (1994) *Approaches to discourse*. Cambridge, MA: Blackwell Publishers.

209. Schiffrin, D. (1987) *Discourse Markers*. New York: Cambridge University Press.

210. Schrire, S. (2006) Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis. *Computers and Education*, Vol. 46: 49-70.

211. Scollon, R. and Scollon,W. (2001) Intercultural Communication: A Discourse Approach. 2nd ed. (first ed.1995), Blackwell Publishers.

212. Segerstad, Y.H. (2002) *Use and Adaptation of Written Language to the Conditions of Computer-Mediated Communication*. Unpublished Ph.D. thesis, Göteborg University, Sweden. Accessed on December 12, 2004, http://www.ling.gu.se/~ylvah/dokument.pdf

213. Seliger, W. and Shohamy, E. (1989) *Second Language Research Methods.* Oxford: Oxford University Press.

214. Shepherd M. and Watters C. (1998) The Evolution of Cybergenre. *Proceedings* of the 31st Hawaii International Conference on System Sciences.

215. Shortis, T. (2000) *The Language of ICT* London: Routledge.

216. Siepmann, D. (2005) Discourse Markers Across Languages. London, New York: Routledge.

217. Silver, M. S. (2006) *Language Across Disciplines: Towards a Critical Reading of Contemporary Academic Discourse.* Brown Walker Press.

218. Sims, B.R. (1996) Electronic mail in two corporate workplaces. In P. Sullivan and J. Dautermann (Eds.), *Electronic literacies in the workplace: Technologies of writing*. Urbana, IL: NCTE, 41-64.

219. Skadiņa, I. (2008) Jaunākie sasniegumi latviešu datorlingvistikas lietojumprogrammās. Valodniecības raksti - 2. Letonika, otrais kongress - Rīga : LU Latviešu valodas institūts

220. Smith, C.S. (2004) *Modes of Discourse: the Local Structure of Texts* Cambridge University Press.

221. Snyder, I. (Ed.) (2002) *Silicon Literacies: Communication, innovation and education in the electronic age* London, New York: Routledge.

222. Snyder, I. (ed.) (1998) *Page to Screen: Taking literacy into the electronic era.* London, New York: Routledge.

223. Sotillo, S. M. (2000) Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language learning and Technology*. Vol. 4 (1):82-119.

224. Spektors, A. (2001) Latviešu valodas datorfonda izveide Latvijas Zinātņu Akadēmijas Vēstis A daļa: Sociālās un humanitārās zinātnes - 2001 - 5 Latvijas Zinātņu Akadēmijas Vēstis - (A daļa: Sociālās un humanitārās zinātnes) - 2001 - 55.sēj., Nr.1/2. - 74.-82.lpp.

225. Stuart, K. (2006) Towards an Analysis of Academic Weblogs. *Revista Alicantina de Estudios Ingleses* 19: 387-404.

226. Stubbs, M. (1983) *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*. University of Chicago Press.

227. Surkova (2008) Creativity in Target Language Use. Unpublished doctoral dissertation, Riga, University of Latvia.

228. Swales, J. M. (2004) *Research Genres: Explorations and Applications* Cambridge: Cambridge University Press.

229. Swales, J. and Malczewski, B. (2001) Discourse management and new-episode flags in MICASE. In Simpson, R. and Swales, J. (Eds) *Corpus Linguistics in North America.* Ann Arbor, MI: University of Michigan Press,145-164.

230. Swales, J. M. (1990) *Genre analysis: English in academic and research settings.* New York: Cambridge University Press.

231. Takahashi, K. (2006) A Study of Register Variation in the British National Corpus. *Literary and Linguistic Computing.* Vol. 21(1): 111-126.

232. Thomas, M. (2002) Learning with Incoherent Structures: The Space of Online Discussion Forums. *Journal of Computer Assisted Learning*. 18(3): 351-66.

233. Thompson, G. and Yiyun, Y. (1991) Evaluation in the reporting verbs in academic papers. *Applied Linguistics*, Vol. 12(4): 365-382.

234. Thompson, J. (1984) *Studies in the Theory of Ideology*. Berkeley, CA: University of California Press.

235. Toyoda, E., Harrison, R. (2002) Categorization of text chat between learners and native speakers of Japanese. *Language Learning and Technology* Vol. 6 (1):82-99, http://llt.msu.edu/vol6num1/pdf/toyoda.pdf.

236. Tudini V. (2003) Conversational elements of online chatting: speaking practice for distance language learners? *Apprentissage des Langues et Systèmes d'Information et de Communication (ALSIC)*, Vol. 6 (2): 63 - 81.

237. Vandergriff, I. (2006) Negotiating common ground in computer-mediated versus face-to-face discussions. *Language Learning and Technology*, Vol. 10 (1):110-138.

238. van Waes, L., Leijten, M., Neuwirth, C. (2006) *Writing and New Media*. Elsevier Science.

239. Veermans, M. and Cesareni, D. (2005) The nature of the discourse in Web-based collaborative learning environments: Case studies from four different countries. *Computers and Education*. Vol. 45 (3): 316-336.

240. Walther, J. B. (1996) Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. Communication Research, 23 (1): 3-43.

241. Warschauer, M. (2002) Networking into academic discourse. *Journal of English for Academic Purposes* 1(1): 45-58. http://www.gse.uci.edu/faculty/markw/networking.html

242. Warschauer, M., Kern, R. (Eds.), (2001). *Network-based language teaching: Concepts and practice.* Cambridge: Cambridge University Press.

243. Warschauer, M. (1996) Comparing face-to-face and electronic discussion in the second language classroom. *CALICO Journal,* 13, (2/3): 7-25.

244. Warren, K. and Rada, R. (1998) Sustaining computer-mediated communication in university courses, Journal of Computer Assisted Learning, 14: 71-80.

245. Weininger, M. and Shield, L. (2004) Promoting oral production in a written channel: an investigation of learner language in MOO. *Computer Assisted Language Learning*, 16(4): 329–349, http://oro.open.ac.uk/983/

246. Werry, C.C. (1996) Linguistic and interactional features of Internet Relay Chat. In Susan C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives.* Amsterdam: John Benjamins Publishing, 29-46.

247. Wysocki, A.F. (2004) *The Multiple Media of Texts: How Onscreen and Paper Texts Incorporate Words, Images, and Other Media.* In Bazerman, C., Prior, P. *What Writing Does and How It Does It: an Introduction to Analysing Text and Textual Practices.* London: Lawrence Erlbaum Associates.

248. Wysocki, A.F., Johnson-Eilola, J., Selfe, C.L., Sirc, G. (2004) *Writing New Media: Theory and Applications for Expanding the Teaching of Composition.* Logan, Utah: Utah State University Press.

249. Yates, S. (1996) Oral and written linguistic aspects of computer conferencing: A corpus-based study. In Susan C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives.* Amsterdam: John Benjamins Publishing, 29-46.

250. Young, A. (2006) *Teaching Writing Across the Curriculum* (4th Ed.) Prentice Hall.

251. Yus, F. (2007) Weblogs: Web pages in search of a genre? In: *The Texture of Internet. Netlinguistics in Progress.* (Eds.) S. Posteguillo, M. Esteve and M. Gea-Valor. Cambridge: Cambridge Scholars Press.

252. Zitzen, M., and Stein, D. (2004) Chat and conversation: a case of transmedial stability? *Linguistics,* 42(5): 983-1021.

Glossaries

GLOSSARY OF TERMS USED IN THE STUDY

Academic discourse	a dynamic process of functional use of language as a means of communication in a situational context marked 'academic.' It means that the language users in academic discourse are the members of an academic discourse community, i.e. educated members of society that represent academic disciplines. Academic discourse is realized in semantically connected and meaningful to the members of a particular academic discourse community verbal instances of spoken or written language longer than a sentence.
Asynchronous interactional CMAD	the process of language use in CMC characterised by an exchange of messages in a way that there is usually some extended period of time between messages though, occasionally, the interchanges may be very fast. It is represented by academic e-mails, discussion forums and weblogs.
Blogspeak	the term refers to a variety of language that is commonly used in weblogs, thus emphasizing specific linguistic characteristics of weblogs distinguishing them from other types of language on the Internet (also called *Netspeak* by Crystal, 2001).
Computer-mediated communication (CMC)	human interaction through networked computers, i.e. the use of computers to create, store, deliver, and process communications.
Computer-mediated academic discourse	a computer-mediated process of functional use of language as a means of communication in academic context that is realised in semantically connected and meaningful to the communicating users of language verbal instances of spoken or written language longer than a sentence (*Lat. elektroniskais akadēmiskais diskurss*).
Computer-mediated discourse	a type of discourse that is mediated by computer. It is a dynamic process of functional use of language as an instrument in different modes of CMC, e.g. chats, discussion forums, e-mails, and Internet publishing, e.g. weblogs, hypertexts. It is realized in textual artefacts (semantically connected and meaningful for the communicating users of language verbal instances of spoken or written language longer than a sentence) that are produced by language users applying computer as a tool for communication. (*Lat. elektroniskais diskurss*)
Corpus	"a collection of naturally occurring language texts in electronic form, often compiled according to specific design criteria and typically containing many millions of words." (Halliday et al. 2005:168).
Corpus linguistics	"Corpus linguistics is simply the study of language through corpus-based research, but it differs from traditional linguistics in its insistence on the systematic study of authentic examples of language in use. ... **Language cannot be invented; it can only be captured." Sinclair (1997: 31)**
Dimension	*(in variation linguistics)* the continuum along which register

	variation occurs and the types of discourse differ from each other in the English language (Biber, 1988). As linguistic features that serve similar discourse functions tend to appear in similar text types, they form strong patterns of co-occurrence in texts. Different groups of co-occurring features constitute different functional (textual) dimensions.
Discourse	*(in linguistics)* a dynamic process of functional use of language as a means of communication in a situational context. It is realised in semantically connected and meaningful to the communicating users of language verbal instances of spoken or written language longer than a sentence (texts).
Discourse analysis	*(in linguistics)* "the analysis of language in use" Brown and Yule, 1983.
Discussion forums	*(academic)* a type of computer-mediated academic discourse. Discussion forums are realisations of asynchronous, interactional, text-based computer-mediated academic discourse with many participants (*Lat. diskusiju grupas, forumi*).
Discourse representation	the patterns of use of linguistic features in texts as linguistic realisations of discourse (Brown and Yule, 1983). Texts representing CMAD are discourse representations.
Electronic discourse	The term used by Davis and Brewer (1997) to mean *computer-mediated discourse.*
Electronic language	The term used by Collot and Belmore (1993) to mean *computer-mediated discourse*.
Electronic networked discourse	The term used by Hawisher and Selfe (1998) to mean *computer-mediated discourse*.
E-mail	*(academic)* a type of computer-mediated academic discourse. E-mails are realisations of asynchronous, interactional, text-based computer-mediated academic discourse. E-mails are usually exchanged between two participants though they may be forwarded to other people *(Lat. elektroniskais pasts; e-pasts).*
Extralinguistic	the term that is used as a synonym of *non-linguistic.*
Extratextual	the term that is used to denote contexts, factors or devices external to the studied text, e.g. specific to the electronic medium use of visual elements, colour, hyperlinks etc. in electronically transmitted texts.
Frequency	an actual count of the linguistic features in a corpus of texts.
Hyperlink	an underlined word, phrase or visual element connected (*hyperlinked*) to other texts or the parts of the same text mediated by computer (*hypertext*).
Hypertexts (academic)	*(academic)* – the term used to mean a type of computer-mediated academic discourse in the present research. Academic hypertexts are realisations of transactional written c*omputer-mediated academic discourse*.
Hypertext	a computer-mediated written text with some coloured or underlined words, phrases or visual elements connected (*hyperlinked*) to other texts or the parts of the same text (Lat. *hiperteksts*).
Interactants	people interacting in *computer-mediated communication.*

Interactional	The term used to mean the type of discourse in which language serves an interactional function by "expressing social relations and personal attitudes" (Brown and Yule, 1983: 1).
Interactional CMAD	a computer-mediated process of academic language use in which text messages are exchanged among the participants in CMC in a form of a dialogue or a *polylogue*.
Interactive networking	The term used by Baron (1984) to mean *computer-mediated discourse*.
Interactive written discourse	The term used by Ferrara, Bruner and Whittemore (1991) to mean *computer-mediated discourse*.
Lexicogrammatical	the term used to mean the linguistic features which represent *lexicogrammar* – the lexicon and grammar of a language, taken together as an integrated system in Systemic Functional Linguistics.
Linguistic data	Linguistic data in discourse analytic methodology are treated as a record (text) of the process of language use (discourse) and are used to find and describe the patterns in the use of linguistic features that people choose from available language resources to achieve their goals in communication (Brown and Yule, 1983).
Linguistic features	lexical and grammatical (lexicogrammatical) forms in a language.
Mean (value)	The arithmetic average, calculated by dividing the sum of all the scores by the number of scores.
Medium of communication	*(computer)* the term emphasizing computer as the agency of by which communication is or transferred. Although, to be precise, language in communication can be conveyed by a variety of substances (e.g. air, paper, a wall etc.), an intermediate agent (e.g. interpreter, reporter, etc.) or a mechanism (e.g. loud-speaker, telephone, SMS messenger, etc.), in this dissertation, the past participle of *mediate* – *mediated* – is used to mean that the communication is indirect and that the action is performed by a computer (as a medium) rather than any other media or humans in the process of human-to-human or human-to-machine communication.
Mode of communication	The terms used by Swales (2004) to mean *mode of discourse* (1), i.e. whether written or spoken language is used.
1. *Mode of discourse*	a linguistic category denoting the physical form in which language is used: written, spoken or signed (used in the present dissertation).
2. *Mode of discourse*	a linguistic category for the local level of analysis introduced by Smith (2004), similar to *text type* in the field of rhetoric. Smith posits five modes: Narrative, Report, Descriptive, Information, and Argument.
Mode of interaction	the way in which the process of language use is realised by material means (Norris, 2004). Mode of interaction in computer-mediated communication is defined by the computer software used for communication, which enables either *synchronous* or *asynchronous interactional CMAD*.
Multidimentional	a research approach to the analysis of linguistic variation in discourse introduced by Biber (1988) which states that it is

	not possible to analyse linguistic variation in discourse along one dichotomous *dimension*, e.g. speaking/ writing.
Networked	The term used by Warschauer and Kern (2001) to mean that computers sued for communication are connected in a net.
Nominalisation	A process which makes frequencies from samples of markedly different sizes comparable by bringing them to a common base (McEnery et al. 2006).
On-line seminars	*(academic)* a type of computer-mediated academic discourse. On-line seminars are the realisations of transactional spoken computer-mediated academic discourse. *(Lat. tiešsaistes semināŗs)*
Operationalisation	turning the research subject into observable and measurable quantities that are statistically countable and comparable, i.e. giving its *operational definition.*
Phatic	the use of language to open or close conversation.
Polylogue	(*in CMC*) a number of simultaneously occurring dialogues in CMC.
Specialised corpus	The corpus compiled for a special research purpose. It is usually domain or genre specific. The specialised corpora compiled for the present study are designed to represent the types of CMAD.
Type-token ratio	The ratio between the number of unique words (types) and the number of occurrences of a given word in the text (tokens).
Synchronicity	the term used by Collot and Belmore (1993) to refer to an immediate exchange of messages CMC.
Synchronous conferencing (chats)	*(academic)* – the term used to mean a type of *computer-mediated academic discourse (CMAD)* in the present research. Academic chats are the realisations of *synchronous interactional* text-based CMAD *(Lat. tērzēšana).*
Synchronous interactional CMAD	the process of language use characterised by an immediate exchange of messages or by relatively short time between them. It is represented by synchronous academic conferences (chats).
Register	*(in linguistics)* a variety of a language used for a particular purpose or in a particular social setting.
Register variation	the process of focusing on the similarities and d**is**similarities between **register** categories in terms of various lingu**is**tic phenomena (Takahashi, 2006).
Telecollaboration	The term used by Belz (2003) and Warschauer (1996) to mean *computer-mediated discourse,* usually applied to a collaboration of students communicating via networked computers in academic settings.
Text	the term denoting "a rich, many-faceted phenomenon that 'means' in many different ways" and referring "to any instance of [spoken or written] language, in any medium, that makes sense to someone who knows the language"(Halliday and Matthiessen, 2004: 3).
Text-based	the term used to mean "visually presented language" (Herring, 2001: 612) in CMC, i.e. messages exchanged in a form of a text typed with the help of computer keyboard.
Text types	a category assigned in linguistics and rhetoric, based on the internal, linguistic characteristics of texts themselves, in

	contrast to *genre*, which is a category "determined on the basis of external criteria relating to the speaker's purpose and topic.((Biber, 1988: 170).
Transactional	is such a process of computer-mediated academic language use in which the transactional function of language is performed by expressing 'content', i.e. conveying factual information.
Transactional CMAD	a process of computer-mediated academic language use in which the transactional function of language is performed by expressing 'content', i.e. conveying factual information.
Type of software	Computer programmes used for CMC (e.g. e-mail protocol, wiki etc.). Different software enables communication to take place in different ways (e.g. synchronous or asynchronous CMC).
Variation (linguistic)	is such a process in which language use is varied by language users to suit different situational contexts. It is realised in fluctuations in the frequency of occurrence of linguistic features in texts. These realisations of linguistic variation as a process are the differences in linguistic characteristics, i.e. different frequency of use of linguistic features that appear within texts, for example, frequency of personal pronouns, passive voice constructions etc. (*Lat. variācija,* Raščevska and Kristapsone, 2000).
Variation linguistics	*(also variationist* or *variational sociolinguistics)* a special sub-field of sociolinguistics that is concerned with the description of linguistic variation using statistical research methods.
Weblogs	*(academic)* asynchronous, rather interactional (i.e. transitional between interactional and transactional) written computer computer-mediated academic discourse (Lat. tīmekļa žurnāls, emuāri).

LATVIAN-ENGLISH GLOSSARY OF THE STUDIED LINGUISTIC FEATURES

Latvian	**English**
1. pirmās personas vietniekvārdi	First person pronouns (PRO1)
2. otrās personas vietniekvārdi	Second person pronouns (PRO2)
3. trešās personas vietniekvārdi	Third person pronouns (PRO3)
4. vietniekvārds *it*	Pronoun *it* (IT)
5. norādāmie vietniekvārdi	Demonstrative pronouns (DEMP)
6. nenoteiktie vietniekvārdi	Indefinite pronouns (INDP)
7. saīsinājumi	Contractions (CONTR)
8. *that* izlaišana	Complementizer *that* deletion (THATD)
9. nolūka prievārdi	Final (stranded) prepositions (FPR)
10. sadalīti palīgdarbības vārdi	Split auxiliaries (SAUX)
11. jautājumi, kas sākas ar "Wh"	WH questions (WHQ)
12. tipa/pazīmes koeficients	Type/token ratio (TTR)
13. vidējais vārdu garums	Mean word length (MWL)
14. vidējais sintaktiskais teikuma garums	Mean syntactic length per sentence/utterance (MSL)
15. Nominalizācijas	Nominalizations (NOM)

16. lietvārdi	Number of nouns (N)
17. darbības vārdi pagātnes laikā	Past tense verbs (PTV)
18. darbības vārdu divdabju formas	Perfect aspect verbs (PAV)
19. darbības vārdi tagadnē	Present tense verbs (PRTV)
20. ciešamā kārta bez subjekta	Agentless passives (ALPASS)
21. ciešamā kārta, izmantojot *by*	*by* passives (BYPASS)
22. varbūtības modālie darbības vārdi	Possibility modals (PMOD)
23. nepieciešamības modālie darbības vārdi	Necessity modals (NMOD)
24. iespēju modālie darbības vārdi	Prediction modals (PRMOD)
25. *be* kā galvenais darbības vārds	*be* as the main verb (BE)
26. Personiski darbības vārdi	Private verbs (PVERB)
27. publiski darbības vārdi	Public verbs (PUBV)
28. pārliecināšanas darbības vārdi	Suasive verbs (SUV)
29. nenoteiksmes formas	Infinitives (INF)
30. atributīvie īpašības vārdi	Attributive adjectives (ATADJ)
31. predikatīvie īpašības vārdi	Predicative adjectives (PRADJ)
32. vietas apstākļa vārdi	Place adverbials (PLADV)
33. laika apstākļa vārdi	Time adverbials (TADV)

34. citi apstākļa vārdi	Other adverbs/adverbials (ADV)
35. savienojumi	Conjuncts (CONJ)
36. Norobežojumi/ norobežotāji	Hedges (HED)
37. Paplašinātāji/pastipriātāji	Amplifiers (AMP)
38. emocionāli ekspresīvie elementi	General emphatics (GENEM)
39. diskursa partikulas	Discourse particles (DPART)
40. kauzāls pakārtojums	Causative adverbial subordination (CSUB)
41. nosacījuma pakārtojums	Conditional adverbial subordinators (COND)
42. Paligteikumu pakārtojums	Other adverbial subordinators (ADVS)
43. teikumi, kas sākas ar "Wh"	WH clauses (WHC)
44. attieksmes teikums ar *that*	*that* relatives (RTHAT)
45. 'Wh' attieksmes teikums kā papildinātājs	WH relatives on object position (WHRCO)
46. 'Wh' attieksmes teikums kā priekšmets	WH relatives on subject position (WHRCS)
47. pagātnes formu divdabja teikumi	Past participial postnominal (reduced relative) clauses (PPC)
48. tagadnes divdabja teikumi	Present participle clauses (PRPCL)
49. analītiskie noliegumi (*not*-)	Analytic (*not*) negation (ANEG)

50. sintētiskie noliegumi (no-)	Synthetic negation (SYNEG)
51. Divdabji	Gerunds (GER)
52. prievārdu skaits	Prepositions (NPR)
53. teikumu attieksmes	Sentence relatives (SREL)
54. frāžu saskaņošana (savienojumi)	Phrasal coordination (PHC)
55. teikumu saskaņošana 56. (savienojumi)	Clausal coordination (CLC)

Appendices

Appendix 1

Empirical Studies of Computer-Mediated Discourse

Publication	Research focus/object	Type of CMD	Research Method	Findings
Holmes (1987)	participation in CMD	synchronous	discourse analysis	Participants adapted to the environment and successfully pursued multiple threads of discourse
Collot and Belmore (1996)	Bulletin board system messages (BBSs) 200, 000-word corpus	Synchronous asynchronous	*Descriptive framework and Comparison* along Biber's (1988) six dimensions	e-language is neither spoken nor written – a discrete 'electronic language' –a 'hybrid' variety of English (p.53)
Ko (1996)	difference between CMD and f-t-f spoken and written discourse	synchronous (*Daedalus, InterChange* software)	discourse analysis	Discourse mode is not merely intermediate between speaking and writing; the electronic medium uniquely fosters some behaviours and inhibits others.
Yates (1996)	Messages from open conferences on British network CoSy Corpus 684,550 words (50 messages)	asynchronous	*Comparison* of samples of spoken and written language with CMC conferencing along four continua: TTR, lexical density, degree of personal reference, modal auxiliary use counted per 1000 words.	CMC is not speech or writing but its own language, more written-like than spoken-like
Werry (1996)	Language of Internet Relay Chat (IRC)	synchronous	1 Qualitative Description	Identifies features unique to the medium: 'high degree of addressivity' (p.52) 'brevity and abbreviation' (p.54), 'paralinguistic and prosodic cues' (56) tendency to produce auditory and visual effects in writing (58) – simulation of discoursive style of f-t-f spoken language.
Blake (2000)	textual CMD entries of fifty intermediate L2 Spanish learners (interlanguage)	synchronous chat program *Remote Technical Assistance* (RTA)	two experiments with different task types	Evidence of predominance of incidental lexical negotiations, in contrast to the paucity of syntactic negotiations. CMC can provide benefits ascribed to the Interaction Hypothesis, but with greatly increased possibilities for access outside of the classroom environment.
Kitade (2000)	eleven advanced Japanese as L2 learners' interactions in CMC	Synchronous chat (50 min sessions); 24 Internet Chat discussions	discourse analysis and conversation analysis	CMC provides potential benefits for learning.

Paramskas, (2000)	intermediate level French as L2 learners Frequency of initiation of topics, interaction within a topic, participation, language functions that are promoted by CMD (e.g. speech acts)	Synchronous (One-hour session conferencing in computer lab)	Qualitative (descriptive)	CMD encourages fluency in writing
Painter, Coffin, Hewings (2003)	the role of the tutor in CMD among three tutorial groups of postgraduate students in Applied Linguistics (UK)	asynchronous	qualitative	the less intervention cause the less results
Stockwell (2003)	sustainability of e-mail interactions among 48 learners of Japanese as L2	asynchronous e-mail (NS-NNS)	qualitative	The strategies to sustain CMC
Belz and Reinhardt (2004)	Transcripts of CMD and e-mails between German as L2 learner and Ns and NNs of German	synchronous chat and asynchronous e-mail	discourse analysis (case study)	important function of language play in CMD
Warner (2004)	the transcripts of CMD (19 students)	synchronous MOO	CMD analysis	theories of communication and communicative competence need to be expanded to include language play
Savignon & Roithmeier (2004).	U.S. German as L2 learners interacting with German English as L2 learners CMD	asynchronous	Quantitative (number of messages)	German learners posted less to the discussion board than did the US learners
Zitzen and Stein (2004)	Linguistic, interactional and pragmatic characteristics of CMC discourse (chat logfiles) 12,422 word corpus	synchronous	*Qualitative case study: description and Comparison* of chat with spoken conversation in "canonical situation of language use' Lyons (1997: 637)	Chat – 'digitality' – neither speech nor writing. Conversation and chat – different genres. There are more features of written than spoken language in chats.
Veermans and Cesareni (2005)	Collaboration in CMD (in Finland, Greece, Italy, and the Netherlands)	Synchronous and asynchronous	a series of case studies of CMD	Computer-mediated collaboration enhances learning
Nguyen, Hanh Thi; Kellogg (2005)	Identity construction and participation patterns in on-line discussions	asynchronous (discussion forums)	discourse analysis and ethnography	language socialization was uniquely facilitated by the affordances of computer technology.

Spiliotopoulos and Carey (2005)	using CMD (WebCT) to improve writing in English as L2 in Canada	asynchronous	relationship between language and identity	monologic and dialogic writing tasks that occur in CMD have important implications for student participation, motivation, and inter-cultural awareness.
Belz and Vyatkina (2005)	German (L2) modal particles use in the context of CMD	Synchronous	Qualitative/ quantitative	the authenticity of the interactions, the corpus-enabled nature of the intervention, the developmental scope of the data, and the potential for "hyper-noticing" in Internet-mediated intercultural foreign language education
Jeong (2005)	group interactions in CMD	asynchronous	sequential analysis and specific software tools and techniques to facilitate the analysis of message-response sequences	latent variables (e.g., message function, response latency, communication style) and exogenous variables (e.g., gender, discourse rules, context) affect how likely a message is to elicit a response, the types of responses elicited by the message, and whether or not the elicited sequence of responses (e.g., claim--challenge--explain) mirror the processes that support group decision-making, problem-solving, and learning.
Moore and Marra (2005)	Transcripts of CMD of thirty-seven graduate students in two course sections	asynchronous	content analysis and Interaction Analysis Model (IAM)	both sections exhibited co-construction of knowledge; less structured section reached the highest phase of knowledge building.
Pelowski, Frissell, Cabral, Yu (2005)	Immediacy of participation in CMD	synchronous	14 immediacy indicators applied	Student immediacy was not related to examination performance, but the immediacy behaviour of support was positively correlated with changes in pre- to post-course test scores.
Bretag, T. (2006)	teacher-student relationship, cultural identity	asynchronous, two-way email exchanges	CMD analysis	CMD reveals cultural identity and differences
Chen (2006)	identity, power relations, culture-specific ideologies, culture-specific politeness in student-professor interaction of Taiwanese graduate student with her American professors	asynchronous (e-mail)	Critical discourse analysis (longitudinal case study)	reveals the complexity of an L2 learner's evolving e-mail practice and struggle for appropriateness. Calculated a proportion of all c-units, reception strategies which occur at a rate of 7.2% in CMC as compared to 5.7% in FTF (similar)
Vandergriff (2006)	reception strategies, such as reprises, hypothesis testing and forward inferencing of 18 students of advanced German	synchronous (InterChange® software - the participant can see only his/her own message while writing)	Qualitative and quantitative	CMD and FTF are very similar with respect to fostering the use of reception strategies; online environment does not necessarily foster the use of communication strategies any more than the conventional classroom.

Appendix 2
1000-word sample of academic discussion forum messages

Each discussion forum message starts with its <tagged> ordinal number in the corpus.The items with numbers above them have been encoded as linguistic features investigated in the present study. These numbers correspond to the numbers of linguistic features in Appendices 8-14.

<358> 1.Dear all[57], during[52] the last[30] days[16] I[1] received[17] some[6] informal[30] questions[16] and[54] suggestions[15]concerning the coordination[15] of working[30] groups[16]. 2. I[1] just[38] would[24] like[29] to let[29] you[2] know[29] that[44] some[6] members[16] would[24] not[49] like[29] to give[29] a single[30] presentation[15] or[54]workshop[16] but[54] prefer[19] to discuss[29] and[54] reflect[29] with[52] colleagues[16] in[52] working[30] groups[16]. 3. In[52] order[16] to coordinate[29] those groups[16] and[54] to facilitate[29] your[2] possible[30] decision[15] making[51] I[1] would[24] like[29] to know[29]: 4. Who[11] is[25] interested[31] to contribute[29] to[52] (one of) these groups[16]? 5. Who[11] would[24] like[29] to take[29] responsibility[15] for[52] guiding[51] the discussions[15] and[54] reflections[15] in[52] the group[16]? 6. Please[39], just[38] give[58] me[1] a short[30] signal[16]. 7. And[55], please[39], do[58] not[49] hesitate to contact[29] us[1] with[52] drafts[16] of[52] ideas[16]. 8. We[1] will[24] collect and[54] coordinate them[3].
9. Looking forward[34] to[52] your[2] reactions[15], on behalf of[52] the organization[15] team.[16]
<Mary Brown>
<353> Dear[57] fellow[30] members[16],
Despite[35] the deceptively[34] attractive[30] offer[16], I[1] must[23] voice[29] my[1] objection[15] to[52] the proposal[15]. As[52] the web[16] master[16] for[52] TESOL Greece I[1] am[8] often[33] contacted by[21] such[30] individuals[16], and[55] we[1] have[19] a strict[30] policy[16] of[52] saying[51] no[50]. The danger[16] is[25] that[44] the information[15] may[22] be resold by[21] Lawrence Erlbaum Associates[16] on[52] to[52] less[34] scrupulous[30] advertisers[16].

However[35], in[52] order[16] not[49] to waste[29] an opportunity[15], why not[49] follow[29] our[1] procedure[16], which[53] is[25] to create[29] a page[16] on[52] the EATAW web[16] site[16] entitled[47] 'for[52] the interest[16] of[52] our[1] members[16]' on[52] which[45] you[2] allow[19] sponsored[30] adds[16], but[54] do not[49] endorse[19] the product[16] or[54] service[16] (thereby[52] avoiding[51] legal[30] problems[16] if[41] they[3] don't[49] live[19] up[52] to[52]
their[3] promises[16]). Lawrence Erlbaum Associates[16] may[22] not[49] want[29] to pay[29] so much as[42] before[52], but[55] at least[36] they[3]'ll[7,24] pay something[6].
I[1] hope[19] my[1] experience[16] has been[18] of[52] use[16].
<John Smith, Athens Greece>
<354 >
Thank you[2] <Bob> for[52] your[2] suggestion[15]. I[1] agree[19] wholeheartedly[34]. I['7]m[25] well[34] aware[31] that[44] Erlbaum may[22] not[49] be[29] 'just'[38] a publisher[16] and[55] that[44] they[3] may[22] be[29] very[37]
trustworthy[31], but[55] I[1] prefer[26] to receive[29] mail[16] only[34] if[41] I[1] really[38] asked[27] for[52] it[4].
Perhaps[36] Erlbaum can[22] be invited[20] to be[29] present[31] in[52] <Bochum>, if[41] there is[25] going to be[29] an information[15] market[16] or[54] something[6] like[36] that[5]?
Best regards[16],
<Tom White>
<351> Dear All[57],
I[1] would[24] agree[29] with[52] John and[54] Mary: set up a separate[30] space[16] where[45] we[1] can[22] go[29] if[41] we['7]re[25] interested[31] in[52] what[45] Lawrence Erlbaum and[54] others have[19] to offer[29] or[55], if[41] that[5]'s[25] somehow[6] not[49] feasible[31], simply[34] say[27] no[50].
<341> Dear[57] Friends[16],
I[1] have[10] recently[33] had[18] a lot[38] of[52] trouble[16] with[52] Information[15] Technology[16]

students[16] and[54] plagiarism[16]. Their[3] main[30] excuse[16] is[25] that[44] they[3] can[22],[7]t[49] summarize[29] information[5] with[52] a lot[38] of[52] formulas[6], etc[35] in[52] the writing[51]. Instead[35] they[3] copy[19] whole[30] pages[16] of[52] text[16] and[54] reference[19] the source[16] at[52] the end[16] of[52] the paper[16].

I[1],[7]m not[49] willing[19] to tell[29] them[3] to put[29] quotation[15] marks[16] around[52] that large[30] chunk[16] of[52] text[16] and[54] cite[29] the reference[16] immediately[33], becuase[40] I[1] still[34] feel[26] the "chunk[16]" is[25] too[30] big[31]. Recently[33], I[1] had[17] a crisis[16] when[53] a sweet[30] young[30] exchange[16] student[16] copied[17] about[36] 12 pages[16] of[52] mathematical[30] text[16] related[47] to[52] fuzzy[30] logic[16]. She[3] had[17] just[38] a few paragraphs[16] at[52] the end[16] of[52] her[3] own.

After[52] this experience[16] I[1] realized[26] that[44] I[1] have[19]n[7]t[49] a clue[16] of[52] how[43] to help[29] these students[16] learn[29] to summarize[29] such[38] complicated[30] math[16] that[44] I[1] have[19] no[50] knowledge[16] of[52] myself[1]. Does[11] anyone[6] have[19] any[6] suggestions[15] of[52] how[43] I[1] can[22] begin[29]

to learn[29] myself[1] so that[35] I[1] can[22] help[29] them[3]? Hint[16]: I[1] am not[49] willing[19] to start[29] taking[51] math[16] and[54] information[15] technology[16] courses[16] myself[1]. :-)

One of[52] the instructors[16] in[52] the IT Department[15] showed[17] me[1] a beautifully[34] simple[30] diagram[16] and[54] told[17] me[1] that[44] if[41] the students[16] can[22] understand[26] this[5] (and[55] they[3] should[23] by[52] the time[6] they[3],[7]re working[19] on[52] master[16] degree[16] papers[16]), they[3] should[23] be[29] able[31] to write[29] straightforward[30] texts[16] using[48] the formulas[16] as[52] illustration[5]. Well[39], if[41] it[4],[7]s[25] so[38] easy[31], why[11] do I[1] have[19] so[38] many students[16] copying?

Help[58]! Any[6] suggestions[5] will[24] be[29] most[38] welcome[31],

<Susan Stone>

<316> Dear[57] <Mary, Ann, Susan,>

This[5] is[25] a very[37] interesting[30] topic[16] you[2] have[19] here[32]. At[52] the Leiden English[16] Department[15] (Netherlands), I[1],[7]ve been involved[18] in[52] designing[51] and[54] teaching[51] a Blackboard[16] course[16] on[52] how[45] to write[29] Masters[16] theses[16] at[52] our[1] department[15]. The problem[16] you[2] mention[27] in[52] your[2] abstract[16] is[25] exactly[34] the one we[1] encountered[17] when[43]

we[1] were designing[17] the course[16]. There were[25] a number[16] of[52] books[16] and[54] articles[16] that[44] were[25] somewhat[6] useful[31] to[52] us[1], as[52] were[25] a number[16] of[52] existing[30] online[30] writing[30] courses[16], mainly[34] OWLs in[52] the U.S., and[54] other related[30] online[30] courses[16]. But[55] none[50] of[52] these[5] hit[17] the spot[16] so to speak[36], because[40] they[3] were[25] either[42] too abstract[31] or they[3] were geared[20] to[52] local[30] conditions[15] which[43] differed[17] considerably[34] from[52] ours[1]. At[52] Leiden we[1] ended up[17] having[51] to develop[29] a lot[38] of[52] material[16] of[52] our[1] own. I[1],[7]d[24] certainly[34] be[25] interested[31] in[52] attending[51] your[2] symposium[16] and[54] making[51] a short[30] contribution[15] about[52] our[1] Leiden experience[16].

Cheers[57]

<Henry Morgan>

<292> I[1],[7]d[24] like[29] to add[29] a little aside[34] to[52] the plagiarism[16] conversation[15]. It[4] could[22] very[37]

well[34] be[25] that[44] second-language[30] writers[16] "borrow[19]" text[16] in[52] much the same way[16] as[52] non-native[30] speakers[16] appropriate[19] text[16] that[44] they[3] hear[27] native[30] speakers[16] say[27] (sometimes[6] called[47] " scaffolding[51]"). If[41] so, "borrowing[51]" written[30] text[16] might[22] be[25] a more[38] or[54] less[34] unavoidable[30] stage[16] in[52] the development[15] of[52] second-language[30] writers[16]. (Unfortunately[34] I[1] lost[17] the reference[16] to[52] a study[16] which[43] proposes[27] something[6] along[52] these lines[16].) One implication[15] would[24] be[25] that[44], in[52] writing[30] courses[16] for[52] beginners[16], students[16] would[24] first be[10] encouraged[20] to cut[29] and[54] paste[29] and[54] generally[34] play[29] around[52] with[52] a selection[15] of[52] source[16] texts[16]. Next[42], they[3] could[22]

move on[29] to[52] writing[51] cut-and-paste[30] papers[16] (sometimes[6] called[47] " patchwriting[51]")

with[52] citations[15] where necessary[31], and[54] a list[16] of[52] works[16] cited[47]. Only[34] after[52] that[5] would[24] students[16] be[10] ready[31] to move on[29] to[52] writing[51] papers[16] of[52] their[3] own, using[48] their[3] own voice[16], without[52] recourse[16] to[52] plagiarism[16], and[54] paraphrases[16] rather than[35] cut-and-paste[16].
<Dan Brown>
<282> <Hi[57] Ann,>
I[1] could[22] n't[7,49] agree[29] more[38], this[5] is[25] a really[38] good[30] way[16] of[52] telling[51] students[16] why[43] plagiarism[16] is[25] undesirable[31]. But[55] still[34], at[52] the risk[16] of[52] sounding[51] repetitive[31], grammar[16] errors[16] are[25] undesirable[31], yet[55] there is[25] a phase[16] in[52] everyone's foreign[30] language[16] learning[51] where[43] you[2] make[19] them[3]. I[1] realize[26] the comparison[15] is[25] not[49] entirely[37] valid[31], but[55] it[4,7]s[25] the best[37] I[1] can[22] come up[29] with[52] right[39] now[33]...
<Steven>

Statistics (sample)**:** 48 sentences, 1000 types, 441 tokens, 4824 characters
Total in the corpus: 76742 words, 362 postings, 391435 characters (no spaces)

Appendix 3

1000-word sample of academic online seminar (transcript)

The items with numbers above them have been encoded as linguistic features investigated in the present study. The numbers correspond to the numbers of linguistic features in Appendices 8-14.

1. HELLO[57] EVERYONE. 2. WELCOME[58] TO[52] WEBCT SEMINAR[16] SERIES[16] . 3. THIS[5] IS[25] THE 19TH SEMINAR[16] IN[52] OUR[1] SERIES[16]. 4. WE[1] ARE[25] DELIGHTED[31] TO HAVE[29] YOU[2] WITH[52] US[1] TODAY[33]. 5. OUR[1] TOPIC[16] IS[25] "AVOIDING[51] PITFALLS[16] WITH[52] YOUR[2] WEBCT CAMPUS[16] EDITION[15] ." 6. OUR[1] SPEAKER[16] TODAY[33] IS[25] SAM BRADLEY OF[52] WEBCT .

7. IN[52] TERMS[16] OF[52] THE FORMAT[16] THAT[44] WE[1] WILL[24] FOLLOW, THE SEMINAR[16] WILL[24] LAST ABOUT[36] ONE HOUR[16] , SCOTT WILL[24] PRESENT[27] 30 TO[52] 40 MINUTES[16] . 8. THEN[33] WE[1] WILL[24] FOLLOW THE PRESENTATION[15] WITH[52] A 10 TO[52] 20 MINUTE[16] QUESTION[16] AND[54] ANSWER[16] SESSION[16] . 9. WE[1] ASK[27] THAT[44] YOU[2] HOLD[29] YOUR[2] QUESTIONS[16] UNTIL[52] THE MODERATOR[16] AND[54] PRESENTER[16] ASKS[27] FOR[52] THEM[3]. 10. WE[1] WILL[24] FOLLOW UP[52] TODAY'S[33] SEMINAR[16] WITH[52] A ONE WEEK[16] DISCUSSION[15] ON[52] WEBCT'S WEBSITE[16]. 11. YOU[2] CAN[22] ACCESS[29] THE SESSION[16] AND[54] AN ARCHIVE[16] OF[52] TODAY'S[33] SEMINAR[16] AT WEBCT.COM / SEMINAR[16]. 12. THE DISCUSSION[15] WILL[24] BE[25] AVAILABLE[31] THIS AFTERNOON[16] AND[55] THE ARCHIVE[16] WILL[24] BE[25] AVAILABLE[31] WITHIN[52] 24 HOURS[16].

13. IF[41] YOU[2] ARE EXPERIENCING[19] TECHNICAL[30] DIFFICULTIES[15], YOU[2] CAN[22] CLICK[29] THE HELP[16] BUTTON[16], WHICH[46] IS[25] RIGHT[38] IN[52] THE MIDDLE[16] OF[52] THE HORIZONLIVE[16] INTERFACE[16]. 14. IF[41] YOU[2] ARE HAVING[19] ANY[6] TROUBLE[16] AT[52] ALL WITH[52] INTERNET[16] AUDIO[16], IT[4] IS[25] BEST[31] TO CALL[29] INTO[52] YOUR[2] TELEPHONE[16] YOU[2] WILL[24] CAST. 15. THE NUMBER[16] FOR[52] NORTH[30] AMERICANS[16] IS[25] 877-825-5810. 16. THAT CALL[16] IS[25] TOLL[16] FREE[31]. INTERNATIONALLY[34], THE NUMBER[16] IS[25] 706-634-8111. THAT NUMBER[16] UNFORTUNATELY[34] IS[25] NOT[49] TOLL[16] FREE[31].

17. CLOSED[30] CAPTIONING[51] IS[25] AVAILABLE[31] TODAY[33] FOR[52] THE HEARING[51] IMPAIRED. 18. TO ACTIVATE[29] THOSE[5] – CLOSED[30] CAPTIONING[51] TYPE[58] /CC INTO[52] THE CHAT[16] BOX[16] TO ENTER[29]. 19. TO DEACTIVATE[29] CLOSED[30] CAPTIONING[51] DO[58] THE SAME THING[16] TYPE[58] /CC THEN[33] HIT[58] RETURN[16]. 20. WEBCT WOULD[24] LIKE TO THANK[29] HORIZONLIVE[16], ONE PARTNERS[16] IN[52] OUR[1] NETWORK[16] FOR[52] MAKING[51] TODAY'S[33] SEMINAR[16] AVAILABLE[31]. 21. HORIZONLIVE IS[25] A WEB[16] BASED[30] TEACHING[51], LEARNING[51], INTERACTIVE[30] SEMINAR[16] OVER[52] THE INTERNET[16]. 22. THEY[3] WORK[19] ON[52] MACK AND[54] UNIX COMPUTERS[16] AND[54] FUNCTION[19] ON[52] A WEB[16] BROWSER[16]. 23. FOR[52] MORE INFORMATION[15] VISIT[58] HORIZONLIVE.COM. NOW[33] THAT[44] WE[1] HAVE[19] THE INTRODUCTION[15] OUT[52] OF[52] THE WAY[16], LET'S[1] GET[29] STARTED[47]. SAM, THANKS[16] FOR[52] JOINING[51] US[1] TODAY[33]. 24. THE FLOOR[16] IS[25] ALL YOURS[2].

25. THANK YOU[2], I[1] WILL[24] MOVE MY[1] SLIDES[16] AND WE[1] CAN[22] GET[29] STARTED[47].

26. OKAY, WE[1] WOULD[24] THANK[29] EVERY ONE FOR[52] COMING[51] TODAY[33], I[1] SEE[26] SOME[6] FAMILIAR[30] NAMES[16], WHICH[53] IS[25] GREAT[31]. 27. OUR[1]

ATTENDANCE[15], AS YOU[2] CAN[22] SEE[26] HERE[32], WE[1] WILL[24] TALK[27] ABOUT[52] THE WEB[16] CAMPUS[16] EDITION[16] ARCHITECTURE[16], GIVE US[1] AN OVERVIEW[16] OF[52] HOW[45] IT[4] IS BUILT[20], SOME[6] OF[52] THE IMPORTANT[30] STRUCTURE[16] FEATURES[16] INSIDE[32] WEBCT. 28. WE[1] WILL[24] TALK[27] ABOUT[52] HARDWARE[16] CONFIGURATIONS[15] TO GO[29] OVER[52] SOME[6] OF[52] THE ISSUES[16] THAT[44] PEOPLE[16] THINK[26] ABOUT[52] OR[54] NEED[19] TO THINK[29] ABOUT[52] WHEN[48] PURCHASING NEW[30] HARDWARE[16]. 29. PLANNERS[16] FOR[52] WEBCT, THINGS[16] WE[1] CAN[22] IDENTIFY[29]. 30. DISCUSS[27] DIFFERENT[30] BACKUP[30] STRATEGIES[16], RECOVERY[16] METHODS[16] THAT[44] WORK[19] WELL[34] WITH[52] THE CAMPUS[16] EDITION[16] APPLICATION[15], SPECIFIC[30] REQUIREMENTS[15] THAT[44] WE[1] NEED[19] TO BE[29] AWARE[31] OF[9]. 31. AND[55], ENDING[51] UP[52] THE PRESENTATION[15] WITH[52] A LITTLE[30] BIT[16] ON[52] SYSTEM[16] MONITORING[51], SUGGESTIONS[15] ON[52] THINGS[16] THAT[44] ARE[25] -- OUR[1] USERS[16] SHOULD[23] BE[29] LOOKING AT[52] WHEN[48] FOLLOWING THE WEB[16] USAGE[16]. 32. OKAY[39]. SO[39], TO START[29] OFF[52] WITH[52] THE WEBCT STRUCTURE[16], THERE IS[25] -- ARE[25] A FEW MAIN[30] COMPONENTS[16] THAT[44] I[1] WOULD[24] LIKE[29] TO POINT[29] OUT[52] AND[54] SORT OF[36] DISCUSS[27] HOW[43] THESE[5] ARE CHANGING[19] IN[52] THE FUTURE[16], THANKS[16] TO--THINGS[16] TO LOOK[29] FORWARD TO[9]. 33. SO[39] THERE IS[25], FIRST OF ALL[35], THE COURSE[16] DATA[16] STRUCTURE[16], WHERE[43] [THERE IS] THE COURSE[16] INFORMATION[15] STORAGE[16]-- STORAGE[16] IN[52] WEBCT, WHICH[46] WOULD[24] INCLUDE[29] DISCUSSION[15] BOARDS[16], MAIL[16] POSTINGS[16] OUR[1] INTERNAL[30] DATA[16] BASES[16]. 34. THE GLOBAL[30] DATA[16] BASE[16] AS A WHOLE IS[25] A SEPARATE[30] SECTION[16] FROM[52] THE COURSE[16], SO[55] WE[1] WANT[19] TO DISCUSS[29] THAT[5], IF[41]-- THE ISSUES[16] INVOLVED[47] WITH[52] POSSIBLY[34] MOVING[51] TO[52] NEW[30] HARDWARE[16] OR[54] YOUR[2] CURRENT[30]--CANCELLATION[15] TO[52] THE OPERATING[30] SYSTEM[16]. 35. TALK[27] ABOUT[52] APACHE[16], HOW[43] WEBCT IS AFFECTED BY[21] IT[4], WEBCT AND[54] THE IMPORTANT[30] - -IMPORTANCE[16] OF[52] LOG[16] FILES[16] AND[54] WATCHING[51] THOSE LOG[16] FILES[16] FOR[52] ANY[6] PECULIAR[30] MESSAGES[16], LASTLY[35] I[1] WANT[19] TO POINT OUT[29] THE WEBCT TOOLS[16] AND[54] RECOVERY[16] KITS[16], WHERE[43] TO FIND[29] THOSE[5], EVERY ONE IS[25] AWARE[31] OF[52] HOW[43] TO USE[29] THEM[3]. 36. OKAY[39]. 37. TO GET[29] INTO[52] A LITTLE BIT[36] MORE DETAIL[16], I[1] BASICALLY[34] LAID OUT[17] THIS[5] SO[55] YOU[2] CAN[22] SEE[26] WHERE[43] SOME[6] OF[52] THESE IMPORTANT[30] FEATURES[16] ARE[25]. 38. SO[55] TO START OFF[29] WITH[52] AT[52] THE TOP[16], THE INSTALLATION[15]

DIRECTORY[16] OUR[1]—THIS[5] EXISTED[17] IN[52] ALL WEBCT. 39. YOU[2] WILL[24] FIND THESE SCRIPTS[16] IN YOUR[2] WEBCT INSTALLATION[15] UNDER[52] WEBCT /SCRIPT[16]. 40. THERE ARE VERSIONS[16] FOR[52] EACH ONE OF[52] THE PLATFORMS[16]. 41. YOU[2] CAN[22] TAKE[29] A LOOK[16] AT[52] EACH ONE OF[52] THESE[5] TO SEE[29] HOW[43] IT[4] IS[25] AVAILABLE[31]--WE[1] ARE[25]--WE[1] WILL[24] TALK[27] ABOUT[52] THE END[16] OF THIS PRESENTATION[15]. 42. THERE WILL[24] BE[25] A PRESENTATION[15] OF[52] A DESCRIPTION[15] OF[52] HOW[45] THE COMPONENT[16] WORKS[19] TOGETHER[34], HOW[45] THESE SCRIPTS[16] ARE USED[20]. 43. I[1] AM[25] SURE[31] THAT[44] MOST OF YOU[2] KNOWS[26], APACHE[16], TO DELIVER[29] TO[52]--CONTENT[16]. 44. WEB[16] APACHE[16] SERVES[19] OUR[1] PAGES[16], EVERYTHING TO DO[29] WITH[52] IT[4] IS STORED[20] UNDER[52] THE WEBCT SERVER[16] DIRECTORY[16]. 45. SO[55] IT[4] IS[25] A VERY[37] IMPORTANT[30] DIRECTORY[16] FOR[52] FINDING[51] THINGS[16] LIKE[52] ACCESS[16] LOG[16], APACHE[16] CONFIGURATION[15] FILE[16] EXISTING[48] WITHIN[52] THIS STRUCTURE[16], HDDTT.COM. 46. SO[55] IF[41] YOU[2] ARE[25] GOING TO BE IMPLEMENTING[29] THINGS[16] LIKE[52]—YOU[2] NEED[19] TO BE[29] FAMILIAR[31] WITH[52] WHERE[43] TO GO[29] IN[52] THESE DIRECTORIES[16] TO FIND[29] THAT INFORMATION[15]. 47. WE[1] WILL[24] FIND OUT ABOUT[52] THAT[5] LATER ON[33] IN[52] THIS DIRECTORY[16] IN[52] THIS PRESENTATION[15]. 48. WITHIN[52] THE WEBCT USER[16] DIRECTORY[16], THIS DIRECTORY[16] CONTAINS[19] MOSTLY[34] THE FILES[16] FROM[52] WITHIN[52] THE FILE[16] MANAGER[16] AREA[16], INSIDE[32] WEBCT COURSES[16]. 49. IT[4] ALSO[34] CONTAINS[19] A LARGE[30] NUMBER[16] OF[52] THE IMAGES[16] THAT[44] ARE USED[20] WITHIN[52] WEBCT. 50. SO[55] IF[41] YOU[2] ARE CUSTOMIZING[19] WEBCT WITH[52] YOUR[2] OWN IMAGES[16], THIS[5] IS[25] THE LOCATION[16] ON[52] WHERE[43] YOU[2] WOULD[24] FIND[29] THOSE IMAGES[16]. 51. DATA[16] BASES[16], UNDER[52] WEBCT COURSES[16], YOU[2] WILL[24] FIND[29] A LIST[16] OF[52] YOUR[2] COURSE[16] ID THAT[44] EXISTS[19] ON[52] YOUR[2] SERVER[16], ALL OF[52] THE DATA[16] WITHIN[52] THIS DIRECTORY[16] ARE[25] THE DATA[16] BASES[16] THAT[44] ARE[25] WITHIN[52] THE COURSES[16] THEMSELVES[3], THIS[5] IS[25] A VERY[37] IMPORTANT[30] AREA[16] WITHIN[52] WEBCT AND[55] IT[4] IS[25] A LARGE[30] REASON[16] OR[54] LARGE[30] THE SIZE[16] OF[52] THE WEBCT STRUCTURE[16]. 52. SO[55], AS[40] WEBCT IS MOVING FORWARD[19], WITH[52] THE VERSION[16] 4.1, WE[1] ARE WORKING[19] TOWARD[34] REDUCING[19] THE NUMBER[16] OF[52] FILES[16] SO[55] THAT[44] IT[4] IS[25] NOT[49] AS[52] CHALLENGING WHEN[43] TRYING[48] TO RECOVER[29] BACK[52] TO[52] WEBCT INSTALLATION[15]. 53. THE PEOPLE[16] WHO[46] HAVE IMPLEMENTED[18] 4.1, KNOW[26] THAT[44] THESE[5] ARE[25]--

THESE AREAS[16] HAVE BEEN[18] OPERATED[20]. 54. WE[1] HAVE TO[19] RUN[29] SCRIPT[16] IN ORDER TO[35] UPGRADE[29] YOUR[2] WEB[16] COURSES[16] DIRECTORY[16]. 55. WHAT[46] IT[4] IS DOING[19], [IS] CHANGING[19] SOME[6] OF[52] THE FORMATS[16] INSIDE[32] THE DIRECTORY[16], INSIDE[32] THE DATA[16] BASES[16] FOR[52] THINGS[16] LIKE[52] MAIL[16], DISCUSSION[15], CALENDAR[16] ENTRY[16],

PUTTING A LOT OF WHAT WAS PREVIOUSLY TEXT FILE INTO A BINARY DATA BASE FORMAT. (end of the last sentence)

Statistics: Types 1003, tokens 361, TTR 35.99 , sentences 55 , characters 4 763 (no spaces), MWL 4.76, MSL 18.2
***Sourse*:** Seminar "Avoiding Common Pitfalls with Your Growth of WebCT Campus Edition" transcript October 22, 2003 Presenter: Scott Baily, WebCT
http://www.webct.com/seminar/viewpage?name=seminar_archive

Appendix 4

1000-word sample of academic e-mails

The items with numbers above them have been encoded as linguistic features investigated in the present study. The numbers correspond to the numbers of linguistic features in Appendices 8-14. All the names have been changed.

<E-mail 18>

1. Dear <Jane>[57]

I[1] am[10] currenly[34] putting[19] together[34] a proposal[16] for[52] an ELTeCS project[16] with[52] some[6] colleagues[16] in Latvia and Estonia. 2. I[1] have[19] a query[16] about[52] the funding[30] rules[16]. 3. Part[16] of[52] our[1] project[16] invovles[19] two face[16] to[52] face[16] meetings[16] of[52] the team[16]. 4. I[1] understand[26] that[44] travel[16] costs[16] are[25] not[49] eligible[31]. 5. Does[11] this[5] simply[34] apply[19] to[52] fares[16] and[55] could[22] we[1] ask[19] for[52] funding[51] for[52] accommodation[15] for[52] members[16] attending these meetings[16]?

6. Thanks[16] for[52] your[2] help[16]

<Mary> (72 words, 6 sentences, MSL 12, characters 349, MWL 4.8)

<E-mail 19>

1.Dear <Susan>[57] -

Thank[39] you[2] for[52] your[2] query[16]. 2. Yes[34], ELTeCS funds[16] are[10] often[33] granted[20] to pay[29] for[52] modest[30] accommodation[15] costs[16], while[42] participants[16] find[19] their[3] own fares[16].

3. I[1] look[19] forward to[52] receiving[51] your[2] proposal[16] -

4. Best[38] wishes[16] -

<Jane> (38 words, 4 sentences, MSL 9.5, characters 184, MWL 4.8)

<E-mail 21>

1. Hi Everyone[57]

I['7]ve tried[18] to address[29] the comments[16] from[52] the British[30] Council[16]. 2. Attached[31] is[25] the latest[30] version[16] - draft[16] 3.

3. Can[22] you[2] let[29] me[1] have[29] any[6] comments[16]?[11]

4. I[1] feel[26] quite[34] optimistic[31]. 5. They[3] don['7]t[49] seem[19] to be ruling[29] us[1] out at[52] this stage[16].

6. Best[38] wishes[16]

<Doroty> (45 words, 6 sentences, MSL 7.5, characters 201, MWL 4.5)

<E-mail 6>

1. Hi All[57]

I[1] wrote[17] to[52] John to ask[29] how[43] much we[1] had missed[18] by in[52] terms[16] of[52] points[16] and[55] this[5] is[25] the reply[16] I[1] got[17]. 2. It[4] sounds[19] as if[42] it[4] is[25] worth[52] trying[51] again[33].

3. Best[38] wishes[16] to[52] you[2] all for[52] the holiday[16] season[16].

<Mary> (43 words, 3 sentences, MSL 14.3, characters 156, MWL 3.6)

<23>

1. Hi <Mary>[57]

thanks[16] for all of this[5]. 2. I[1] also[34] feel[26] very[37] optimistic[31]. 3. I[1] have made[18] some[6] adjustments[15] and[54]will[24] send out next version[16] when[43] I[1] have made[18] all adjustments[15].

4. All the best[57]

<Jane> (32 words, 4 sentences, MSL 8, characters 144, MWL 4.5)

<24>

1. Hi Everyone[57]

I[1] received[17] the following[30] feedback[16] on[52] the draft[16] of[52] our[1] proposal[16] from[52] the British[30] Council[16]. 2. i[1] am going[19] to adapt[29] the proposal[16] in[52] the light[16] of[52] their[3] comments[16] and[54] will[24] send it[4] to[52] you[2] later[33] today[33]. 3. meanwhile[42], if[41] you[2] have[19] any[6] views[16] on[52] their[3] comments[16], it[4] would[24] be[25] useful[31] for[52] me[1] to have[29] them[3] as soon as[33] possible[31].

4. Thanks[39]
<Mary> (61 words, 4 sentences, MSL 15.3, characters 266, MWL 4.4)
< 27>
1. <Mary>,
Sorry it[4] took[17] so[38] long[34] to reply[29]. 2. I[1] think[26] this[5] is[25] a wonderful[30] idea[16] and[55] I[1] would[24] be[25] interested[31] in[52] hearing[51] more[34] about[52] the work[16] and[54] in[52] participating[51] if[41] I[1] can[22].
3. A colleague[16] of[52] mine[1] here[32], who[46] is[25] out[52] of[52] town[16], might[22] also[34] be[25] interested[31] in[52] this[5] and[55] I[1] will[24] talk to[52] him[3] to see[29] if[41] any[5] of[52] his[3] work[16] would[24] be[25] useful[31] for[52] us[1] all.
4. The best[16] to[52] you[2] on[52] your[2] work[16],
<Mary> (75 words, 4 sentences, MSL 18.8, Characters 278, MWL 3.7)
< 28>
1. Dear <Mary>[57],
It[4] was[25] a pleasure[16] to hear[29] your[2] presentation[15] in[52] Athens and[54] to talk[29] with[52] you[2] while[42] I[1] was[25] there[32]. 2. If[41] you[2] do get[19] a group[16] going[48] to investigate[29] the area[16] you[2] propose[19], please[39] count[19] me[1] in[52].
3. Wishing you[2] well[34],
<Mary Brown> (43 words, 3 sentences, MSL 14.3, characters 180, MWL 4.2)
<38>
1. Hello <Mary>[57]
Thank[39] you[2] very[38] much[39] for[52] the photos[16]. 2. I[1] have[10] just[33] got[18] back[52] from[52] the conference[16]
after[52] quite[34] a journey[16] via[52] Ethiopia and[54] South Africa.

3. The photos[16] brought[19] back[52] great[30] memories[16] of[52] wonderful[30] people[16].
4. Take[58] care[16]...I[1] look[19] forward[32] to[52] reading[51] your[2] writing[52].
. Best[38] wishes[16]
<John> (46 words, 5 sentences, MSL 9.2, characters 228, MWL 4.96)
< 52>
1. Dear <John>[57]
Can[22] you[2] please[39] be[25] more[34] specific[31]?[11]
<John Smith> (8 words, 1 sentence, MSL 8, characters 37, MWL 4.6)
<65>
1. Hi John[57],
thanks[16] for[52] the tips[16].
2. Actually[34] I[1]'[7]ve got[18] no[50] problem[16] with[52] Clan[16], I[1] think[26] it[4]'[7]s[25] great[31].
3. But[55] what[43] I[1] am[10] really[38] talking[19] about[52] is[25] when[43] you[2] are putting[19] sections[16] of[52] a completed[30] transcript[16] into[52] an article[16] or[54] a thesis[16].
4. However[35] I[1] think[26] I[1]'[7]ve[10] kind of[36] answered[18] my[1] own question[16].
5. My[1] version[16] of[52] Word[16] doesn'[7]t[49] handle[19] unicode[16] but[55] I[1] believe[26] the latest[30] version[16] does[19]. 6. I[1]'[7]ve got[18] around[52] that problem[16] by[52] using[51] NeoOffice[16] the Mac[16] version[16] of[52] OpenOffice[16] which[45] does handle[19] Unicode[16].
(81 words, 6 sentences, MSL 13.5, characters 373, MWL 4.6)
<74>
1. <Mary Ann>, in[52] that case[16] I[1]'[7]d[24] beg[29] to differ[29]. I[1] read[17] the <Swales> and[54] <Feak> book[16]and[55] I[1] based[17] our[1] online[30] thesis[16] training[30] course[16] on[52] genre[16] analysis[16] and[54] the
"moves[16]" <Swales> and[54] <Feak> identify[19] (with[52] some[6] ideas[16] of[52] my[1] own added[47]). 2. What[45] I[1] didn'[7]t[49] do was[25] include[29] "trawling[30]" exercises[16] and[55] I[1]'[7]m thinking[19] about[52] including[51] them[3] in[52] next[30] year's[16] version[16]. 3. What[45] I[1] did[17] do was[25] include[29] spot-the-

move[30] exercises[16] and[54] is-this-the-right-move[30] exercises[16], using[48] examples[16] from[52] old[30] theses[16]. 4. I[1] agree[19] with[52] you[2] entirely[37] that[44] this[5] has[19] nothing[50] to do[29] with[52] plagiarism[16] of[52] patchwriting[51].
5. Cheers[57]
<John> (93 words, 5 sentences, MSL 18.6, characters 468, MWL 5
< 75>
1. Hi <Mary>[57],
I[1] love[19] that Graham Greene quote[16]... It[4,7]s[25] so[38] true[31]. 2. Language[16] that[44] you[2] really[38] like[19] might[22] not[49] have[29] the intended[30] effect[16], i.e.[35] it[4] might[22] draw[29] attention[15] to[52] itself[4] rather than to[52] the message[16]. 3. Or[55] there[7]s[25] the ever[33] lurking[30] danger[16] of[52] purple[30] prose[16]. 4. As for[42] the gentleman[16] with[52] English[16] as[52] a fourth language[16] who[45] collects[19] nice[30] turns[16] of[52] phrase[16] to use[29] in[52] his[3] own work[16], I[1] think[26] this[5] is[25] a good[30] example[16] of[52] one of[52] the less[34] desirable[30] strategies[16] that[44] John mentioned[17] – you[2] don[7]t[49] learn[19] much that way[16] do you[2].
5. Cheers[57]
(92 words, 5 sentences, MSL 18.4, characters 405, MWL 4.4)
< 76>
1. Hi John[57]
The gentleman[16] you[2] refer[19] to[52] is[25] a senior[30] and[54] much[34] published[30] (in[52] English[16]) professor[16]! 2. He[3] certainly[34] learned[17] that way[16]. 3. He[3] made[17] the comment[16] in[52] response[16] to[52]a question[16] about[52] how[43] he[3] learned[17] to write[29] in[52] English[16] for[52] publication[15] in[52]
international[30] journals[16] and[54] in[52] his[3] books.[16]
<Mary > (49 words, 3 sentences, 16.3, characters 226, MWL 4.6)
<77>
1. Hi Mary[57],
Ah[39], now[33] I[1] get[19] it[4]. 2. Thanks[39]. 3. Would[24] this[5] imply[29] that[44] one kind of[36] class[16] activity[16]
could[22] be trawling[19] publications[15] for[52] nice[30] turns[16] of[52] phrase[16] to use[29] later[33]?[11] 4. I[1] can[22]
see[26] how[43] students[16] might[22] enjoy[29] doing[51] this[5].
5. Cheers[57]
<John> (40 words, 5 sentences, MSL 8, characters 175, MWL 4.4)
<86>
1. Dear Colleagues[57],
I[1] am[10] currently[33] working[19] on[52] a project[16] in[52] which[45] I[1] propose[27] interactive[30] tasks[16] for[52] the development[15] of[52] academic[30] reading[30] and[54] writing[30] skills[16] in[52] electronically[34] delivered[47] English[16] for[52] Academic[30] Purposes[16] programs[16].
2. I[1] am looking for[19] references[16] of[52] studies[16] involving[48] the evolution[15] of[52] different[30] approaches[16] to[52] the teaching[16] of[52] reading[51] and[54] writing[51] in[52] online[30] EAP programs[16].
3. All studies[16] I[1] have found[18] so far[33] are[25] concerned[31] with[52] online[30] teaching[51] in[52] general, not[49] online[30] language[16] teaching[51].
4. I[1] would[24] appreciate[19] any[6] information[15] you[2] can[22] provide[29] me[1] with[9].
5. Thank[39] you[2],
<Mary Brown> (88 words, 5 sentences, MSL 17.6, 472 characters, MWL 5.4)
<87>
1. Dear colleagues[57],
a colleague[16] of[52] mine[1] sent[17] me[1] the information[15] to share[29] about[52] the conference[16] bellow[32]. 2. I[1] believe[26] Florence in[52] September[16] is[25] a very[38] good[30] place[16] for[52] meetings[16]:-))

3. Enjoy[58]!
4. Best[38] wishes[16],
<Mary> (32 words, 4 sentences, MSL 8, characters 164, MWL 5.1)
<89>
1. Thank[39] you[2] very[38] much to[52] all those[5] who[45] answered[17] my[1] query[16] about[52] uses[16] of[52] "home[16]" and[54] "home[16]" as[52] compound[30] terms[16].
2 All the best[16]
<John> (24 words, 2 sentences, MSL 12, characters 102, MWL 4.25)
< 91>
1. Good[30] morning[16] Langusers[16],
I[1,7]m searching[19] some[6] theory[16], thesis[16], model[16], paradigm[16] about[52] the role[16] of[52] personal[30] experience[16] in[52] face-to-face[30] dialogue[16], particularly[34] in[52] institutional[30] context[16].
2. I[1] wonder[26] if[41] someone[6] could[22] suggest[29] me[1] (or[55] everyone who[45] is[25] interested[31] in[19]) the maior[30] theory[16] or[54] studies[16] about[52] this matter[16].
3. Thank[39] you[2] very[38], very[38] much[39]!
<Mary Brown>, PhD (52 words, 3 sentences, MSL 17.3, characters 290, MWL 5.6)
Statistics (in the sample): 19 e-mails, 52.6 words per email (min 8, max 93, SD 24.3), tokens 1000, types 339, TTR 33.9, MWL 4.6 (min 3.6, max 5.6, SD 0.5), sentences 79, MSL 12.7 (min 7.5, max 18.5, SD 4.2)

Appendix 5
1000-word sample of academic weblog text

The items with numbers above them have been encoded as linguistic features investigated in the present study. The numbers correspond to the numbers of linguistic features in Appendices 8-14.

1. Great[31] to see[29] your[2] hypertext[16] work[16] continuing to delelop[29]! 2. I[1] like[19] your[2] thought[16] bubble[16]/ hypertext[16] comparison[15]. 3.The illustration[15] really[38] helps[19] make[29] the point[16]. 4.[I[1] don't[49] have[19] my[1] copy[16] of[52] Landow[16] with[52] me[1]... was[11] that[44] the source[16] of[52] the graphic[16]? 5. Could[22] have been[18] cited[20] more[38] clearly[34].]

6.You[2] probably[34] know[26] me[1] well[34] enough[34] to predict[29] that[44] I[1]'[7]d[24] have[29] some[6] critiques[16] to offer[29], but[55] don't[49] worry – they[3] are[25] minor[30] details[16].

7. There are[25] some[6] who[46] say[27,19] that[44] the first webpage[16] ever[34] was[25] also[34] the first weblog[16] – it[4] was[25] pretty[30] much[34] a list[16] of[52] links[16] to[52] other universities[16] with[52] web[16] pages[16].

8. You[2] call[19] Vannevar[16] Bush[16]'s "items[16]" similar[30] to[52] what[45] we[1] call[19] "files[16]", but[55] I[1] think[26,8] it[4]'s[7] probably[34] fairer[31] to say[29] that[44] Bush[16] was talking[17] about[52] a photocopy[16] of[52] what[45] we[1] would[24] call[29] a file[16]; his[3] was[25] a photomechanical[30] storage[16] system[16] that[5] wouldn't[24,49] have let[18] the user[16] edit[29] the contents[16] (other than[52] by[52] somehow[6] getting[51] a printout[16], cutting[51] and pasting[51] with[52] scissors[16] and[54] glue[16], and[54] then[33] storing[51] the result[16]).
9. Although[35] I[1]'ve seen[18,8] similar[30] figures[16] used[47] elsewhere[32], I[1]'m[25] sure[31,8] there were[25] more[38] than[52] 23 weblogs[16] in[52] 1999 (I[1] was blogging[17] then[33], and[55] I[1] don't[49] think[26,8] I[1] was[25] that cutting[30] edge[16]). 10. I[1] don't[49] know[26]... maybe[34] early[33] in[52] 1999 the term[16] "weblog[16]" hadn't[49] spread[18] very[37] far[34]. 11. I[1] don't[49] think[26,8] I[1] actually[34] used[17] the term[16] "weblog[16]" for[52] several months[16] – not[49] until[33] I[1] realized[17] that[44] there was[25] a name[16] for[52] what[45] I[1] was doing[17]. 12. (I[1] began[17] by[52] pretty[30] much[34] shamelessly[34] copying[51] the layout[16] and[54] methodology[15] of[52] Arts[16] & Letters[16] Daily[16].) 13. But[55] certainly[34] by[52] the summer[16] of[52]1999, when[53] the first really[38] good[30] blogging[30] tools[16] went[17] mainstream[16], things[16] took[17] off[9]. 14. Certainly[34] there had[10] previously[34] been[18] tons[16] of[52] "Link[16] of[52] the Day[16]" sites[16], but[55] those pages[16] rarely[34] had[17] annotations[15] along[52] with[52] their[3] links[16], and[55] the webmasters[16] probably[34] coded[17] them[3] by[52] hand[16] (manually[34] inserting[48] new[30] entries[16] at[52] the top[16] and[54] cutting[48] and[54] pasting[48] to move[29] old[30] entries[16] to[52] archive[16] pages[16]... 15. I[1] spent[17] HOURS[16] doing[48] that stuff[16], before[52] I[1] finally[42] wrote[17] my[1] own lame[30] blogging[30] tool[16], which[53] I[1] happily[34] abandoned[17] for[52] the one I[1]'[7]m using[19] now[33]).
16. I[1]'[7]m[25] amused[31] to see[29] my[1] blog[16] listed[47] in[52] the appendix[16] as[52] the utter[16] and[54] complete[30] opposite[16] of[52] a personal[30]/journal[16] weblog[16]. 17. (Mary[16], I[1] remember[19,26] bugging[51] you[2] to put[29] more[38] links[16] in[52] your[2] blog[16]!)
18. Although[35] I[1] have[19] a few minor[30] quibbles[16], this[5] is[25] an excellent[30] overview[16] of[52] the blogging[30] phenomenon[16].
19. Mary[16], after[52] then[33] the basic[30] introduction[15] I[1] gave[17] you[2] last[30] year[16], you[2] seem[19] to have[29] really[38] taught[17] yourself[2] well[34]. 20. I[1]'[7]m[25] sure[31,8] I[1] wouldn't[24,49] be[25] as[52] invovled[31] with[52] blogging[51] now[33] if[41] I[1] hadn't[49] been[18] able[31] to practice[29] with[52] you[2] last[30] year[16] -- thanks[16] for[52] being[51] an eager[30] and hard-working[30] student[16].

21. It[4]'s[7] been[18] a pleasure[16] to read[29] your[2] paper[16] and[54] contemplate[29] your[2] achievements[15].

22. I[1]'[7]m[25] glad[31] that[44] you[2] read[17] this paper[16]. I[1] was going[17] to send[29] you[2] a link[16] to[52] the site[16] when[42] I[1] was[25] finished[31], but[55] I[1] guess[19,27] you[2] found[17] it[4] on[52] your[2] own. 22. Just[38] a few responses[16]...

23. First[35], the thought[16] bubble[16] was[25] my[1] own "brilliant[30]" creation[15] (yes[34], I[1] can[22] use[29] Photoshop[16]). 24. I[1]'[7]m[25] not[49] sure[31] if[41] you[2] are being[19] sarcastic[31] when asking[48] for[52] the source[16],

but[55] I[1,7]m[25] not[49] sure[31] how[42] much Landow[16] used[17] diagrams[16] that[5] eventually[34] lead[17] to[52] going out[51] (although[35] I[1,7]m[25] sure[31,8] he[3] occasionally[34] thought[17] about[52] it[4]). 25. It[4,7]s[25] possible[31] that[44] I[1] subconsciously[34] mimicked[17] a diagram[16] in[52] Landow's[16] text[16], although[35] I[1] am[25] honestly[34] not[49] sure[31]. 26. I[1] actually[34] asked[17,27] one of[52] my[1] friends[16] in[52] the computer[16] lab[16] to tell[29] me[1] exactly[34] what[43] he[3] was thinking[17], exactly[34] as[52] he[3] was thinking[17] about[52] it[4], in[52] the exact[30] way[16] that[44] the thoughts[16] came[17] to[52] him[3]. 27. Although[35] I[1,7]m[25] sure[31] the string[16] of[52] thought[16] is[25] not[49] perfectly[37] accurate[31], I[1] think[26,8] it[4] makes[19] a valid[30] point[16].

28. My[1] capstone[16] professor[16] commented[17] on[52] the Jorn[16] Barger[16] "web[16] + blog[16]" thing[16] too[34]. 29. I[1] guess[19,27,8] I[1] am[25] confused[31] about[52] where[43] the term[16] "blog[16]" came[17] from[9]?? 30. Was[11] it[4] originally[34] "webblog[16]," as[52] a single[30] word[16]? 31. "Weblog[16]" seems[19] more[38] intuitive[31] to[52] me[1]...since[40] it[4] is[25] sort[16] of[52] a log[16] for[52] links[16] on[52] the web[16]. 32. That[5]'s[7,25] something[6] that[44] I[1] didn't[7] really[38] pay enough[34] attention[15] to[52] in[52] the paper[16], and[55] that[44] I[1] will[24] hopefully[34] revisit during[52] my[1] upcoming[30] 9-month[30] break[16].

33. I[1] went[17] back[52] to[52] the research[16,8] I[1] did[25] last[30] year[16] for[52] your[2] class[16], and[55] I[1] found[17] the "23 weblogs[16] in[52] 1999" comment[16] on[52] Mary[16] Brown's[16] site[16] originally[34], and[54] have found[18] similar[30] numbers[16] since[42]. 34. That[5] seems[19] like[52] a pretty[30] hard[30] thing[16] to verify[29], though[35], considering[48] that[44] weblogging[51] was[25] still[34] a pretty[30] new[30] phenomenon[16]. 36. I[1] might[22] also[34] spend[29] some[6] time[16] researching[51] that[5] in[52] the future[16] why[11]? because[40] I[1] am[25] obsessive[30] compulsive[31]. 37. I[1]'ve[7] always[34] wondered[26] when[43] you[2] began[17] your[2] weblog[16]...?

38. Finally[35], I[1] guess[19,27] that[44] I[1] always[34] resisted[17] adding[51] more[38] links[16] because[40] I[1] am[25] extremely[37] lazy[31]. 39. I[1] didn't[7,49,10] really[38] become an Internet[16] geek[16] until[52] after[52] I[1] began[17] my[1] weblog[16], and[54] still[34] have[19] difficulty[15] finding[51] the time[16] and[54] patience[16] to read[29] even[34] the mainstream[30] news[16] over[52] the computer[16]. 40. We[1] now[33] get[19] free[30] print[30] versions[16] of[52] the NYTimes[16] at[52] school[16], and[55] I[1] really[38] like[19] how[43] important[31] I[1] look[19] when[43] I[1,7]m reading[19] it[4], drinking[19] coffee[16], with[52] my[1] PowerBook[16] sitting[48] next[32] to[52] me[1]. Kidding[19]. 41. But[55], I[1] do like[19] the observation[15] John[16] Smith[16] makes[19] in[52] "Blogger[16] Manifesto[16]," when[43] he[3] says[19,27], "I[1] liked[17] the vanity[16] of[52] a site[16] devoted[47] to[52] ME[1]." 42. I[1] am[25] all about[52] the site[16] devoted[47] to[52] me[1] 43. (Heck[39], I[1,7]m[25] all about[52] ANYTHING[6] devoted[47] to[52] me[1]!) 44. I[1] think[26,8] this[5] is[25] a very[37] apt[30] point[16], and[54] a very[37] common[30] motivation[15] for[52] weblogs[16] maintained[47] by[52] students[16] and[54] people[16] my[1] age[16]. 45. We[1] have[19] very[37] little opportunity[15] to write[29], let alone write[29] about[52] ourselves[1], and[55] weblogs[16] have offered[18] an unrivaled[30] freedom[16] for[52] young[30] writers[16].

46. Thank you[2] for[52] all of[52] your[2] help[16]. 47. You[2] have[10] really[38] been[18] a huge[30] influence[16] on[52] my[1] educational[30] experience[16]. 48. I[1] wish[26] that[44] you[2] had been[18] here[32] to help[29] me[1] with[52] this paper[16]--it[4] was[25] quite[34] difficult[31] to do[29] on[52] my[1] own, but[54] well-worth[31] the effort[16] in[52] the end[16]. 49. You[2]'ll[7] be[25] happy[31] to know[29] that[44] I[1] got[17] an A (which[53] is[25] possibly[34] due[31] to[52] the fact[16] that[44] noone[50] had[17] absolutely[37] any[6] idea[16] of[52] what[43] I[1] was talking[17] about[9]).

***Statistics*:** Types 1000, tokens 417, TTR 41.7, sentences 49, characters 4567, MWL 4.6 MSL 20.4.
Source: weblog address: http://jerz.setonhill.edu/weblog/essays (file *literacy weblog* p.16), accessed 12th December, 2003.

Appendix 6
1000-word sample of academic hypertext

The items with numbers above them have been encoded as linguistic features investigated in the present study. The numbers correspond to the numbers of linguistic features in Appendices 8-14.

<Hypertext 3>

1. Capillary[16] filtration[15] is[25] a key[16] area[16] in[52] the understanding[16] of[52] cardiovascular[30] function[16] and[54] has[19] both physiological[30] and[54] pathophysiological[30] relevance[16] in[52] nearly[36] every organ[16] system[16]. 2. This article[16] describes[19] how[43] classic[30] papers[16] in[52] the Legacy[16] collection[15] of[52] American[30] Physiological[30] Society[16] publications[15] can[22] be[19] used[20] in[52] a teaching[30] symposium[16] exploring[48] the evidence[16] supporting[48] current[30] concepts[16] of[52] capillary[16] fluid[16] exchange[16]. 3. Individual[30] students[16] are[19] given[20] papers[16] to read[29], edit[29], and[54] present[29] to[52] the class[16]. 4. The appropriate[30] selection[16] and[54] sequencing[51] of[52] these papers[16] allows[19] the development[15] of[52] important[30] physiological[30] concepts[16] to be[29] tracked[20]. 5. A series[16] of[52] papers[16] concerned[47] with[52] capillary[16] filtration[15] is[19] suggested[20,27]. 6. The contribution[15] of[52] each to[52] the developing[30] story[16] is[19] outlined[20]. 7. This approach[16] allows[19] students[16] to develop[29] critical[30] and presentation[15] skills[16] and[54] provides[19] them[3] with[52] a case[16] study[16] of[52] the scientific[30] method[16] as[55] it[4] is[19] applied[20] to physiology[16] as well as[35] establishing[48] an appropriate[30] knowledge[16] base[16] concerning[48] the role[16] of[52] hydrostatic[30] and[54] oncotic[30] forces[16] in[52] capillary[16] fluid[16] exchange[16]. 8. Relevant[30] teaching[30] points[16] are[19] explored[20] further[34] using[48] questions[16] based[47] on[52] a figure[16] from[52] one of[52] the three classic[30] papers[16] used[47]: microinjection[15] studies[16] of[52] capillary[16] permeability[15], the relationship[15] between[52] capillary[16] pressure[16] and[54] the rate[16] at[52] which[45] fluid[16] passes[19] through[52] the walls[16] of[52] single[30] capillaries[16].

9. In[52] addition[15] to[52] gaining[51] relevant[30] knowledge[16] and[54] understanding[51] of[52] physiological[30] concepts[16], our[1] students[16] also[34] need[23] to develop[29] a range[16] of[52] general[30] skills[16] that[44] have[19] wide[30] applications[15] within[52] both science[16] and[54] the wider[30] world[16]. 10. At[52] the highest[30] level[16], these skills[16] include[19] the ability[16] to analyze[29] and[54] synthesize[29] complex[30] ideas[16], a critical[30] approach[16] to evidence[16], and[54] the ability[16] to communicate[29] difficult[30] ideas[16] clearly[34], both orally[34] and[54] in writing[51]. 11. The approach[16] outlined[47] here[32] describes[19] how[43] original[30] research[16] papers[16] can[22] be[19] used[20] to facilitate[29] the development[15] of[52] these skills[16]. 12. Topics[16] are[19] explored[20] through[52] a series[16] of[52] oral[30] presentations[15] given[47] by[21] the students[16], each outlining[48] an important[30] study[16] related[47] to[52] the theme[16]. 13. This approach[16] emphasizes[19] evidence[16] over[52]

information[15] while[42] seeking[48] to ensure[29] that[44] students[16] gain[19] a clear[30] grasp[16] of[52] the basic[30] concepts[16].

14. The class[16] described[47] here[32] deals[19] with[52] capillary[16] filtration[15] and[54] is[19] taught[20] within[52] a module[16] on[52] advanced[30] cardiovascular[30] physiology[16]. 15. The teaching[30] method[16] could[22] be applied[20] to[52] most[30] areas[16] of[52] study[16] and[54] is[25,10] particularly[34] appropriate[31] for[52] postgraduate[30] students[16]. 16. By[52] the end[16] of[52] the module[16], each student[16] has[19] read[18], interpreted[18], edited[18], and[54] presented[18] separate[30] papers[16], which[46] are[25] the papers[16] relating[48] to[52] each of[52] five different[30] topics[16] in[52] the module[16], as well as[34] listened[18] to[52] and[54] discussed[18] other presentations[15] by[52] their[3] fellow[30] students[16]. 17. The selected[30] papers[16] should[23] have[29] made[18] an important[30] contribution[15] to[52] the field[16] and[54] should[23] stand[29] as[52] good[30] examples[16] of[52] clear[30] scientific[30] writing[51]. 18. In[52] this regard[16], the classic[30] papers[16] identified[47] in[52] the American[30] Physiological[30] Society[16] Legacy[16] archive[16] represent[19] an excellent[30] resource[16] from[52] which[45] classic[30] papers[16] can[22] be selected[20] secure[30] in[52] the knowledge[16] that[44] their[3] quality[15] is[25] not[49] in[52] doubt[16]. 19. This style[16] of[52] teaching[51] has[18] been[25] a core[30] element[16] of[52] teaching[51] in[52] our[1] department[15] since[42] taught[47] here[32] and[55] the main[30] steps[16] are[19] summarized[20] in[52] Table[16] 20. It[4] is[25] our[1] practice[16] to introduce[29] each topic[16] in[52] the module[16] through[52] a lecture[16] in[52] which[45] the range[16] of[52] the material[16] to be[29] considered[20] can[22] be outlined[20] and[55] issues[16] that[44] are[25,10] likely[34] to give[29] students[16] particular[30] problems[16] are[19] addressed[20]. 21. It[4] is[25] often[33] prudent[31] to explain[29,27] the principles[16] underpinning[48] the methods[16] used[47] in[52] specific[30] papers[16] because[40] these[5] may[22] not[49] be easily[34] sourced[20] from[52] monographs[16] or[54] other papers[16]. 22. At[52] the end[16] of[52] this session[16], each student[16] is[19] assigned[20] a paper[16] from[52] which[45] they[3] prepare[19] a presentation[15] with[52] computer[16] graphics[16]. 23.The presentations[15] are[19] delivered[20] to[52] the whole[30] class[16] as[52] a teaching[30] symposium[16], with[52] five minutes[16] set[47] aside for[52] questioning[51] after[52] each talk[16].24. Students[16] are[19] given[20] early[30] feedback[16] based[47] on[52] presentation[15] style[16], content[16], and[54] their[3] response[16] to[52] questions[16] along[52] with[52] their[3] mark[16] for[52] the session[16].

25. A proper[30] grasp[16] of[52] the mechanisms[16] determining[48] capillary[16] filtration[15] is[25] crucial[31] if[41] students[16] are[25] to understand[29,26] how[43] normal[30] tissue[16] hydration[15] is[19] maintained[20]. 26. It[4] also[34] provides[19] a necessary[30] foundation[15] when[43] considering[48] the specialization[15], e.g.[35], to favor[29] filtration[15] in[52] the kidney[16] or[54] absorption[15] in[52] the gastrointestinal[30] tract[16] and[54] lungs[16]. 27.The significance[15] of[52] this[5] as[52] a clinical[30] marker[16] of[52] disease[16], a major[30] contributor[15] to[52] the functional[30] deficit[16] that[44] results[19] from[52] those diseases[16], and[54] a therapeutic[30] target[16] in[52] the relief[16] of[52] symptoms[16] further[34] underlines[19]

the importance[15] of[52] the ideas[16] involved[47] for[52] students[16] and[54] researchers[16]. 28. Since[42] the turn[16] of[52] the century[16], the basic[30] model[16] of[52] capillary[16] fluid[16] exchange[16] driven[47] by[21] hydrostatic[30] and[54] oncotic[30] gradients[16] across[52] the capillary[16] wall[16] has[18] been generally[34] accepted[20]. 29. In[52] mathematical[30] form[16], this[5] predicts[19] that[44], for[52] a single[30] capillary[16], the fluid[16] flux[16] per[52] unit[16] surface[16] area[16] should[23] satisfy[29] the relationship[15] where[43] A is[25] a measure[16] of[52] capillary[16] fluid[16] permeability[15] known[47] as[52] hydraulic[30] conductivity[15]; and[55] B, C are[25] the hydrostatic[30] pressures[16] in[52] the capillary[16] and[54] surrounding[30] interstitium[16], respectively[34]; and[55] D, E are[25] the colloid[16] osmotic[30] or oncotic[30] pressures[16] exerted[47] by[21] protein[16] in[52] the plasma[16] and[54] interstitial[30] fluid[16], respectively[34] and[55] F is[25] the reflection[15] coefficient[16], a measure[16] of[52] how[43] closely[34] the capillary[16] wall[16] approximates[19] to[52] a perfect[30] semi-permeable[30] membrane[16] for[52] protein[16]. 30. Early[30] quantitative[30] studies[16], including[48] the classic[30] papers[16], were[25, 17] consistent[31] with[52] this model[16] if[41] it[4] was[17,25] assumed[20, 26] that[44] interstitial[30] hydrostatic[30] and[54] oncotic[30] pressures[16] were[25, 17] small[31]. 31. Measurements[15] of[52] interstitial[30] hydrostatic[30] and[54] oncotic[30] forces[16], however[35], suggested[17,28] that[44] there should[23] be[29] a considerable[30] filtration[15] gradient[16] along[52] the entire[30] length[16] of[52] most capillaries[16], even[34] in[52] tissues[16] that[44] are[25] in[52] fluid[16] balance[16]. 32. This[5] has led[18] to[52] proposed[30] modifications[15] of[52] the original[30] model[16], which[46] retain[19] this concept[16] of[52] hydrostatic[30] and[54] colloid[16] osmotic[30] pressures[16] as[55] the driving[30] forces[16] determining[48] capillary[16] fluid[16] exchange[16] but[54] emphasize[19] that[44] it[4] may[22] be[29] the values[16] of[52] these forces[16] within[52] very[37] specific[30] compartments[15] of[52] the extracapillary[16] space[16] that[44] matter[19], e.g.[35], just[38] outside[32] the endothelial[30] gylcocalyx[16] within[52] capillary[16] pores[16]. 33. Guiding[51] students[16] through[52] the key[16] stages[16] in[52] the development[15], testing[51], and[54] refinement[15] of[52] the hypothesis[16] provides[19] an excellent[30] case[16] study[16] in[52] the scientific[30] method[16]. 34. In[52] any[6] class[16] such as[52] this[5], the selection[15] and[54] sequence[16] of[52] the papers[16] is[25] crucial[31], and[55] each should[23] add[29] something[6] to[52] the development[15] of[52] the story[16]. 35. For[52] capillary[16] filtration[15], the papers[16] listed[47] provide[19] a useful[30] set[16] of[52] landmarks[16].

***Statistics*:** Types 1001, tokens 401, TTR 40, sentences 35, characters 5512, MWL 5,5 MSL 28.

Source: J. Graham McGeown (2006) Passing on the legacy: teaching capillary filtration and developing presentation skills using classic papers. *Advan. Physiol. Edu.* 30: 108-112; doi:10.1152/advan.00032.2006, http://advan.physiology.org/cgi/content/full/30/3/108

Appendix 7

1000-word sample of academic 'chat' text

The items with numbers above them have been encoded as linguistic features investigated in the present study. The numbers correspond to the numbers of linguistic features in Appendices 8-14. All names have been changed, nicknames removed.

_where[11] are[25] they[3] available[31]?
_you[2,7]re[25] from[52] Barry U[56][niversity]?
_Hello everyone[57]. Apologies[16] for[52] arriving[51] late[33]. Better late[33] than never[33] is[25] what[43] I[1] am shooting[19] for[52] everything these days[16] it[4] seems[19].
_Do we[1] also[34] get[19] to use[29] these los [learning objects]?
_It[4,7]s[25] like[52] a lot[38] of[52] our[1] traditional[30] classes[16], eh[39]?
_Swish[16] utilities[15] is[25] one component[16] the flash[16] [is] important[31] for[52] users[16] inexpert[30] [inexpert users].
_Yes[34] I[1] am[25] in[52] the School[16] of[52] Ed[56][ucation] at[52] Barry [University]. I[1] work[19] as[52] the instructional[30] designer[16], distance[30] edu[56][cation].
_Why[11] can[22,7]t[49] the user[16] setup[19] the puzzle[16] without[52] Flash[16]?
_Mary Are you[2] Dr. McGregor?
_where[11] you[2] can[22] get[29] this puzzle[16] game[16]?
_Absolutely[37] not[49] I[1]am[25] Dr,. Mary Tate-White, Dr. McGregor is[25] actually[34] "Maria"
_Why[11] can[22,7]t[49] they[3] use[29] an XML file[16]?
_A fine[30] professor[16] here[32] at[52] BU[56] [Barry University].
_I[1] see[26]... I[1,7]m[25] a PhD student[16] at[52] the school[16] of[52] Ed[ucation]. (Leadership[15] and Ed[56][ucation])
_mgermain99 do you[2] have[19] a name[16].[?]
_Mary Barry[56].[Mary answers that she is at Barry University]
_[It is] good[31] to meet[29] you[2] [.] are[25] you[2] here[32] at[52] BU or[54] where[43] [you are?].
_Ann- what[11] school[16] are[25] you[2] at[9] - I[1] am[25] also[34] [a] PhD student[16] at[52] UNL.
_[It is] Nice[31] to meet[29] you[2] too[34].
_[I] lost[17] audio[16].
_Dana: I[1,7]m[25] at[52] Barry University[16] in[52] Miami.
_For[52] that much[38] people[16], this[5] is[25] pretty[34] smooth[31].
_PLEASE[39] SEND[58] HER[3] URL[56].
_indeed[39] indeed[39].
_Ann what[11] do you[2] do[19] for[52] work[16] while[42] being[148] a student[16]?
_Where[11,7]s[25] the simulcast[16]?
_I[1] am[25] a Department[15] Chair[16] for[52] a small[30] college[16] in[52] Miami.
_that[5] is[25] pretty[34] good[31].
_the e learning[31] squares[16].
_Ann is[25] that[5] an online[30] program[16] or[54] f2f? I[1] am[25] in[52] Ed[56][ucation] Leadership[15] online[30] program[16].
_yes[34], do you[2] do[19] some[6] kind of[36] distance[16] ed[56][ducation] work[16].[?]
_What[11] is[25] the lag[16] time[16] here[32] [?] my[1] screen[16] hasn'[7]t[49] changed[18] in[52] 4 min.
_where[11] do i[1] go[19] to stop[29]/start[29] audio[16]?
_[Are there] any[6] Moodlers[16]?
_at[52] bu[56] [Barry University] we[1] have[19] online[30] courses[16] but[54] no[50] programs[16] yet[33].
_Dana: It[4] is[25] essentially[34] a f2f program[16]. I[1] don'[7]t[49] know[19] of[52] any[6] good[30] and[54] reputable[30] PhD program[16].
_is anyone[6] seeing[19] the screen[16] shots[16] that[44] they[3,7]re talking[19] about[9]. i[1] still[33] have[19] the puzzle[16] on[52] my[1] screen[16].
_At[52] Walden[16] they[3] seem[19] to have[29] a pretty[30] decent[30] program[16] Dana .

_The e-learning[51] squares[16] should[23] be[29] on[52] screen[16] now[33].
_Tom: yes[34].
_[I] Still[33] have[19] the puzzle[16] screen[16] shot[16].
_Hi[57], just[38] want[19] to know[29], what[43,7]s[25] a "Moodler[16]. "
_Can[22] you[2] minimize[19] the puzzle[16]?
_I[1] still[33] have[19] the text[16] puzzle[16] also[34].
_martinez, go[58] back[52] to[52] the original[30] browser[16] window[16] in[52] which[45] horizonlive[16] began[17]...
_someone[6] who[45] Moos[19].
_The college[16, 8] I[1] work[19] for[52] (City[16] College[16]) offers[19] a few online[30] courses[30], no[50] programs[16].
_i[1,7]ve lost[18] sound[16] and[54] still[33] have[19] text[16] puzzle[16] game[16].
_Minimize[58] the puzzle[16] window[16].
_horhe, the puzzle[16] window[16] might[22] be covering[29] the HorizonLive[16] windows[16]. Kill[58] the puzzle[16] screen[16] shot[16] by[52] closing[51] [it] if[41] you[2] like[19].
_Someone[6] using[48] the open[30] source[30] platform[16] Moodle[16], I[1] guess[26].
_i[1] don[7]t[49] have[19] that screen[16] anymore[34].. but[55] i[1,7]ll[24] try[29] it[4].
_Moos?
_Mary: I[1] am[25] a already[33] enrolled[31] in[52] University[16] of[52] Nebraska- Lincoln distance[30] PhD – it[4] is[25] a very[37] good[30] program[16].
_[I've] got[18] it[4]. thanks[39].
_There[32] you[2] go[19] Dana another one for[52] you[2].
_You[2] can[22] also[34] get[29] back[52] in[52] synch[ronous] if[41] you[2] go[19] to[52] the "lobby[16]" button[16] in[52] the middle[16] of[52] the screen[16] and[54] re-enter[19] the room[16].
_horhe, have[58] a colada!
_elearnig[51] squares[16] on[52] screen[16] now[33] but[54] [they are] static[31] and[55] [there is] no[50] sound[16] .
_lol[56][laughing out loud] ... [It] sounds[19] like[52] a great[30] idea[16] .
_[It's] Cool[31], Dana. Was[25] it[4] selective[31]?
_ahhh.. i[1] hit[17] 'lobby[16]' [but] lost[17] audio[16]!
_Are you[2] sharing[19] these objects[16] for[52] free[31]?
_Do we[1] as[52] participants[16] get[19] to use[29] some[6] of[52] these Los[56]?
_as[40] i[1,7]m[25] @[52] home[16] can[22] i[1] use[19] the learning[51] objects[16] later on[33] this week[16].[?]
_You[2] mentioned[19] small[30] groups[16]. What[11] size[16] is[25] good[31].[?] I[1] have found[18] over[52] 20 is[25] BAAAD[31]
_is[25] there any[6] Director[16] applications[5] w/learning[51] objects[16]?
_Is[25] the Flash[16] plug[16] in[52] free[31] for[52] students[16]?
_what[11] is[25] the process[16] for[52] accessing[51] and[54] using[51] Los[56].
_I[1] missed[17] a few minutes[16] changing[48] the video[16] for[52] my[1] daughter[16], but[55] did I[1] miss[17] how[43] this[5] connects[19] to[52] the WebCT database[16] /grade[16] book[16]?
_How[11] does the score[16] of[52] a learning[51] object[16] get[19] placed[47] into[52] the WebCT gradebook[16]?
_I[1] have[19] the stuff[16] so[55] [I've] never[33] thought[18] about[52] cost[16], but[55] I[1] know[19] that[44] we[1] have[19] to think[29] of[52] the end[16] user[16] everytime[33] we[1] develop[19] […] for[52] students[16].
_The flash[16] player[16] is[25] free[31], the full[30] Flash[16] development[15] system[16] costs[19] but[54] reasonable[30] education[15] rates[16].
(Thanks[39] John1, I[1] had[17] the same question[16]...)
_Thanks Bill.
_Benny, [it's] good[31] to "see[29]" you[2] again[33].
_very[38] frustrating[30] sound[16] comes[19] and[54] goes[19] .
_gotta go[56] – by[56] from[52] the UK.

_Hi[57] lawrence – [I] hope[19] to 'see'[29] you[2] at[52] the Digital[30] Games[16] Community[16] @[52] WebCT.com .
_Bye[57].
_Maria, I[1] believe[26] the program[16] was[25] pretty[37] selective[31].
_For[52] starters[16] what[11] type[16] of[52] "Flash[16]" are you[2] recommending[19]?
_bye[57] Mary.
_Dana: how[11] long is[25] the program[16]?
_Bye[57]!
_Yeah[39] […] is[25] a very[37] most[37] utilities[15] in[52] flash[16]
_Now[33] exist[s][19] the swish[16].
_The phonetic[30] tool[16] looks[19] like[52] the learning[30] object[16] (quiz[16]) [that][8] you[2] can[22] download[29] from[52] Flash[16] but[55] getting[51] the score[16] in[52] the gradebook[16] is[25] waht[43] I[1] need[19] to know[29].
_Anyone[6] using[48] swish[16]?
_newblie ? - does Flash[16] support[19] options[16] for[52] adaptability[15] for[52] students[16] with[52] learning[30] differences[16]?
_why[11] don'[7]t[49] you[2] want[19] to do[29] [to] Director[16] again[33]?
_I[1,7]ve used[18] Swish[16]—they[3] hav[e][56,19] a new[30] version[16] [of] SwishMAX.
_The program[16] can[22] be completed[20] in[52] 3 years[16] but[55] students[16] usually[33] take[19] a bit longer[33].
_Director[16] is[25] so[38] difficult[31], flash[16] is[25] easy[31] and[55] it[4] works[19]!
_What[11] kind[16] of[52] adaptability[15] are[25] you[2] particularly[34] interested[31] in[8]?
_Do you[2] all use[19] WebCT?
_yes.
_no[50].
_yes[34].
_Dana: is[25] there a required[30] residency[16]? (at[52] the campus[16]).
_yes[34].
_yes[34].
_yes[34].
_I[1] find[26] it[4] the opposite[31] - Director[16] seems[19] very[37] easy[31].
_yes[34].
_yes[34].
_not[49] yet[33].
_yes[34].
_yes[34].
_I[1] think[26] learning[51] objects[16] (reusable[30]) [are] a great[30] idea[16]. Are[25] there issues[16] about[52] copyright[16] and[54] adaptability[15]?
_are[25] the learning[51] objects[16] shown[47] here[32] for[52] sale[16] or[54] lease[16]?
_Well[39], any[6] programming[51] [that][8] I[1] have tried[18] is[25] extremely[37] difficult[31]. Flash[16] scripting[31] is[25] intuitive[31].
_not[49] flash[16] is[25] very[37] easy[31]
_How[11] do you[2] get[19] the little[30] smilies[16] into[52] your[2] message[16]?
_No[50] there is[25] not[49].
_No[50], upstate[32] New York.
_no[50], class[16] synchronous[30] size[16].
_I[1] am[25] unable[31] to access[25] the WebCT games[16] page[16] even[34] though[42] I[1] have registered[18] and[54] logged[18] in[9]?
_thank you[2] for[52] an interesting[30] presentation[15] I[1] must[23] go[29] now[33].
_I[1] keep losing[19] the audio[16].
_yes[34], thanks[39].

Statistics: TTR 39,3, MWL 4,5, MSL 7,4
***Sourse*:** Seminar "Avoiding Common Pitfalls with Your Growth of WebCT Campus Edition" chatroom October 22, 2003 WebCT http://www.webct.com/seminar/viewpage?name=seminar_archive

Appendix 8
Descriptive statistics for the sub-corpus of on-line seminars

	2 Code	3 Linguistic features	Mean	Min. value	Max. value	Sum	Standard deviation
1.	PRO1	First person pronouns	38.2	19.0	52.0	382.0	9.8
2.	PRO2	Second person pronouns	18.6	9.0	26.0	186.0	4.7
3.	PRO3	Third person pronouns	17.0	4.0	26.0	170.0	6.2
4.	IT	Pronoun *it*	14.0	10.0	19.0	140.0	2.9
5.	DEMP	Demonstrative pronouns	12.2	6.0	17.0	122.0	3.5
6.	INDP	Indefinite pronouns	7.5	4.0	11.0	75.0	2.4
7.	CONTR	Contractions	13.7	10.0	22.0	137.0	3.8
8.	THATD	Complementizer *that* deletion	1.1	0.0	3.0	11.0	1.2
9.	FPR	Stranded prepositions	1.8	0.0	4.0	18.0	1.5
10.	SAUX	Split auxiliaries	1.0	0.0	3.0	10.0	1.1
11.	WHQ	WH questions	0.6	0.0	2.0	6.0	0.8
12.	TTR	Type/token ratio	35.6	29.7	39.4	n/a	3.0
13.	MWL	Mean word length	4.6	4.1	4.8	n/a	0.2
14.	MSL	Mean syntactic length	18.2	14.4	22.1	n/a	2.2
15.	NOM	Nominalizations	25.7	17.0	35.0	257.0	5.2
16.	N	Nouns	194.6	164.0	237.0	1946.0	23.6
17.	PTV	Past tense verbs	2.4	0.0	6.0	24.0	2.0
18.	PAV	Perfect aspect verbs	2.5	1.0	8.0	25.0	2.2
19.	PRTV	Present tense verbs	24.3	18.0	38.0	243.0	6.4
20.	ALPASS	Agentless passives	5.6	3.0	11.0	56.0	2.5
21.	BYPASS	*By* passives	0.6	0.0	1.0	6.0	0.5
22.	PMOD	Possibility modals	7.5	4.0	13.0	75.0	2.9
23.	NMOD	Necessity modals	1.2	0.0	4.0	12.0	1.2
24.	PRMOD	Prediction modals	20.7	15.0	25.0	207.0	3.3
25.	BE	*Be* as main verb	35.7	22.0	52.0	357.0	8.0
26.	PVERB	Private verbs	6.5	3.0	12.0	65.0	3.0
27.	PUBV	Public verbs	8.6	5.0	16.0	86.0	3.3
28.	SUV	Suasive verbs	0.7	0.0	2.0	7.0	0.8
29.	INF	Infinitives	41.4	29.0	53.0	414.0	7.0
30.	ATADJ	Attributive adjectives	28.6	17.0	41.0	286.0	7.1
31.	PRADJ	Predicative adjectives	14.5	9.0	23.0	145.0	4.6
32.	PLADV	Place adverbials	5.6	3.0	9.0	56.0	1.9
33.	TADV	Time adverbials	11.8	5.0	18.0	118.0	3.5
34.	ADV	Other adverbs	9.7	6.0	15.0	97.0	2.7
35.	CONJ	Conjuncts	3.3	0.0	8.0	33.0	2.4
36.	HED	Hedges	3.4	2.0	6.0	34.0	1.5
37.	AMP	Amplifiers	2.0	1.0	4.0	20.0	1.1
38.	GENEM	General emphatics	0.9	0.0	3.0	9.0	1.0
39.	DPART	Discourse particles	4.1	3.0	6.0	41.0	1.0
40.	CSUB	Causative adverbial subordinator	1.4	0.0	4.0	14.0	1.5
41.	COND	Conditional adverbial subordinator	4.9	185 2.0	7.0	49.0	1.6
42.	ADVS	Other adverbial	1.1	0.0	3.0	11.0	1.0

		subordinator					
43.	WHC	WH clauses	10.3	7.0	14.0	103.0	2.4
44.	3.1.1.1	*That* relatives	13.6	8.0	18.0	136.0	3.3
45.	WHRCO	WH relatives on object position	3.0	1.0	5.0	30.0	2.0
46.	WHRCS	WH relatives on subject position	4.0	2.0	6.0	40.0	1.2
47.	PPC	Past participial clauses	3.3	1.0	7.0	33.0	1.9
48.	PRPCL	Present participle clauses	4.4	1.0	8.0	44.0	2.3
49.	ANEG	Analytic negation	5.6	1.0	11.0	56.0	3.0
50.	SYNEG	Synthetic negation	1.0	0.0	3.0	10.0	1.0
51.	GER	Gerunds	15.2	10.0	27.0	152.0	5.5
52.	NPR	Number of prepositions	112.9	89.0	140.0	1129.0	18.1
53.	SREL	Sentence relatives	1.3	0.0	4.0	13.0	1.4
54.	PHC	Phrasal coordination	12.9	9.0	21.0	129.0	3.9
55.	CLC	Clausal coordination	11.1	7.0	16.0	111.0	2.6

Appendix 9

Descriptive statistics for the sub-corpus of academic discussion forum messages

	4 Code	5 Linguistic features	Mean	Min. value	Max. value	Sum	Standard deviation
56.	PRO1	First person pronouns	39,6	29	50	396	6,1
57.	PRO2	Second person pronouns	12,6	8	20	126	3,7
58.	PRO3	Third person pronouns	17,2	12	22	172	3,3
59.	IT	Pronoun *it*	12,9	8	18	129	3,8
60.	DEMP	Demonstrative pronouns	8,4	5	13	84	2,6
61.	INDP	Indefinite pronouns	11,8	5	18	118	3,5
62.	CONTR	Contractions	13,3	10	17	133	2,6
63.	THATD	Complementizer *that* deletion	5	3	7	50	1,3
64.	FPR	Stranded prepositions	0,1	0	1	1	0,3
65.	SAUX	Split auxiliaries	4,5	1	8	45	2,1
66.	WHQ	WH questions	3,1	1	6	31	1,7
67.	TTR	Type/token ratio	43,97	37,8	49,2	439,7	4,0
68.	MWL	Mean word length	4,8	4,6	5,1	48	0,1
69.	MSL	Mean syntactic length	19,86	17	22,6	198,6	1,9
70.	NOM	Nominalizations	30,8	21	42	308	6,8
71.	N	Nouns	140,2	127	160	1402	11,2
72.	PTV	Past tense verbs	12,4	9	17	124	2,5
73.	PAV	Perfect aspect verbs	3,5	0	7	35	2,2
74.	PRTV	Present tense verbs	24,2	18	29	242	3,8
75.	ALPASS	Agentless passives	4,1	1	7	41	1,6
76.	BYPASS	*By* passives	1,8	0	4	18	1,2
77.	PMOD	Possibility modals	15,5	11	20	155	3,2
78.	NMOD	Necessity modals	2,5	1	6	25	1,7
79.	PRMOD	Prediction modals	12,5	8	17	125	2,7
80.	BE	*Be* as main verb	28,6	19	39	286	6,0
81.	PVERB	Private verbs	5,1	2	7	51	1,8
82.	PUBV	Public verbs	5,7	1	9	57	2,5
83.	SUV	Suasive verbs	0,2	0	1	2	0,4
84.	INF	Infinitives	53,4	40	66	534	7,7
85.	ATADJ	Attributive adjectives	44,4	38	56	444	5,4
86.	PRADJ	Predicative adjectives	18,8	13	25	188	3,4
87.	PLADV	Place adverbials	1,2	0	3	12	1,0
88.	TADV	Time adverbials	5,2	3	8	52	1,6
89.	ADV	Other adverbs	21	15	30	210	4,5
90.	CONJ	Conjuncts	5,7	2	9	57	2,2
91.	HED	Hedges	5	1	9	50	2,5
92.	AMP	Amplifiers	5,1	3	8	51	1,9
93.	GENEM	General emphatics	14,5	11	19	145	2,3
94.	DPART	Discourse particles	4,3	1	8	43	2,3
95.	CSUB	Causative adverbial subordinator	1,9	0	4	19	1,3
96.	COND	Conditional adverbial subordinator	7,7	3	12	77	2,9
97.	ADVS	Other adverbial	2,8	1	6	28	1,8

		subordinator					
98.	WHC	WH clauses	6,5	3	10	65	2,2
99.	5.1.1.1	*That* relatives	12,5	10	16	125	2,0
100	WHRCO	WH relatives on object position	3,8	1	6	38	1,4
101	WHRCS	WH relatives on subject position	0,2	0	1	2	0,4
102	PPC	Past participial clauses	5	1	7	50	1,7
103	PRPCL	Present participle clauses	1,7	0	2	17	0,7
104	ANEG	Analytic negation	8,6	6	15	86	3,0
105	SYNEG	Synthetic negation	2,8	1	5	28	1,6
106	GER	Gerunds	19,2	15	26	192	3,2
107	NPR	Number of prepositions	97,7	86	111	977	7,6
108	SREL	Sentence relatives	1,7	0	4	17	1,4
109	PHC	Phrasal coordination	25,6	17	35	256	5,6
110	CLC	Clausal coordination	7,8	4	11	78	2,1

Appendix 10

Descriptive statistics for the sub-corpus of academic synchronous discussions ('chats')

	6 Code	7 Linguistic features	Mean	Min. value	Max. value	Sum	Standard deviation
1.	PRO1	First person pronouns	51,1	47	57	511	3,4
2.	PRO2	Second person pronouns	8,8	4	14	88	3,4
3.	PRO3	Third person pronouns	21,5	16	27	215	3,8
4.	IT	Pronoun *it*	4,1	3	6	41	1,0
5.	DEMP	Demonstrative pronouns	8,4	5	13	84	2,6
6.	INDP	Indefinite pronouns	11,8	5	18	118	3,5
7.	CONTR	Contractions	13,3	10	17	133	2,6
8.	THATD	Complementizer *that* deletion	5	3	7	50	1,3
9.	FPR	Stranded prepositions	0,1	0	1	1	0,3
10.	SAUX	Split auxiliaries	4,5	1	8	45	2,1
11.	WHQ	WH questions	3,1	1	6	31	1,7
12.	TTR	Type/token ratio	43,97	37,8	49,2	439,7	4,0
13.	MWL	Mean word length	4,76	4,6	4,9	47,6	0,1
14.	MSL	Mean syntactic length	19,86	17	22,6	198,6	1,9
15.	NOM	Nominalizations	30,8	21	42	308	6,8
16.	N	Nouns	140,2	127	160	1402	11,2
17.	PTV	Past tense verbs	12,4	9	17	124	2,5
18.	PAV	Perfect aspect verbs	3,5	0	7	35	2,2
19.	PRTV	Present tense verbs	24,2	18	29	242	3,8
20.	ALPASS	Agentless passives	4,1	1	7	41	1,6
21.	BYPASS	*By* passives	2,1	1	4	21	1,1
22.	PMOD	Possibility modals	15,5	11	20	155	3,2
23.	NMOD	Necessity modals	2,5	1	6	25	1,7
24.	PRMOD	Prediction modals	12,5	8	17	125	2,7
25.	BE	*Be* as main verb	26	16	35	260	5,6
26.	PVERB	Private verbs	5,1	2	7	51	1,8
27.	PUBV	Public verbs	5,7	1	9	57	2,5
28.	SUV	Suasive verbs	0,2	0	1	2	0,4
29.	INF	Infinitives	53,4	40	66	534	7,7
30.	ATADJ	Attributive adjectives	44,4	38	56	444	5,4
31.	PRADJ	Predicative adjectives	18,8	13	25	188	3,4
32.	PLADV	Place adverbials	1,2	0	3	12	1,0
33.	TADV	Time adverbials	5,2	3	8	52	1,6
34.	ADV	Other adverbs	21	15	30	210	4,5
35.	CONJ	Conjuncts	5,7	2	9	57	2,2
36.	HED	Hedges	5	1	9	50	2,5
37.	AMP	Amplifiers	5,1	3	8	51	1,9
38.	GENEM	General emphatics	14,5	11	19	145	2,3
39.	DPART	Discourse particles	4,3	1	8	43	2,3
40.	CSUB	Causative adverbial subordinator	1,9	0	4	19	1,3
41.	COND	Conditional adverbial subordinator	7,7	3	12	77	2,9

42.	ADVS	Other adverbial subordinator	2,8	1	6	28	1,8
43.	WHC	WH clauses	6,5	3	10	65	2,2
44.	7.1.1.1	*That* relatives	12,5	10	16	125	2,0
45.	WHRCO	WH relatives on object position	3,8	1	6	38	1,4
46.	WHRCS	WH relatives on subject position	0,2	0	1	2	0,4
47.	PPC	Past participial clauses	5	1	7	50	1,7
48.	PRPCL	Present participle clauses	1,7	0	2	17	0,7
49.	ANEG	Analytic negation	14,6	10	19	146	3,1
50.	SYNEG	Synthetic negation	1,9	0	5	19	1,7
51.	GER	Gerunds	19,2	15	26	192	3,2
52.	NPR	Number of prepositions	97,7	86	111	977	7,6
53.	SREL	Sentence relatives	1,7	0	4	17	1,4
54.	PHC	Phrasal coordination	25,6	17	35	256	5,6
55.	CLC	Clausal coordination	7,8	4	11	78	2,1

Appendix 11
Descriptive statistics for the sub-corpus of academic e-mails

	8 Code	9 Linguistic features	Mean	Min. value	Max. value	Sum	Standard deviation
1.	PRO1	First person pronouns	45,1	37	64	451	10,2
2.	PRO2	Second person pronouns	25,2	16	32	252	5,8
3.	PRO3	Third person pronouns	8,7	4	13	87	2,8
4.	IT	Pronoun *it*	12,8	7	20	128	3,8
5.	DEMP	Demonstrative pronouns	7,5	4	11	75	2,5
6.	INDP	Indefinite pronouns	10,7	7	15	107	2,5
7.	CONTR	Contractions	14,1	10	19	141	3,0
8.	THATD	Complementizer *that* deletion	1,7	0	5	17	1,6
9.	FPR	Stranded prepositions	1,4	0	3	14	1,0
10.	SAUX	Split auxiliaries	5,4	2	9	54	2,2
11.	WHQ	WH questions	0,8	0	2	8	0,8
12.	TTR	Type/token ratio	33,48	30,2	36,2	n/a	2,3
13.	MWL	Mean word length	4,82	4,4	5,3	n/a	0,3
14.	MSL	Mean syntactic length	12,72	11,3	15,4	n/a	1,1
15.	NOM	Nominalizations	12	10	14	120	1,4
16.	N	Nouns	176,8	165	188	1768	7,5
17.	PTV	Past tense verbs	12,8	9	17	128	2,6
18.	PAV	Perfect aspect verbs	9,7	7	13	97	2,1
19.	PRTV	Present tense verbs	17,6	10	23	176	3,5
20.	ALPASS	Agentless passives	0,7	0	2	7	0,8
21.	BYPASS	*By* passives	1,5	0	4	15	1,4
22.	PMOD	Possibility modals	11,9	10	15	119	1,7
23.	NMOD	Necessity modals	0,6	0	2	6	0,8
24.	PRMOD	Prediction modals	8,9	4	14	89	3,2
25.	BE	*Be* as main verb	16,3	8	24	163	5,6
26.	PVERB	Private verbs	11,1	6	16	111	3,0
27.	PUBV	Public verbs	0,7	0	3	7	1,1
28.	SUV	Suasive verbs	0,3	0	1	3	0,5
29.	INF	Infinitives	27,6	21	34	276	3,7
30.	ATADJ	Attributive adjectives	44,2	38	50	442	4,0
31.	PRADJ	Predicative adjectives	13,8	10	18	138	3,0
32.	PLADV	Place adverbials	4,2	1	8	42	2,3
33.	TADV	Time adverbials	11,5	8	17	115	3,1
34.	ADV	Other adverbs	16,1	10	21	161	3,5
35.	CONJ	Conjuncts	1,6	0	3	16	1,1
36.	HED	Hedges	2,3	0	5	23	1,5
37.	AMP	Amplifiers	1,9	1	4	19	1,2
38.	GENEM	General emphatics	13,9	10	19	139	2,9
39.	DPART	Discourse particles	11	7	17	110	3,3
40.	CSUB	Causative adverbial subordinator	0,2	0	1	2	0,4
41.	COND	Conditional adverbial subordinator	4,9	2	7	49	1,9

42.	ADVS	Other adverbial subordinator	5,5	3	11	55	2,6
43.	WHC	WH clauses	6,5	4	8	65	1,2
44.	9.1.1.1	*That* relatives	5,5	3	9	55	2,0
45.	WHRCO	WH relatives on object position	7,5	4	13	75	3,1
46.	WHRCS	WH relatives on subject position	1,1	0	2	11	0,7
47.	PPC	Past participial clauses	2,2	0	4	22	1,3
48.	PRPCL	Present participle clauses	3,1	0	6	31	2,0
49.	ANEG	Analytic negation	5,7	3	11	57	2,6
50.	SYNEG	Synthetic negation	1,8	0	3	18	0,9
51.	GER	Gerunds	14	10	20	140	3,0
52.	NPR	Number of prepositions	119,4	99	140	1194	13,7
53.	SREL	Sentence relatives	0,1	0	1	1	0,3
54.	PHC	Phrasal coordination	13,6	6	19	136	4,1
55.	CLC	Clausal coordination	10	5	15	100	3,4

Appendix 12
Descriptive statistics for the sub-corpus of academic weblogs

	10 Code	11 Linguistic features	Mean	Min. value	Max. value	Sum	Standard deviation
1.	PRO1	First person pronouns	42,9	31	90	429	5,8
2.	PRO2	Second person pronouns	12	6	25	120	3,8
3.	PRO3	Third person pronouns	20,7	17	29	207	4,2
4.	IT	Pronoun *it*	10,5	6	15	105	2,9
5.	DEMP	Demonstrative pronouns	4,9	2	7	49	1,7
6.	INDP	Indefinite pronouns	7,7	3	14	77	3,8
7.	CONTR	Contractions	20,1	13	28	201	4,5
8.	THATD	Complementizer *that* deletion	10,5	6	17	105	3,7
9.	FPR	Stranded prepositions	2,2	0	6	22	2,2
10.	SAUX	Split auxiliaries	2,6	0	6	26	2,3
11.	WHQ	WH questions	2,4	0	7	24	2,4
12.	TTR	Type/token ratio	40,9	39,1	42,8	409	1,1
13.	MWL	Mean word length	4,87	4,7	5,4	48,7	0,2
14.	MSL	Mean syntactic length	20,58	17,3	27	205,8	2,9
15.	NOM	Nominalizations	12,8	9	18	128	2,9
16.	N	Nouns	181,6	162	201	1816	15,0
17.	PTV	Past tense verbs	37,5	27	44	375	5,9
18.	PAV	Perfect aspect verbs	11,1	7	15	111	2,3
19.	PRTV	Present tense verbs	28,1	19	33	281	4,0
20.	ALPASS	Agentless passives	1,1	0	3	11	1,0
21.	BYPASS	*By* passives	2,1	0	7	21	2,1
22.	PMOD	Possibility modals	3,4	0	8	34	2,6
23.	NMOD	Necessity modals	0,7	0	3	7	1,0
24.	PRMOD	Prediction modals	4,5	0	8	45	2,4
25.	BE	*Be* as main verb	20,2	15	34	202	2,3
26.	PVERB	Private verbs	10,7	5	16	107	3,6
27.	PUBV	Public verbs	4,4	0	8	44	2,6
28.	SUV	Suasive verbs	0,2	0	1	2	0,4
29.	INF	Infinitives	26,6	16	36	266	7,3
30.	ATADJ	Attributive adjectives	44,7	33	52	447	6,3
31.	PRADJ	Predicative adjectives	26,3	17	39	263	6,8
32.	PLADV	Place adverbials	3,5	0	8	35	2,4
33.	TADV	Time adverbials	8,2	2	14	82	3,9
34.	ADV	Other adverbs	42,4	33	53	424	6,2
35.	CONJ	Conjuncts	8,2	2	13	82	3,4
36.	HED	Hedges	0,9	0	5	9	1,7
37.	AMP	Amplifiers	7,8	3	14	78	3,8
38.	GENE	General emphatics	13	10	17	130	2,6

	M						
39.	DPART	Discourse particles	1,7	0	6	17	2,1
40.	CSUB	Causative adverbial subordinator	2,7	1	5	27	1,6
41.	COND	Conditional adverbial subordinator	1,7	0	3	17	1,2
42.	ADVS	Other adverbial subordinator	1,6	0	4	16	1,2
43.	WHC	WH clauses	7,1	3	11	71	2,6
44.	11.1.1.1	*That* relatives	14,9	7	20	149	4,3
45.	WHRCO	WH relatives on object position	2,6	0	5	26	1,7
46.	WHRCS	WH relatives on subject position	0,9	0	3	9	1,1
47.	PPC	Past participial clauses	6,7	3	11	67	2,7
48.	PRPCL	Present participle clauses	4,5	1	8	45	2,3
49.	ANEG	Analytic negation	4,7	1	15	47	2,5
50.	SYNEG	Synthetic negation	2,2	0	4	22	1,3
51.	GER	Gerunds	12,9	8	18	129	2,8
52.	NPR	Number of prepositions	96,6	89	102	966	4,3
53.	SREL	Sentence relatives	2,7	0	5	27	1,8
54.	PHC	Phrasal coordination	13,9	7	22	139	4,4
55.	CLC	Clausal coordination	12,9	9	18	129	2,9

Appendix 13
Descriptive statistics for the sub-corpus of academic hypertexts

	12 Code	13 Linguistic features	Mean	Min. value	Max. value	Sum	Standard deviation
1.	PRO1	First person pronouns	4,2	0	9	42	3,0
2.	PRO2	Second person pronouns	0	0	0	0	0,0
3.	PRO3	Third person pronouns	5,7	0	11	57	3,6
4.	IT	Pronoun *it*	5	2	9	50	2,1
5.	DEMP	Demonstrative pronouns	1,5	0	5	15	1,8
6.	INDP	Indefinite pronouns	0,5	0	2	5	0,8
7.	CONTR	Contractions	0	0	0	0	0,0
8.	THATD	Complementizer *that* deletion	0,1	0	1	1	0,3
9.	FPR	Stranded prepositions	0	0	0	0	0,0
10.	SAUX	Split auxiliaries	0,9	0	3	9	1,1
11.	WHQ	WH questions	0,3	0	2	3	0,7
12.	TTR	Type/token ratio	40,4	38,9	42,4	n/a	1,2
13.	MWL	Mean word length	5,47	4,7	6,2	n/a	0,5
14.	MSL	Mean syntactic length	28,78	26,5	32,7	n/a	1,8
15.	NOM	Nominalizations	45,7	35	57	457	6,5
16.	N	Nouns	239,1	199	259	2391	18,4
17.	PTV	Past tense verbs	4,1	2	6	41	1,4
18.	PAV	Perfect aspect verbs	10,2	7	15	102	2,3
19.	PRTV	Present tense verbs	40,9	37	52	409	4,5
20.	ALPASS	Agentless passives	36	27	42	360	4,8
21.	BYPASS	*By* passives	4,7	1	9	47	2,6
22.	PMOD	Possibility modals	6,1	2	9	61	2,8
23.	NMOD	Necessity modals	2,2	0	6	22	2,1
24.	PRMOD	Prediction modals	0,3	0	2	3	0,7
25.	BE	*Be* as main verb	13,7	9	19	137	3,8
26.	PVERB	Private verbs	0,8	0	3	8	1,1
27.	PUBV	Public verbs	1,3	0	3	13	1,1
28.	SUV	Suasive verbs	0,3	0	1	3	0,5
29.	INF	Infinitives	22,7	16	29	227	4,1
30.	ATADJ	Attributive adjectives	103,5	96	109	1035	4,6
31.	PRADJ	Predicative adjectives	10,2	5	17	102	3,9
32.	PLADV	Place adverbials	1,4	0	4	14	1,3
33.	TADV	Time adverbials	0,3	0	1	3	0,5
34.	ADV	Other adverbs	9,6	4	15	96	3,4
35.	CONJ	Conjuncts	7,8	4	12	78	2,6
36.	HED	Hedges	0,1	0	1	1	0,3
37.	AMP	Amplifiers	0,1	0	1	1	0,3
38.	GENEM	General emphatics	0,6	0	2	6	0,7
39.	DPART	Discourse particles	0	0	0	0	0,0
40.	CSUB	Causative adverbial subordinator	0,7	0	3	7	0,9
41.	COND	Conditional adverbial	0,6	0	2	6	0,8

		subordinator					
42.	ADVS	Other adverbial subordinator	5,3	3	7	53	1,3
43.	WHC	WH clauses	0	0	0	0	0,0
44.	13.1.1.1	*That* relatives	10,5	6	15	105	3,0
45.	WHRCO	WH relatives on object position	1,1	0	4	11	1,4
46.	WHRCS	WH relatives on subject position	2,1	1	4	21	1,0
47.	PPC	Past participial clauses	19,1	14	25	191	3,3
48.	PRPCL	Present participle clauses	13,2	9	18	132	3,3
49.	ANEG	Analytic negation	2,6	1	5	26	1,4
50.	SYNEG	Synthetic negation	0,5	0	2	5	0,7
51.	GER	Gerunds	10,5	7	15	105	2,5
52.	NPR	Number of prepositions	136,2	127	146	1362	6,6
53.	SREL	Sentence relatives	2	0	4	20	1,2
54.	PHC	Phrasal coordination	32,3	27	43	323	4,8
55.	CLC	Clausal coordination	8,8	6	13	88	2,1

Appendix 14

Sample output of data computer-processing with FREQUENCY(on-line seminars)

```
Input file was:       C:\Documents and Settings\RANGE32\10_000_transcr.txt
Outfile file was:     C:\Documents and Settings\RANGE32\semin_10000.txt

Total tokens:         9988
Total types:          1408

 ... in frequency order ...
```

Word Type	Rank	Frequency	Cumulative Percent
THE	1	568	5.69
AND	2	327	8.96
OF	3	307	12.03
TO	4	266	14.70
A	5	221	16.91
THAT	6	217	19.08
WE	7	206	21.15
IS	8	195	23.10
YOU	9	168	24.78
IN	10	160	26.38
ARE	11	154	27.92
IT	12	140	29.33
FOR	13	120	30.53
COURSE	14	112	31.65
THIS	15	112	32.77
I	16	104	33.81
HAVE	17	95	34.76
ON	18	77	35.53
AT	19	75	36.28
THEY	20	75	37.03
STUDENTS	21	73	37.77
S	22	68	38.45
SO	23	65	39.10
ABOUT	24	63	39.73
COURSES	25	62	40.35
WILL	26	59	40.94
NOT	27	56	41.50

WHAT	28	55	42.05
AS	29	52	42.57
DO	30	51	43.08
OR	31	51	43.59
WITH	32	51	44.10
FROM	33	48	44.58
HOW	34	47	45.05
IF	35	47	45.52
SEE	36	46	45.99
AN	37	45	46.44
CAN	38	45	46.89
THERE	39	45	47.34
T	40	44	47.78
OUR	41	43	48.21
MORE	42	41	48.62
SOME	43	41	49.03
VERY	44	41	49.44
ONE	45	40	49.84
BUT	46	39	50.23
CONTENT	47	39	50.62
AVAILABLE	48	37	50.99
JUST	49	37	51.36
WEBCT	50	37	51.73
LOOK	51	36	52.09
BE	52	35	52.44
USE	53	34	52.78
EXAMPLE	54	33	53.11
THEIR	55	32	53.43
GO	56	30	53.73
WAS	57	30	54.03
GOING	58	29	54.33
ALL	59	27	54.60
WITHIN	60	27	54.87
YOUR	61	25	55.12
OUT	62	24	55.36
PART	63	24	55.60
RUBRIC	64	23	55.83
VIDEO	65	23	56.06
GET	66	22	56.28

INSTANCE	67	22	56.50
INTO	68	22	56.72
KNOW	69	22	56.94
NEXT	70	22	57.16
WELL	71	22	57.38
BECAUSE	72	21	57.59
GIVE	73	21	57.80
HAS	74	21	58.01
HE	75	21	58.22
AGAIN	76	20	58.42
ALSO	77	20	58.62
DISCUSSION	78	20	58.82
LIKE	79	20	59.02
QUESTION	80	20	59.22
STUDENT	81	20	59.42
TODAY	82	20	59.62
WHO	83	20	59.82
YEAR	84	20	60.02
AM	85	19	60.21
LITTLE	86	19	60.40
THINK	87	19	60.59
THOSE	88	19	60.78
THROUGH	89	19	60.97
UP	90	19	61.16
HERE	91	18	61.34
INFORMATION	92	18	61.52
MASIE	93	18	61.70
TECHNOLOGY	94	18	61.88
TERMS	95	18	62.06
THEN	96	18	62.24
THESE	97	18	62.42
BACK	98	17	62.60
EXEMPLARY	99	17	62.77
FACULTY	100	17	62.94

Appendix 15

List of linguistic features in five textual dimensions

Dimension 1. ***Involved/ informational production***

31. PVERB private verbs 32. CONTR contractions 33. PRTV present tense verbs 34. PRO2 second person pronouns 35. ANEG analytic (*not*-)negation 36. DEMP demonstrative pronouns 37. GENEM general emphatics 38. PRO1 first person pronouns 39. IT pronoun *it* 40. BE *be* as main verb 41. CSUB causative subordination 42. DPART discourse particles 43. INDP indefinite pronouns 44. AMP amplifiers 45. SREL sentence relatives 46. WHQ 'Wh' questions 47. PMOD possibility modals 48. WHC 'Wh' clauses 49. FPR final (stranded) prepositions	50. TTR type/token ratio 51. PLADV Place adverbs 52. N nouns 53. MWL mean word length 54. NPR number of prepositions 55. ATADJ attributive adjectives

Dimension 2 ***Intergrated/fragmented*** information production

Features positively associated with intergrated information production	Features associated fragmented information production
9. PTV past tense verbs 10. PRO3 third person pronouns 11. PAV perfect aspect verbs 12. PUBV public verbs 13. SYNEG synthetic (*no*-)negation 14. PRPCL present participle clauses	15. PRTV present tense verbs 16. ATADJ attributive adjectives

Dimension 3. ***Explicit versus Situation-Dependent Reference***

Features positively associated with explicit reference	Features associated situation dependent reference
8. WHRCO 'Wh' relative clauses on object position 9. WHRCS 'Wh'relative clauses on subject position 10. PHC phrasal coordination 11. NOM nominalizations	12. TADV time adverbials 13. PLADV place adverbials 14. ADV other adverbs

Dimension 4. ***Overt expression of persuasion***

Features positively associated with overt expression of persuasion	No negative features
6. PRMOD Prediction modals 7. SUV Suasive verbs 8. COND Conditional subordination 9. NMOD Necessity modals 10. SAUX Split auxiliaries 11. INF Infinitives	

Dimension 5. ***Abstract Versus Non-abstract Information***

Features positively associated with abstract information production	Features associated with non-abstract information production
8. CONJ Conjuncts 9. ALPASS Agentless passives 10. PPC Past participial clauses 11. BYPASS By-passives 12. PRADJ Predicative adjectives	13. TTR type/token ratio

Appendix 16

Sorted frequency values for six CMAD types:
15 the most frequent and 15 the least frequent linguistic features in each CMAD type.

chats	
WHQ	1,935603
PRTV	1,704639
SYNEG	1,679483
BE	1,603905
TADV	1,599166
AMP	1,106814
PRO1	0,914156
PLADV	0,888558
PRO2	0,888112
DPART	0,834451
CONTR	0,613184
INDP	0,612331
FPR	0,511143
PAV	0,433106
IT	0,37783
BYPASS	-0,68855
SAUX	-0,69887
PPC	-0,70625
WHRCS	-0,7236
SREL	-0,8104
INF	-0,848
PUBV	-0,85773
NOM	-0,8953
MWL	-0,89924
PRMOD	-0,92718
GER	-1,11187
CONJ	-1,16623
NPR	-1,29335
RTHAT	-1,47245
MSL	-1,50746

semimars	
PRMOD	1,688389
WHRCS	1,638744
PUBV	1,398977
WHC	1,21701
DEMP	1,005688
IT	0,933883
SUV	0,795574
PLADV	0,770955
INF	0,716345
RTHAT	0,699416
BE	0,634102
HED	0,581406
PRO3	0,533554
COND	0,437724
CLC	0,429373
PRTV	-0,53554
WHQ	-0,58703
BYPASS	-0,59358
AMP	-0,5939
SAUX	-0,61823
SYNEG	-0,62257
THATD	-0,64302
PHC	-0,64928
ADV	-0,75543
ADVS	-0,81079
TTR	-0,8122
PTV	-0,83484
ATADJ	-0,85378
GENEM	-1,10893
PAV	-1,29932

weblogs	
PTV	2,033303
ADV	1,969128
THATD	1,681749
PRADJ	1,372606
PRO3	1,076898
CSUB	1,032356
CONJ	1,005369
CONTR	0,998206
CLC	0,994889
AMP	0,971837
PVERB	0,956631
RTHAT	0,93869
SREL	0,921413
PAV	0,829089
GENEM	0,789849
INF	-0,42981
NPR	-0,46111
PRMOD	-0,46248
DEMP	-0,48246
NMOD	-0,49629
BE	-0,5222
HED	-0,52253
ALPASS	-0,53361
PHC	-0,53504
MWL	-0,54252
ADVS	-0,59742
COND	-0,63975
DPART	-0,743
NOM	-0,75452
PMOD	-1,17808

emails	
WHRCO	1,529997
DPART	1,383131
SAUX	1,155819
ADVS	1,066826
PVERB	1,049583
PRO2	1,012034
GENEM	0,931081
IT	0,677243
PMOD	0,648122
NPR	0,564577
INDP	0,527384
PAV	0,482604
COND	0,437724
PRO1	0,416999
TADV	0,258689
ALPASS	-0,56447
PHC	-0,56931
AMP	-0,6209
PPC	-0,65882
PRO3	-0,6853
MSL	-0,74679
CONJ	-0,76408
CSUB	-0,76826
PUBV	-0,77517
RTHAT	-0,79144
SREL	-0,8104
BE	-0,81314
NOM	-0,8138
PRTV	-1,00458
TTR	-1,31209

discussions	
INF	1,645657
PMOD	1,421571
COND	1,380515
HED	1,287924
GER	1,237737
TTR	1,200555
DEMP	1,148386
GENEM	1,025235
ANEG	0,888204
PHC	0,801604
SAUX	0,792945
INDP	0,76099
NMOD	0,644126
PUBV	0,600872
PRMOD	0,599677
ATADJ	-0,22877
PRO2	-0,28915
ALPASS	-0,30216
PVERB	-0,3447
NPR	-0,41162
PRTV	-0,54254
CLC	-0,60741
TADV	-0,63026
PRPCL	-0,77074
WHRCS	-0,78745
FPR	-0,82227
PLADV	-0,95389
PAV	-1,05183
IT	-1,18339
N	-1,33685

hypertexts	
ALPASS	2,158854
ATADJ	2,109063
PPC	2,013345
PRPCL	1,861988
NOM	1,683167
N	1,648801
MWL	1,597828
PHC	1,578454
MSL	1,558173
BYPASS	1,353351
NPR	1,329343
ADVS	0,98148
CONJ	0,898129
PRTV	0,647553
PAV	0,606349
BE	-1,0071
COND	-1,01013
PRMOD	-1,02011
AMP	-1,10681
PRO3	-1,12585
DPART	-1,13165
GENEM	-1,156
DEMP	-1,17557
TADV	-1,32166
PVERB	-1,34393
PRO2	-1,59034
WHC	-1,61048
INDP	-1,63878
PRO1	-1,79319
CONTR	-1,86807

Appendix 17
List of Analysed Academic Hypertexts
(Articles published in electronic journals)

1. J. Graham McGeown Passing on the legacy: teaching capillary filtration and developing presentation skills using classic papers *Advances in Physiological Education* 30: 108-112, 2006; retrieved on 10th October 2006 from http://advan.physiology.org/
2. Gregory A. Brown Teaching skeletal muscle adaptations to aerobic exercise using an American Physiological Society classic paper by Dr. Philip Gollnick and colleagues *Advan. Physiol. Edu*. 30: 113-118, 2006, retrieved on 10th October 2006 from http://advan.physiology.org/
3. Harry R. Goldberg, Eileen Haase, Artin Shoukas, and Lawrence Schramm Redefining classroom instruction
Advan. Physiol. Edu. 30: 124-127, 2006, retrieved on 10th October 2006 from http://advan.physiology.org/
4. Michael D. Bentley, Maria C. Ortiz, Erik L. Ritman, and J. Carlos Romero The use of microcomputed tomography to study microvasculature in small rodents *AJP - Regulatory, Integrative and Comparative Physiology*, 282: R1267-R1279, 2002. Retrieved on 10th October, 2006 from http://ajpregu.physiology.org/content/vol282/issue5/
5. Sharon Judge, Young Mok Jang, Anthony Smith, Colin Selman, Tracey Phillips, John R. Speakman, Tory Hagen, and Christiaan Leeuwenburgh Exercise by lifelong voluntary wheel running reduces subsarcolemmal and interfibrillar mitochondrial hydrogen peroxide production in the heart *AJP- Regulatory, Integrative and Comparative Physiology* 289: R1564-R1572, 2005. First published July 28, 2005; retrieved on 10th October, 2006 from http://ajpregu.physiology.org/cgi/content/full/289/6/R1564
6. Tim Edwards 'Generalised Swan modules and the D(2) problem' *Algebraic & Geometric Topology* 6 (2006) 71–89. Retrieved on 30th September from http://www.emis.de/journals/UW/agt/ftp/main/2006/agt-06-02.pdf
7. Dietrich Notbohm, Nigel Ray On Davis-Januszkiewicz homotopy types I; formality and rationalisation *Algebraic and Geometric Topology* 5 (2005), paper no. 3, pages 31-51. Retrieved on 30th September, 2006 from http://www.emis.de/journals/UW/agt/ftp/main/2005/agt-5l3.pdf
8. Dietrich Burde Left-symmetric algebras, or pre-Lie algebras in geometry and physics *Central-European Journal of Mathematics* Volume 4, Number 3 / September, 2006: 323- 357; 10.2478/s11533-006-0014-9 Retrieved on 30th September, 2006 from http://www.springerlink.com/content/u31168v17010w00g/fulltext.pdf
9. Paul Loya; Jinsung Park On gluing formulas for the spectral invariants of Dirac type operators *Electronic Research Announcements of the American Mathematic Society* 11, 2005:1-11. Retrieved on 30th September, 2006 from http://www.ams.org/era/2005-11-01/S1079-6762-05-00141-1/S1079-6762-05-00141-1.pdf
10. Thomas Schank Dorothea Wagner Approximating Clustering Coefficient and Transitivity *Journal of Graph Algorithms and Applications* vol. 9, no. 2, 2005: 265–275. Retrieved on 30th September from http://jgaa.info/
11. Yang T, Sauve AA. NAD Metabolism and Sirtuins: Metabolic Regulation of Protein Deacetylation in Stress and Toxicity. *AAPS Journal.* 2006; 8(4): E632-E643. Retrieved on 30th September, 2006 from http://www.aapspharmsci.org/view.asp?art=aapsj080472

12. Sandeep Kumar,Rajesh Kumar, Ranvinder Singh, Rakesh Kumar, Awdhesh Kumar Shukla, V.K. Jindal and Lalit M. Bharadwaj Binding of Carbon Nanotubes Dispersed by Optical Tweezer on Silicon Surface AZojono *Journal of Nanotechnology Online* June 10th, 2006. Retrieved on 30th September, 2006 from http://www.azonano.com/nanotechnology.asp
13. Christiane Goerke, Johanna Köller, and Christiane Wolz Ciprofloxacin and Trimethoprim Cause Phage Induction and Virulence Modulation in Staphylococcus aureus *Antimicrob Agents Chemother.* 2006 January; 50(1): 171–177. Retrieved on 30th September, 2006 from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1346766
14. Nicola G. A. Abrescia, Hanna M. Kivelä, Jonathan M. Grimes, Jaana K. H. Bamford, Dennis H. Bamford and David I. Stuart Preliminary crystallographic analysis of the major capsid protein P2 of the lipid-containing bacteriophage PM2 *Structural Biology and Crystallization Communications Online* Volume 61 Part 8 Pages 762-765 August 2005. Retrieved on 30th September, 2006 from http://journals.iucr.org/f/issues/2005/08/00/en5116/index.html
15. Tomoko Adachi, Atsushi Izumi, Dean Rea, Sam-Yong Park, Jeremy R. H. Tame and David I. Roper 'Expression, purification and crystallization of 2-oxo-hept-4-ene-1,7-dioate hydratase (HpcG) from Escherichia coli C' *Structural Biology and Crystallization Communications Online* Volume 62 Part 10 Pages 1010-1012 October 2006 http://journals.iucr.org/f/services/readerservices.html
16. S. K. Srivastava *, M. Pramanik, H. Acharya Ethylene/vinyl acetate copolymer/clay nanocomposites *Journal of Polymer Science Part B: Polymer Physics* Volume 44, Issue 3, Pages 471-480 Published Online: 16 Dec 2005. Retrieved on 30th September, 2006 from http://www3.interscience.wiley.com/cgi-bin/jabout/36698/HighlightedPapers2006.html
17. Shigeru Yamago Development of organotellurium-mediated and organostibine-mediated living radical polymerization reactions *Journal of Polymer Science Part A: Polymer Chemistry* Volume 44, Issue 1, Pages 1-12 Published Online: 10 Nov 2005 http://www3.interscience.wiley.com/cgi-bin/fulltext/112141174/HTMLSTART
18. Luke J. Engelking, Bret M. Evers, James A. Richardson, Joseph L. Goldstein, Michael S. Brown, and Guosheng Liang Severe facial clefting in Insig-deficient mouse embryos caused by sterol accumulation and reversed by lovastatin *The Journal of Clinical Investigations*, 2006; N 116 (9): 2356–2365. Retrieved on 30th September, 2006 from http://www.pubmedcentral.nih.gov/tocrender.fcgi?action=archive&journal=120
19. te Beek LAM, van der Werf MJ, Richter C, Borgdorff MW. Extrapulmonary tuberculosis, by nationality, the Netherlands 1993–2001. *Emerg Infect Dis* [serial on the Internet]. 2006 Sep [date cited]. Available from http://www.cdc.gov/ncidod/EID/vol12no09/05-0553.htm
20. Gottfreðsson M, Diggle MA, Lawrie DI, Erlendsdóttir H, Harðardóttir H, Kristinsson KG, et al. Neisseria meningitidis sequence type and risk for death, Iceland. Emerg Infect Dis [serial on the Internet]. 2006 Jul [date cited]. Available from http://www.cdc.gov/ncidod/EID/vol12no07/05-1624.htm
21. Amany Saleh, Ph.D, Marcia Lamkin, Ed.D., and David Cox, Ed.D.The Role of Higher Education in America: A Spa or a Smörgåsbord? *Academic Leadership – The Online Journal* Vol. 4 (3) 2006. Retrieved on 30th September, 2006 from http://www.academicleadership.org/volume4/issue3/index.html
22. Szafron, D., Carbonaro, M., Cutumisu, M., Gillis, S., McNaughton, M., Onuczko, C., Roy, T., and J. Schaeffer (2005) Writing Interactive Stories in the Classroom. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning* (USA),

Vol. 7, No 1, retrieved on 9th September, 2006 from http://imej.wfu.edu/articles/2005/1/02/index.asp

23. Charles ROSS Underwater Women in Shakespeare Films *Comparative Literature and Culture A WWWeb Journal* Vol. 6 (1) 2004, ISSN 1481-4374 http://clcwebjournal.lib.purdue.edu/
24. Oliver Warren, James Kinross, Paraskevas Paraskeva, Ara Darzi Emergency laparoscopy – current best practice *World Journal of Emergency Surgery* 2006, 1:24. Retrieved on 30th September, 2006 from http://www.wjes.org/content/1/1/24
25. Kelly Edmonds Off with their heads! Copyright infringement in the Canadian online higher educational environment *Canadian Journal of Learning and Technology* 2006, Vol. 32(2) Retrieved on 30th September, 2006 from http://www.cjlt.ca/content/vol32.2/edmonds.html
26. Ascuena, A. and M. Mattison, '(Re)Wiring Ourselves: The Electrical and Pedagogical Evolution of a Writing Center *Computers and Composition Online, Theory into Practice* Spring 2006 http://www.bgsu.edu/cconline/ascuenamattison_rewiring/ascuenamattison_rewiring.htm
27. Whithaus, C. 'Think Different/Think Differently: A Tale of Green Squiggly Lines, or Evaluating Student Writing in Computer- Mediated Environments', *Kairos*, Vol 7 (10), 2002 Retrieved on 30th September, 2006 from http://english.ttu.edu/Kairos/7.x
28. Joseph Claudet A Multimedia Approach to Enhancing School Leaders' Reflective Thinking and Decision Making *Interactive Educational Multimedia*, Number 13, 2006: 1-10. Retrieved on 18th September, 2006 from http://www.ub.edu/multimedia/iem
29. Dong-Shin Shin 'ESL Students' Computer-Mediated Communication Practices : Context Configuration' *Language Learning & Technology*, Vol.10, No.3, 2006: 65-84 Retrieved on 18th September, 2006 from http://llt.msu.edu/vol10num3/shin/default.html
30. Christine Hult and Ryan Richins 'The Rhetoric and Discourse of Instant Messaging' *Computers and Composition Online, Theory into Practice* Spring 2006. Retrieved on 18th September, 2006 from http://www.bgsu.edu/cconline/hultrichins_im/hultrichins_im.htm