

the compilation of a corpus that is able to capture the studies, diachronic corpus linguistics is a very promising to have access to a corpus as a representative sample advances in English corpus linguistics include the following between a monitor corpus for lexicographical description was observed in the corpus of Old English texts on the iterative analysis of corpus data may yield interesting the kind of reference corpus is represented by the British

Ewa Jonsson

Conversational Writing

A Multidimensional Study of Synchronous and Supersynchronous Computer-Mediated Communication

st it out on a test corpus of 50,000 words of spontaneous the influence of the corpus revolution on applied linguistics composition of this corpus shows its representativeness the compilation of a corpus that is able to capture the studies, diachronic corpus linguistics is a very promising to have access to a corpus as a representative sample advances in English corpus linguistics include the following between a monitor corpus for lexicographical description was observed in the corpus of Old English texts on the iterative analysis of corpus data may yield interesting the kind of reference corpus is represented by the British st it out on a test corpus of 50,000 words of spontaneous the influence of the corpus revolution on applied linguistics composition of this corpus shows its representativeness the compilation of a corpus that is able to capture the studies, diachronic corpus linguistics is a very promising to have access to a corpus as a representative sample advances in English corpus linguistics include the following between a monitor corpus for lexicographical description was observed in the corpus of Old English texts on the iterative analysis of corpus data may yield interesting the kind of reference corpus is represented by the British st it out on a test corpus of 50,000 words of spontaneous the influence of the corpus revolution on applied linguistics composition of this corpus shows its representativeness the compilation of a corpus that is able to capture the studies, diachronic corpus linguistics is a very promising to have access to a corpus as a representative sample

The author analyses computer chat as a form of communication. While some forms of computer-mediated communication (CMC) deviate only marginally from traditional writing, computer chat is popularly considered to be written conversation and the most “oral” form of written CMC. This book systematically explores the varying degrees of conversationality (“orality”) in CMC, focusing in particular on a corpus of computer chat (synchronous and supersynchronous CMC) compiled by the author. The author employs Douglas Biber’s multidimensional methodology and situates the chats relative to a range of spoken and written genres on his dimensions of linguistic variation. The study fills a gap both in CMC linguistics as regards a systematic variationist approach to computer chat genres and in variationist linguistics as regards a description of conversational writing.

Ewa Jonsson is a researcher in English linguistics at Mid-Sweden University.

Conversational Writing

ENGLISH CORPUS LINGUISTICS

Thomas Kohonen / Joybrato Mukherjee (eds.)

VOLUME 16



PETER LANG
EDITION

Ewa Jonsson

Conversational Writing

A Multidimensional Study of
Synchronous and Supersynchronous
Computer-Mediated Communication



PETER LANG
EDITION

Bibliographic Information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the internet at <http://dnb.d-nb.de>.

Library of Congress Cataloging-in-Publication Data

Names: Jonsson, Ewa, 1968- author.

Title: Conversational writing : a multidimensional study of synchronous and supersynchronous computer-mediated communication / Ewa Jonsson.

Description: Frankfurt am Main ; New York : Peter Lang, [2016] | Series:

English Corpus Linguistics ; Volume 16 | Includes bibliographical references.

Identifiers: LCCN 2016000923 | ISBN 9783631671535 (Print) | ISBN 9783653065121 (E-Book)

Subjects: LCSH: Conversation analysis--Data processing. | Authorship--Computer network sources. | Digital communications. | Online social networks.

Classification: LCC P95.45 .J66 2016 | DDC 302.2/24--dc23 LC record available at <http://lcn.loc.gov/2016000923>

Published with financial support from Mid-Sweden University.

This book is an open access book and available on www.oapen.org and www.peterlang.com. It is distributed under the terms of the Creative Commons Attribution Noncommercial, No Derivatives (CC-BY-NC-ND).



ISSN 1610-868X

ISBN 978-3-631-67153-5 (Print)

E-ISBN 978-3-653-06512-1 (E-Book)

DOI 10.3726/978-3-653-06512-1

© Peter Lang GmbH

Internationaler Verlag der Wissenschaften

Frankfurt am Main 2015

Alle Rechte vorbehalten.

All rights reserved.

Peter Lang Edition is an Imprint of Peter Lang GmbH.

Peter Lang – Frankfurt am Main · Bern · Bruxelles · New York ·
Oxford · Warszawa · Wien

All parts of this publication are protected by copyright. Any utilisation outside the strict limits of the copyright law, without the permission of the publisher, is forbidden and liable to prosecution. This applies in particular to reproductions, translations, microfilming, and storage and processing in electronic retrieval systems.

This publication has been peer reviewed.

www.peterlang.com

Acknowledgments

Without generous support from the English section of the Department of Humanities at Mid-Sweden University this book might not have seen the light of day. Thus, I wish to express my deep gratitude to my colleagues there for encouraging its publication, especially Terry Walker and Anders Olsson. The book is based on my doctoral thesis, written and presented at Uppsala University. I would like to acknowledge my Uppsala University colleagues for their constructive advice and scholarly support during the writing process. Most importantly, I am grateful to my doctoral supervisors Merja Kytö and Christer Geisler in this respect. Thank you all for your professional assistance and inspiration.

Preface

This book presents a linguistic investigation of two genres of computer-mediated communication (CMC), namely two modes of conversational writing: “Internet relay chat” (synchronous CMC) and “split-window ICQ chat” (supersynchronous CMC). The investigation employs Douglas Biber’s multifeature multidimensional methodology, taking into account the six dimensions of textual variation in English identified in his 1988 book *Variation across speech and writing*.

The book came about as an attempt to disentangle my puzzlement in the early 21st century with some fellow university students’ frequent propensity to prefer written conversation (computer chat) to spoken conversation. I was a member of the board of the university’s computer society and one of few in the society from outside the technological sphere. At board meetings, I noticed a reluctance among board members to sit down and discuss face-to-face. It seemed as if the members had a lack of practice and rather wished to meet and discuss in chat room channels or in Unix Talk. Occasionally, items on the agenda were left unfinished or postponed to discussions in the online environments, and several board members appeared to be more comfortable conversing in writing.

I became curious about the board members’ choice of modality – opting for writing instead of speech. Much like the interlocutors in social media today, they appeared to feel safer in the graphemic interface, while still being able to solve issues of the computer society efficiently because of the real-time communication. Conversation in writing seems to filter away a number of cues that users potentially find threatening in face-to-face communication. If I was a psychologist, I might have embarked on a study involving in-depth interviews with chat room users like the board members, but since I am a linguist, I decided to limit my scope to the language communicated in each respective medium.

Questions that I address in this book are what the most salient linguistic features of computer chat are, how synchronous writing is similar to speech and how written conversations differ from spoken conversations. My study does not involve any of the individuals described above, but chat room conversationalists in international, public channels (for synchronous chat) and adolescents in an English-speaking country (for supersynchronous chat). The multidimensional methodology chosen for the investigation identifies, among other things, the most salient linguistic features of their computer chats (features conspicuous either by their high relative frequency or by their relative rarity), and the procedure of positioning the two genres represented by the chats on Biber’s (1988)

dimensions enables a systematic lexico-grammatical description of the genres relative to other genres of writing, and speech.

Although none of Biber's (1988) dimensions constitutes a dichotomous distinction between writing and speech, they all differentiate among literate and oral genres in various respects. Among the genres studied by Biber are face-to-face and telephone conversations. By relating the CMC genres to the oral conversational genres on the dimensions, it is possible to assess the degree of orality in computer-mediated conversational writing, another undertaking of the study. The investigation presented here considers previous assumptions that synchronously mediated texts display more speech-like properties than asynchronous texts, and discusses whether supersynchronously mediated conversational writing texts are more speech-like than synchronously mediated ones.

The study further employs M. A. K. Halliday's model of semiotics, among other reasons to explain differences in the outcome of subtly divergent communicative settings, and argues for the inclusion of Halliday's measure of lexical density in studies of linguistic variation involving conversational writing. Finally, two features not included in Biber's (1988) methodology are here found to be particularly indicative of conversational writing texts: inserts, specified in Biber et al.'s (1999) *Longman grammar of spoken and written English*, and "emotives" (comprising emoticons and sentiment initialisms), a feature introduced in this study.

Why, then, is it important to study conversational writing genres from such an in-depth linguistic point of view? Firstly, linguistic research has found register/genre variation to be a fundamental aspect of human language. Biber & Conrad (2009: 23) note that "all humans control a range of registers/genres" and that "[g]iven the ubiquity of register/genre variation, an understanding of how linguistic features are used in patterned ways across text varieties is of central importance for both the description of particular languages and the development of cross-linguistic theories of language use." Biber & Conrad call register/genre variation a linguistic universal. In the light of this, a study of conversational writing genres is as natural, relevant and important as the study of other genres of language. Variationists aim to describe language adequately, to enable the comparison across genres, to map out language users' competence and to eventually facilitate, for instance, cross-linguistic comparisons. A thorough description of conversational writing may in turn facilitate the development of computational tools for automatic genre classification, editing and translation, as well as the development of new software for digital communication. And last but not least, it may lend a clue to psychologists' and sociologists' investigation of people's motivations for opting for written, rather than spoken, conversations.

Table of Contents

Tables	13
Figures	15
Abbreviations	17
Chapter 1. Introduction	19
1.1 Speech vs. writing vs. conversational writing.....	19
1.2 Aim and scope of the study.....	23
1.3 Synchronicity of communication.....	30
1.4 Notes on terminology.....	34
1.5 Outline of the study.....	37
Chapter 2. Background	39
2.1 Introductory remarks.....	39
2.2 Survey of the literature on speech and writing.....	39
2.3 Biber's (1988) dimensions of textual variation.....	50
2.4 Halliday's and others' essentially qualitative approaches.....	60
2.5 Survey of the literature on CMC.....	66
2.6 Description of the media for conversational writing.....	77
2.7 Chapter summary.....	81
Chapter 3. Material and method	83
3.1 Introductory remarks.....	83
3.2 Creating and annotating a corpus of Internet relay chat.....	84

3.3	Creating and annotating a corpus of split-window ICQ chat.....	93
3.4	The Santa Barbara Corpus subset.....	96
3.5	Standardization and dimension score computation	100
3.6	Average figures for writing and speech, respectively	103
3.7	Chapter summary.....	106
	Chapter 4. Salient features in conversational writing	109
4.1	Introductory remarks.....	109
4.2	Distribution of modal auxiliary verbs and personal pronouns.....	111
4.3	Word length, type/token ratio and lexical density.....	130
4.4	The most salient features	149
4.5	Paralinguistic features and extra-linguistic content.....	172
4.6	Inserts and emotives	189
4.7	Chapter summary.....	204
	Chapter 5. Conversational writing positioned on Biber's (1988) dimensions	205
5.1	Introductory remarks.....	205
5.2	Dimension plots	211
5.2.1	Dimension 1: Informational versus Involved Production.....	212
5.2.2	Dimension 2: Narrative versus Non-Narrative Concerns.....	222
5.2.3	Dimension 3: Explicit/Elaborated versus Situation-Dependent Reference.....	228
5.2.4	Dimension 4: Overt Expression of Persuasion/Argumentation.....	234
5.2.5	Dimension 5: Abstract/Impersonal versus Non-Abstract/Non-Impersonal Information.....	241
5.2.6	Dimension 6: On-Line Informational Elaboration	247
5.3	Chapter summary.....	254

Chapter 6. Discussion	255
6.1 Introductory remarks.....	255
6.2 Hypotheses revisited quantitatively.....	256
6.3 From genres to text types.....	263
6.4 Research questions revisited	275
6.5 Chapter summary.....	289
Chapter 7. Conclusion	291
7.1 Summary of the study.....	291
7.2 Suggestions for further research	295
Appendices	297
Appendix I. Texts used in Biber's (1988) study	297
Appendix II. Descriptive statistics for genres studied.....	299
Appendix III. Raw frequencies of linguistic features.....	313
Appendix IV. Examples of excluded material	319
Appendix V. Features with a standard score >2.0	320
Appendix VI. Statistical tests of salient features.....	322
Appendix VII. Word lists for the corpora studied.....	323
Appendix VIII. Dimension score statistics for Biber's (1988) genres.....	325
Appendix IX. Computation of cluster affiliations.....	330
Appendix X. Dimension scores for individual texts	332
List of References	333

Tables

Table 1.1:	Principal synchronicity and direction of communication in various genres.....	31
Table 2.1:	Linguistic features studied in Biber (1988)	51
Table 2.2:	Summary of co-occurring features on each dimension	53
Table 2.3:	Halliday’s three metafunctions in language and related concepts.....	64
Table 3.1:	Size of corpora compiled/sampled and annotated for the present study	83
Table 3.2:	Tags used in the annotation of the first twelve turns in Internet relay chat text 4a (UCOW).....	88
Table 4.1:	Frequencies of possibility, necessity and prediction modals per 1,000 words.....	114
Table 4.2:	Frequencies of first, second and third person pronouns per 1,000 words.....	122
Table 4.3:	Type/token ratio, with standard deviation	135
Table 4.4:	Unweighted lexical density for five corpora.....	140
Table 4.5:	Unweighted lexical density per clause and related measures.....	146
Table 4.6:	Frequencies per 1,000 words for the most salient linguistic features	150
Table 4.7:	Frequencies of inserts.....	190
Table 4.8:	Frequencies of emotives.....	190
Table 4.9:	Examples of turns with inserts in the three annotated corpora.....	194
Table 4.10:	Individuals’ emotives usage in the split-window ICQ corpus, by gender.....	202
Table 5.1:	Descriptive dimension statistics for the UCOW genres and the SBC subset.....	206

Table 5.2	Results from ANOVA among the new genres and from Biber's (1988: 127) tests among his genres.....	208
Table 5.3:	Results from t-tests among the new genres.....	208
Table 5.4:	Summary of co-occurring features on each dimension.....	210
Table 5.5:	Corrected dimension scores for the "ELC other" corpus of BBS conferencing presented in Collot (1991).....	211
Table 6.1:	Distance of the three CMC genres to oral conversations measured as standard deviation units on each dimension.....	258
Table 6.2:	Distance of the conversational writing genres to oral conversations indicated as t-values on each dimension.....	259
Table 6.3:	Results from t-tests among the conversational writing genres and the conversational spoken genres.....	261
Table 6.4:	Summary of English text types.....	267

Figures

Figure 1.1:	Examples of asynchronous, synchronous and supersynchronous modes of written CMC.....	27
Figure 1.2:	Working relationship between modalities, media and genres/modes in the present study	29
Figure 2.1:	Metafunctions in relation to register and genre in semiotics.....	65
Figure 2.2:	Approximate emergence of modes for written CMC.....	67
Figure 2.3:	Screenshot of Internet relay chat window (SCMC).....	79
Figure 2.4:	Screenshot of split-window ICQ chat (SSCMC).....	80
Figure 4.1:	Distribution of possibility, necessity and prediction modals per 1,000 words.....	114
Figure 4.2:	Distribution of first, second and third person pronouns per 1,000 words.....	122
Figure 4.3:	Proportions for first, second and third person pronouns of total personal pronoun use.....	125
Figure 4.4:	Average word length in the five media.....	131
Figure 4.5:	Type/token ratio, with standard deviation	135
Figure 4.6:	Direct WH-questions.....	152
Figure 4.7:	Analytic negation.....	152
Figure 4.8:	Demonstrative pronouns	156
Figure 4.9:	Indefinite pronouns	156
Figure 4.10:	Present tense verbs.....	159
Figure 4.11:	Predicative adjectives.....	159
Figure 4.12:	Contractions.....	163
Figure 4.13:	Prepositional phrases.....	163

Figure 4.14:	Standard score distribution of the linguistic features that, in SCMC or SSCMC, deviate by more than 2 s.d. from Biber's (1988) mean.....	171
Figure 4.15:	Inserts	190
Figure 4.16	Emotives.....	190
Figure 4.17	Distribution of emotives in the conversational writing corpora.....	201
Figure 5.1a:	Mean scores on Dimension 1 for all genres.....	214
Figure 5.1b:	Spread of scores along Dimension 1 for all genres.....	220
Figure 5.2a:	Mean scores on Dimension 2 for all genres.....	223
Figure 5.2b:	Spread of scores along Dimension 2 for all genres.....	226
Figure 5.3a:	Mean scores on Dimension 3 for all genres.....	229
Figure 5.3b:	Spread of scores along Dimension 3 for all genres.....	234
Figure 5.4a:	Mean scores on Dimension 4 for all genres.....	236
Figure 5.4b:	Spread of scores along Dimension 4 for all genres.....	240
Figure 5.5a:	Mean scores on Dimension 5 for all genres.....	243
Figure 5.5b:	Spread of scores along Dimension 5 for all genres.....	247
Figure 5.6a:	Mean scores on Dimension 6 for all genres.....	249
Figure 5.6b:	Spread of scores along Dimension 6 for all genres.....	254
Figure 6.1:	Matrix combining the degree of shared context and the synchronicity of communication in the genres studied.....	283
Figure 6.2:	Relationships found between modalities, media and the genres investigated.	286

Abbreviations

ACMC	Asynchronous computer-mediated communication
AIM	America Online instant messenger
ASCII	American standard code for information interchange
BBS	Bulletin board system(s)
CMC	Computer-mediated communication
EFL	English as a foreign language
ELC	Electronic Language Corpus
FAQ	Frequently asked questions
ICE	International Corpus of English
ICQ	“I seek you” (chat)
IM	Instant messaging
IRC	Internet relay chat
LLC	London-Lund Corpus (of Spoken English)
LOB	Lancaster-Oslo/Bergen Corpus
LSWE	Longman Spoken and Written English Corpus
MD	Multidimensional
MF/MD	Multifeature/multidimensional
MMORPG	Massive multiplayer online role-playing game
MOO	MUD object-oriented
MSN	Microsoft network instant messenger
MUD	Multi-user dungeon
OED	<i>Oxford English dictionary</i>
SBC	Santa Barbara Corpus (of Spoken American English)
SCMC	Synchronous computer-mediated communication
SMS	Short message service (message)
SSCMC	Supersynchronous computer-mediated communication
TTR	Type/token ratio
UCOW	Uppsala Conversational Writing Corpus
1PP	First person pronoun(s)
2PP	Second person pronoun(s)
3PP	Third person pronoun(s)

Chapter 1. Introduction

1.1 Speech vs. writing vs. conversational writing

Every day millions of Internet users converse in real time by exchanging messages over computer chat systems. This study documents an investigation of conversational writing as carried out in text-based online chat in the early 21st century. More precisely, it presents a lexico-grammatical and functional linguistic analysis of features in a corpus of conversational writing consisting of synchronous and supersynchronous computer-mediated communication (CMC). Synchronous and supersynchronous computer chat differ in that the former is carried out one turn at a time, in for instance chat channels, whereas the latter is carried out in a split window, in which turns are realized keystroke by keystroke, so that completely overlapping turns are possible. The conversational writing corpus, compiled for the present study, is contrasted with existing corpora of various genres of speech and writing to elucidate the relationship between conversational writing and the spoken and written genres. The study is multidimensional in that it applies Biber's (1988) dimensions of linguistic variation to investigate the discourse. Biber's (1988) methodology lies at the heart of the study, as it enables the systematic assessment of lexico-grammatical patterns. In addition, certain textual, interpersonal and modal aspects of the communication are discussed in the light of e.g. Halliday's model of semiotics (see e.g. Halliday 1985a, 2004).

Previous studies have invariably pointed to the dual nature of computer chat, to its oral and written properties (e.g. Ko 1996, Mar 2000, Crystal 2001, Dresner 2005). The present study acknowledges this characterization, discussing the orality and writtenness of conversational writing, but also attempts to rise above the duality. The foremost aim of the study is to position two modes of computer-mediated communication (one synchronous and the other supersynchronous) on Biber's (1988) dimensions of linguistic variation, using Biber's multifeature multidimensional model. A mode is defined as "a genre of CMC that combines messaging protocols and the social and cultural practices that have evolved around their use" (Herring 2002: 112, drawing on Murray 1988). The synchronous mode investigated in the present study is Internet relay chat (IRC), and the supersynchronous mode is split-window ICQ ("I seek you") chat (for descriptions and screenshots of the modes, see section 2.6). Biber's (1988) dimensions are continua along which spoken and written genres vary with respect to more than their oral and written character, for instance with regard to their informational vs. involved focus. To adapt to Biber's framework, "modes" are also termed

“genres” in this study. Positioning the genres of conversational writing on Biber’s dimensions is expected not just to provide a clearer picture of conversational writing as a whole, but also to enable the detailed linguistic description of the discourse in the individual conversational writing genres.

From the inception of human computer-mediated communication, the synchronous and supersynchronous modes of computer chat, with their simultaneously oral and written properties, have puzzled linguists and laymen alike. While linguists carefully analyze the oral and written features of the discourse, the chatters themselves conceive of their communication as “talk.” Below are ten turns (examples 1 a–j) sampled from various computer chat channels and private chats in which chatters’ metalanguage reveals the perceived nature of the communication.

- (1) a. hey i’ll *talk* to ya all later i need to jet for a lil while
- b. ah been *talking* while ive been away have you ?
- c. i like it how you *talk* with me
- d. anyone want to *talk*
- e. i was trying to *talk* french..... ouoooo fuooooo pou shou
- f. where *talking* about hat in a chat
- g. i wasnt *talkin* to you!!
- h. jim what r u *talking* about?
- i. you’re not *saying* anything to me except hi
- j. so you wanna *hear* the rest of the v day story?

Internet relay chat and split-window ICQ chat (UCOW)

Computer chatters arguably perceive their communication as *talk*, as *saying* things and *hearing* each other’s utterances (cf. Giese 1998, Herring 2011b). Just as in an oral situation, their “talk” occurs in real time; it is spontaneous, interactive and immediately revisable. The oral nature of the communication is also reflected in the very denomination of the medium they use: computer chat. *The Oxford English dictionary* (OED) defines the verb “chat” as “to talk in a light and informal manner; to converse familiarly and pleasantly” and the noun “chat” as “familiar and easy talk or conversation.” The conversations in computer chat are fluid; topics evolve and evanesce; feedback is immediate; questions are answered (or not) and emotive content abounds (ranging from affective to adversarial). Even so, it is only through writing that the conversation is made possible; it is conveyed by keystrokes of letters and punctuation, and decoded visually by the recipient. The interlocutors depend on the encoding and decoding of graphemes, much like writers and readers in the written media (books, journals, magazines, hypertext, notes, etc.). Demonstrably, real-time text-based computer-mediated communication is a coin with two sides – the oral and the literate. The present study inhabits

this borderland of speech and writing, the field of tension that constitutes the interface between the spoken and the written, endeavoring to map it out.

Conversational writing is by no means a creation of the computer – it presumably appeared long before this invention. During classroom lessons, for instance, when silence is preferred, students sharing a desk may pass notes between themselves carrying out a silent, written conversation. Such a conversation relies on interlocutors' mutual awareness of each other's presence and immediate attention to the message. The students' messages thus constitute conversational writing (as does their act).¹ The computer, however, has made conversational writing (the textual product) amenable to large-scale study, or rather, the Internet and logging software have. Conversational writing texts in this study are chatted² texts produced for social interaction in synchronous and supersynchronous computer-mediated communication (SCMC and SSCMC).³ As mentioned, the two modes of computer-mediated conversational writing differ in that the conversational writing in SCMC is carried out in, for instance, chat channels to which participants submit their entire turn, one turn at a time, whereas the conversational writing in SSCMC is carried out between two or three interlocutors in a split window, in which turns are realized keystroke by keystroke with possible complete overlap. In either mode, the chatters producing the computer-mediated texts, like the students passing notes, rely on the simultaneous presence of a recipient and expect the recipient's immediate feedback. The synchronicity of their communication enables the interlocutors to affect each other's line of thought before or during its formulation into words, thereby, just as in oral interaction, enabling interlocutors to stake out the direction of the conversation. The following four characteristics in combination, then, provide a working definition of conversational writing: it is written communication 1) for social interaction 2) which requires the simultaneous presence

-
- 1 "Conversational writing" may refer to both a textual product (a noun) and an act (a verbal noun). The present study is primarily concerned with conversational writing in the former sense, i.e. with the texts themselves.
 - 2 "Chatted" is recurrently used as an adjective in this study, by analogy with the adjectives "spoken" and "written," to denote the texts of computer-mediated conversational writing (thus, "chatted" texts/corpora/words etc. are contrasted with "spoken" and "written" texts/corpora/words etc.).
 - 3 Other texts may also be produced in synchronous and supersynchronous CMC, for instance in office suites for collaborative writing (e.g. in Google Docs). Documents co-authored in the document window of such collaborative writing software, however, are typically expository prose, spreadsheets and presentations, and not conversational writing. Collaborative writing texts are not considered in this study. The present study is concerned only with conversational writing intended for social interaction.

(physical or virtual) of producer and recipient, 3) in which interlocutors expect immediate feedback (i.e. within seconds) and 4) during which the discourse may be reconfigured by the participants while under construction (e.g. as interlocutors are able to influence each other's line of thought).

Linguists studying computer-mediated discourse have characterized both asynchronous texts (such as e-mail and computer conferencing texts) and synchronous texts (computer chat) as intermediate between speech and writing. Investigating asynchronous CMC (ACMC), Collot & Belmore (1996: 28) call the communication a "hybrid" variety of English, Yates (1996: 46) concludes that it is "neither simply speech-like nor simply written-like," and Davis & Brewer (1997: 2) call it "writing talking." Studying synchronous CMC, Ferrara et al. (1991: 10) call the "interactive written discourse" a "hybrid register that resembles both speech and writing, yet is neither." Similarly, Foertsch (1995: 304) finds the electronic discourse to occupy "the middle ground between oral and written discourse," but also makes clear that the "most compositional formats" (cf. ACMC) fall closer to the written side whereas the "most interactive formats" (cf. SCMC) fall closer to the oral side. A number of empirical investigations of SCMC have shown that its discursive content is intrinsically oral in nature. Werry (1996) exemplifies richly from SCMC to show how the discursive style of the communication simulates face-to-face spoken language. Schulze (1999) points to the inherent interactivity as the most important characteristic of SCMC, presenting non-verbal and paraverbal properties as well as means for signaling presence cues and status information as features that make SCMC similar to spoken communication. Hård af Segerstad (2002: 246) characterizes SCMC as "a form of conversation, which happens to be written down instead of spoken."

Although most previous studies show that computer-mediated discourse defies simple classification into speech or writing, they point to the important assumption that synchronously mediated texts display more speech-like properties than asynchronous texts (cf. Korsgaard Sorensen 1993, Herring 2001, Sveningsson 2001, Hård af Segerstad 2002, Condon & Čech 2010, Georgakopoulou 2011a). Very few linguists have studied supersynchronously mediated conversational texts, even though several have suggested such studies, e.g. Hård af Segerstad (2002: 269) and Freiermuth (2003: 183). Herring (2004a, 2007) suggests synchronicity as a useful parameter for distinguishing among modes of CMC, seeing that synchronicity is a "robust predictor of structural complexity, as well as many pragmatic and interactional behaviors, in computer-mediated discourse" (Herring 2007: 14). Given its greater interactivity, supersynchronous conversational CMC might thus display even more speech-like properties than synchronous conversational CMC,

a possibility that makes a contrastive study of the two highly desirable. However, as Herring (2011b) points out, genres of conversation (oral as well as computer-mediated) should be studied not only with regard to the oral vs. written dimension, but could be situated along various other dimensions (cf. Biber's 1988 study). The present investigation, consequently, takes all of Biber's six dimensions into account (e.g. informational vs. involved production and narrative vs. non-narrative concerns, dimensions further described in section 2.3) for the classification of synchronous and supersynchronous conversational CMC.

Herring (2011b) notes that the scholarly assessment of relative degrees of conversationality in different CMC modes is straggling; "no single set of methods is employed, or questions asked, across the collection that would make the results of the individual studies directly comparable with one another" (2011b: 7). Calling for research in the field, she emphasizes that the "systematic consideration of what it means for CMC to be 'conversational' is still lacking" (2011b: 3), as is the systematic comparison of multiple modes of CMC using a "common set of methods" (2011b: 7). The present study is a first step towards remedying these shortcomings; it intends not only to describe conversational writing, but intends to do so using Biber's (1988) systematic multifeature multidimensional (MF/MD) methodology. Positioning the two modes of CMC on Biber's dimensions enables not just the systematic comparison of the modes, but also the systematic comparison of the modes (genres) relative to other genres of writing and speech. Although none of Biber's (1988) dimensions makes a simple, dichotomous distinction between writing and speech, the dimensions differentiate among literate and oral genres in different respects. Among the genres situated by Biber (1988) on the dimensions are face-to-face and telephone conversations. By relating the conversational writing genres to these conversational genres on the different dimensions, lexico-grammatically, situationally and functionally, it is possible to determine the degree of orality in conversational writing. A high degree of orality means that the conversational writing genre displays features with great resemblance to spoken conversations, or even displays features or levels beyond current notions of orality, thus re-defining what it means to converse in real time. In the next section, the hypotheses regarding the relationship between the conversational writing genres and oral conversations are presented, along with the research questions to be addressed in the study.

1.2 Aim and scope of the study

The principal aim of the present study is to position two genres of conversational writing, one of synchronous and the other of supersynchronous CMC, on Biber's

(1988) six dimensions of textual variation; Biber's methodology and dimensions are described in section 2.3, and the positions of conversational writing genres are presented and discussed in chapter 5. The dimensions distinguish spoken and written genres but are not strictly scales of variation between speech and writing; rather, they are "fundamental parameters of linguistic variation among English texts" (Biber 1988: 200).⁴ Biber's (1988) *multi*-dimensional model, by definition, substantiates that there is no single dimension of orality vs. literacy (writtenness); rather, texts vary on several dimensions at one and the same time. Three of the dimensions (1, 3 and 5) in themselves can be said to distinguish between spoken and written discourse (despite evincing some overlapping genres of speech and writing), but the other three dimensions (2, 4 and 6) do not correspond to this distinction. Determining the degree to which conversational writing resembles oral conversation therefore imperatively entails consideration of the genres' positions on all six dimensions. It is simply not adequate to equate the two on the basis of one dimension only; rather, in the multidimensional model, "two genres are 'similar' to the extent that they are similarly characterized with respect to all dimensions; they are 'different' to the extent that they are distinguished along all dimensions" (Biber 1988: 168).

It was mentioned in the previous section, and above, that the degree of orality in conversational writing can be determined by considering how conversational writing relates to oral conversations on Biber's (1988) dimensions. Systematic correspondence between the conversational writing genres (from SCMC and SSCMC) and oral conversations (face-to-face and telephone conversations) on all dimensions should then suggest a high degree of orality in conversational writing (although such a correspondence has to be functionally attested to be conclusively established). Biber (1988) mentions face-to-face conversations as a stereotypically oral genre, as "having the characteristic situational features that are most typical of speech" (1988: 162). His analyses characterize discourse as highly oral when it displays characteristics of involved production, situation-dependent reference and non-abstract content, as opposed to highly "literate" discourse, which displays features of informational production, explicit, elaborated reference and abstract content (1988: 162–163), all characterizations based on dimensions 1, 3 and 5. Special attention is thus paid to these dimensions in the investigation of the orality of conversational writing, even though, admittedly, the full picture of the nature of conversational writing emerges only through the

4 In variation studies, studying speech means studying transcribed speech, hence spoken "texts."

overall consideration of all dimensions together, and such overall consideration is also forthcoming, in the penultimate chapter of this study.

The assumption underlying the two hypotheses to be tested in this study is that most written discourse is conveyed in one-way communication or in asynchronous exchanges (with delay between production and reception), whereas most oral discourse is conveyed in synchronous exchanges (with no delay between production and reception), and that the degree of orality increases with the degree of synchronicity (cf. Korsgaard Sorensen 1993, Condon & Čech 2010). As mentioned, Herring (2007: 14) suggests synchronicity as “a useful dimension for comparing different types of CMC with spoken and written discourse.” The present study consequently acknowledges the importance of the synchronicity of communication in the various genres under study, alongside the analysis of the genres’ positions on Biber’s dimensions. By virtue of being communicated in real time, the discourse of the conversational writing genres is expected to approximate the discourse of oral conversations, despite its not being spoken. The two hypotheses underlying the present investigation are presented below. The first is derived from previous research on CMC (e.g. Foerstch 1995, Sveningsson 2001, Hård af Segerstad 2002, Herring 2007), and it forms the point of departure for the second hypothesis.

- Synchronous conversational writing displays a higher degree of orality than asynchronous CMC
- Supersynchronous conversational writing displays a higher degree of orality than synchronous conversational writing

None of Biber’s dimensions explicitly distinguishes between asynchronous and synchronous discourse; instead, as described, “orality” in the present study is determined by a genre’s similarity to oral conversations. A factor concomitant with such similarity, however, is synchronicity; oral conversations are indeed synchronous. Conversational writing, as mentioned, comprises synchronous and supersynchronous communication. The supersynchronous mode surpasses oral conversations in that interlocutors in SSCMC can carry out conversations in complete overlap for an extended period of time (as supersynchronous conversational writing is carried out in a split window into which both interlocutors type at once), even if this opportunity is not taken at all times. Such complete overlap is possible in oral conversations too, but is usually avoided as extended overlap renders the communication incomprehensible (cf. Herring 1999). In SSCMC, by contrast, extended complete overlap does not affect the comprehensibility of the communication. Supersynchronous conversational writing can thus be regarded as exceeding oral conversations in synchronicity (which explains its denotation as *supersynchronous*); see also table 1.1 below (to be explained in section 1.3).

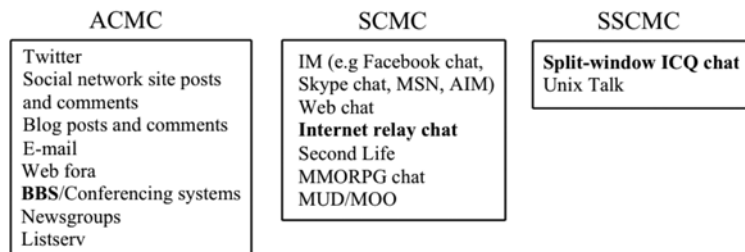
Determining the degree of orality in SSCMC consequently entails taking into account not just the similarity of supersynchronous conversational writing to oral conversations, but also the paradoxical possibility of the former exceeding oral conversations in “orality,” as seen from the perspective of synchronicity. This will be borne in mind in the interpretation of the results of the study, especially in the consideration of the positions of the supersynchronous conversational writing genre on Biber’s dimensions, more precisely, on dimensions 1, 3 and 5, the dimensions with an “oral” end.

The corpus of conversational writing to be investigated in the present study, the “Uppsala Conversational Writing Corpus” (UCOW), was recorded and annotated in a research project culminating in this book, and will be described in detail in sections 3.1–3.3. The corpus consists of conversational writing from SCMC, as instantiated in Internet relay chat (IRC) chat channels, and from SSCMC, as instantiated in private split-window ICQ chats. The two genres to be positioned on Biber’s dimensions are therefore labeled “Internet relay chat” and “split-window ICQ chat,” respectively. The genres are exemplars of SCMC and SSCMC, much as face-to-face conversations and a range of other genres exemplify speech, and as e.g. academic prose and a host of other genres exemplify writing. The categories speech, writing, ACMC, SCMC and SSCMC each have the working label of “medium” in the present study. In a slightly opportunistic account of Internet language, Crystal (2001) subsumes the various modes of textual CMC under one linguistic variety labeled “Netspeak.” “Netspeak,” according to Crystal, “is something completely new [...] something fundamentally different from both writing and speech [...] in short, a fourth medium,” the first three being speech, writing and sign language (Crystal 2001: 238). The present study recognizes the relative novelty of textual CMC, but stresses the heterogeneity of this communication, above all hesitating to draw conclusions as to the linguistic nature of all CMC. Rather, conclusions are drawn only with respect to conversational writing, as instantiated through IRC and split-window ICQ chat, or with respect to one of these modes.

To understand the diversity of CMC modes for social interaction, see figure 1.1. Out of all these modes (further explained in sections 2.5 and 2.6), the present study covers only one mode of SCMC (Internet relay chat) and one mode of SSCMC (split-window ICQ chat). With regard to ACMC, the study comments on various previous research, especially that of Collot (1991) and Collot & Belmore (1996) on bulletin board system (BBS) communication, as well as Yates (1993, 1996). ACMC is brought in mostly as a quantitative point of reference in this study, as the ACMC data to be compared to conversational writing consists of linguistic frequency counts, derived from previous research, more than actual corpus texts (since few texts have been made available from the comparable studies).

The hands-on empirical investigations in this study thus focus primarily on conversational writing, that is, on the conversational discourse in SCMC and SSCMC.

Figure 1.1: Examples of asynchronous, synchronous and supersynchronous modes of written CMC.⁵



The research questions to be answered in the course of the present study are the following four (parentheses indicating chapters in which the questions are addressed):

-
- 5 Some modes deserve explanation as they may be unfamiliar to present-day readers. A BBS is a bulletin board system run on a server to which users log in to post and exchange messages (Collot 1991, Collot & Belmore 1996). BBSs peaked around 1996 but were rapidly replaced by web fora upon the popularization of web browsers (the hypertext protocol). Computer conferencing systems have come in an abundance of modes (besides BBSs), e.g. CoSy, VAX Notes, Confer, First Class; those investigated by linguists include CoSy (Yates 1993, 1996) and VAX Notes (Davis & Brewer 1997). Newsgroups are hosted by Usenet servers and accessed in client programs via users' selective subscription (Herring 2002, Paolillo 2011). Listservs are electronic mailing list software applications that allow users access to global interest groups, also by subscription (Herring 1996b). Like BBSs, newsgroups and listservs have largely been superseded by web fora and other web-based applications. IM is an umbrella term for instant messaging software, of which early applications include Microsoft Messenger (MSN) and America Online instant messenger (AIM). Linguistic studies of IM include Baron (2004, 2010) and Tagliamonte & Denis (2008). Second Life is a graphic online virtual world in which users interact as avatars. Also graphic, MMORPGs are massive multiplayer online role-playing games, e.g. World of Warcraft. MMORPGs have largely superseded MUDs, multi-user dungeon games, and MOOs, i.e. MUD object-oriented applications, which are text-only virtual worlds (Reid 1994, Cherny 1994, 1999, Herring et al. 2009). Second Life, MMORPGs, MUDs and MOOs all allow synchronous chat among participants in the virtual worlds. A predecessor of split-window ICQ chat is Unix Talk, also carried out in a split window and realized character by character with possible complete overlap.

- What is the linguistic nature of conversational writing and the genres studied here, IRC and split-window ICQ chat? (Chapters 4, 5 and 6)
- How does conversational writing carried out in SCMC and SSCMC, respectively, relate to writing and speech? (Chapters 4, 5 and 6)
- How do the genres of SCMC, SSCMC and ACMC relate to oral conversations on Biber's (1988) dimensions? (Chapters 5 and 6)
- Does conversational writing carried out in SCMC and SSCMC constitute a modality of its own? (Chapter 6)

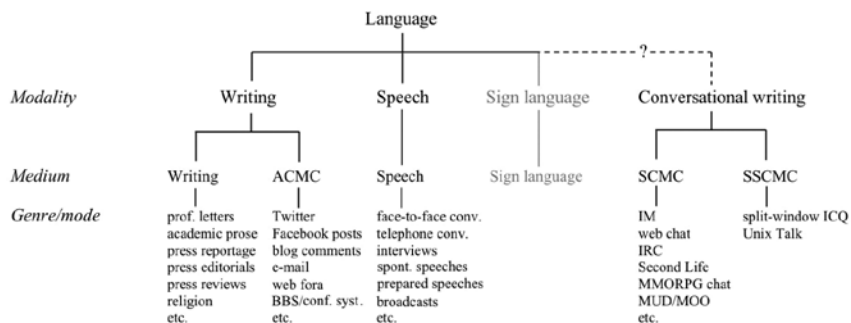
The first two research questions are treated extensively in chapter 4, which contrasts the conversational writing genres with the media of speech, writing and ACMC (the latter represented by Collot's 1991 BBS conferencing genre), in chapter 5 with regard to Biber's dimensions and in chapter 6, summarizing and discussing the results. The third question is partly addressed in chapter 5, in which the positions of Internet relay chat (SCMC) and split-window ICQ chat (SSCMC) on Biber's dimensions are presented, as well as those of BBS conferencing (ACMC), and partly in chapter 6, which discusses the CMC genres' similarity to oral conversations on Biber's dimensions. By relating the CMC genres to oral conversations on Biber's dimensions it is also possible to address the two hypotheses posed at the beginning of this section, which suggest different degrees of orality in texts from the three CMC media.

As mentioned, the working label for speech, writing, ACMC, SCMC and SSCMC is "media." Speech and writing are known from previous research to be separate modalities, as is sign language; see figure 1.2.⁶ The present study leaves sign language out of account, but attempts to answer the fourth research question, as to whether conversational writing constitutes a fourth modality, as it is conceptualized along the dashed line in figure 1.2. This fourth research question will not be addressed until chapter 6, when all results have been presented, as the answer must be backed up by substantial evidence. In the meantime, the genres are subsumed under their media categories when compared to speech and writing. This means that, in chapter 4, the media of ACMC, SCMC and SSCMC are compared to the media of speech and writing whereas, in chapter 5, the genres

6 A modality is a "means of production/reception" (Herring 2007: 5), i.e. a means of materializing a linguistic message. Three modalities are regularly recognized in linguistics: speech, writing and sign language (cf. Baron 1981). Crystal (2001) calls each of the three a medium. Crystal (2008b: 300) notes that speech is regarded as "the 'primary medium'" and "writing the 'secondary' or 'derived' medium," and that various branches of linguistics may denominate these modalities instead of media.

of IRC (representing SCMC) and split-window ICQ chat (representing SSCMC) are compared to various genres of speech and writing. ACMC is subsumed under the written modality throughout this study, as ACMC is not defined as conversational writing.⁷ Owing to the lack of corpus texts (from Collot’s 1991 ACMC study), however, the investigation of ACMC is limited; instead, discussions mainly revolve around conversational writing as compared to traditional writing and speech (cf. the genres studied in Biber 1988). Figure 1.2 illustrates the working relationship between modalities, media and genres/modes in the present study, in which genres are subcategories of the media writing and speech, and modes are subcategories of the three CMC media.

Figure 1.2: Working relationship between modalities, media and genres/modes in the present study.



The primary purpose of the first results chapter, chapter 4, is to document the features that are salient in conversational writing when the genres of SCMC and SSCMC studied are compared to speech and writing at the level of medium. The quantitative findings presented in the chapter utilize the mean frequencies and standard deviations of Biber’s (1988) linguistic features, and interpretations draw on the fact that the features in conversational writing that deviate most from the mean of all spoken and written genres (considered in Biber 1988) are those that

7 Drawing on Morrisett (1996), Mann & Stewart (2000: 182) note that ACMC “has been associated with [...] characteristics found in traditional writing forms,” including “having the time to study, analyse and reflect on incoming messages and being able to compose responses carefully” (ibid.), but also acknowledge that other analyses have characterized ACMC as a hybrid variety between writing and speech (as noted in section 1.1 above).

most distinctively characterize conversational writing. In addition, the chapter takes up salient features in the conversational writing corpus that are not included among the features studied in Biber's (1988) multidimensional methodology, among them paralinguistic features. The analysis of conversational writing thus sets out broadly, relating computer-mediated communication to speech and writing, and proceeds to more fine-grained scrutiny, comparing the genres of conversational writing to the multiple genres of speech and writing, all in order to adequately answer the research questions posed above.

1.3 Synchronicity of communication

The present study recognizes synchronicity as a useful construct for classifying text-based computer-mediated communication (as suggested in Herring 2004a, 2007). Accordingly, figure 1.1 illustrated the synchronicity of communication in a number of CMC modes. As mentioned, the analysis of conversational writing in this study is based on the UCOW components Internet relay chat, representing SCMC, and split-window ICQ chat, representing SSCMC. The UCOW findings are related to the findings in previous research on APMC, especially on BBS conferencing (Collot 1991, Collot & Belmore 1996), but more importantly, to the genres studied by Biber in his account of textual variation in English (Biber 1988). Concomitant to the classification of synchronicity in the CMC genres/modes (figure 1.1), therefore, is the consideration of the synchronicity of communication in Biber's (1988) genres. Biber studied six genres of speech (face-to-face conversations, telephone conversations, interviews, broadcasts, spontaneous speeches and prepared speeches) and 17 genres of writing (including professional letters, academic prose, press reportage, press editorials, popular lore, general fiction and official documents). The spoken texts derived from the London-Lund Corpus, LLC (Svartvik 1990), and the written texts from the Lancaster-Oslo/Bergen Corpus, LOB (Johansson et al. 1978), and two collections of letters. A complete list of the texts included in Biber's (1988) study is given in Appendix I. To begin to relate the CMC genres/modes to Biber's genres with regard to synchronicity of communication, they are here conflated into one list, table 1.1, along with a number of other existing and hypothetical genres (not yet classified as such).

Table 1.1: *Principal synchronicity and direction of communication in various genres. Genres studied appear in bold script, other genres in normal script*⁸

Type of communication	corpus	asynchronous		synchronous		supersyn-
		one-way	two-way	one-way	two-way	chronous two-way
Speech (conversation)						
face-to-face	LLC, SBC	A				● (●)
audiovisual telephone		A				● (●)
audiovisual Skype telephone		A				● (●)
telephone	LLC	A				● (●)
Skype telephone		A				● (●)
Ventrilo		A				● (●)
Conversational writing						
split-window ICQ chat	UCOW	V				● ●
Internet relay chat	UCOW	V			(●)	● ●
Second Life		V			(●)	● ●
MMORPG chat		V			(●)	● ●
notes passed face-to-face		V			(●)	● ●
IM, e.g. Facebook chat		V	(●)	(●)	(●)	● ●
Speech – continued						
interv., publ. conv., debates	LLC	A				●
spontaneous speeches	LLC	A				(●)
prepared speeches	LLC	A				●
TV/web broadcasts ⁸		A	●			●
audio broadcasts	LLC	A	●			●
voicemail, Heytell		A	●	(●)		
Asynchronous CMC						
SMS		V	(●)	●	●	(●)
Twitter		V	●	●	(●)	(●)
e-mail		V	●	●	(●)	(●)
Facebook posts, comments		V	●	●		
blog posts and comments		V	●	●		
newsgroups, BBS, web fora	ELC other	V	●	●		
Writing						
personal letters	Grabe	V	(●)	●		
professional letters	Biber	V	●	(●)		
FAQs		V	●	(●)		
Wikis		V	●	(●)		
posted personal notes		V	●	(●)		
academic prose	LOB	V	●	(●)		

8 TV/web and audio broadcasts, of course, may contain e.g. conversations – the categorization here indicates only the communicative purpose of the broadcast to the public and the conventional response from the public.

Type of communication	corpus	asynchronous		synchronous		supersyn-
		one-way	two-way	one-way	two-way	chronous two-way
commercial web pages	V	●				
direct mail letters	V	●				
press reportage LOB	V	●				
press editorials LOB	V	●				
press reviews LOB	V	●				
religion LOB	V	●				
hobbies LOB	V	●				
popular lore LOB	V	●				
humor LOB	V	●				
biographies LOB	V	●				
general fiction LOB	V	●				
mystery fiction LOB	V	●				
science fiction LOB	V	●				
adventure fiction LOB	V	●				
romantic fiction LOB	V	●				
official documents LOB	V	●				

A – mainly auditory reception

V – mainly visual reception

Table 1.1 outlines the distribution of genres according to their principal synchronicity and direction of communication.⁹ Communication may be asynchronous, synchronous or supersynchronous. Asynchronous and synchronous communication can vary in direction, i.e. it can be one-way or two-way. Professional letters that elicit no response from the recipient, for instance, are communicated one way, presumably the default direction of such letters (indicated by a bullet in table 1.1). A professional letter that is responded to, however, becomes part of a two-way asynchronous transaction (indicated by a parenthesized bullet). Supersynchronous communication, however, is by default two-way.¹⁰ In individual genres, texts may be communicated with different synchronicity and direction.

9 Table 1.1 includes SMS among asynchronous CMC. Although similar to APMC modes (cf. figure 1.1), SMS is, strictly speaking, not CMC, but rather telecommunication, and therefore not included among the APMC modes in figure 1.1.

10 The delineation of two-way communication here differs from Herring's (2001) definition of two-way transmission, as two-way communication here includes asynchronous writing, whereas Herring (2001) regards asynchronous writing as one-way transmission. Herring's definition of two-way transmission includes only modes (genres) in which speaker and addressee perceive the message as it is produced, such as oral conversations and supersynchronous conversational writing.

The bullets in table 1.1 indicate the main types of communication carried out in the genres, i.e. the types of communication used to fulfill the communicative purposes of the genre. By inference, face-to-face conversations are usually synchronous (as indicated by a bullet), but may be supersynchronous for very limited periods of time (as indicated by a parenthesized bullet). Split-window ICQ chat, by contrast, is both synchronous *and* supersynchronous, as its keystroke-by-keystroke means of transmission enables the communication to fluctuate between being realized in consecutive, synchronous, turns up to extensively overlapping, supersynchronous, turns.

Internet relay chat, on the other hand, is only carried out synchronously, i.e. turn by turn, and one-way messages, i.e. turns not responded to, are more likely in the Internet relay chat channels than in the split-window ICQ mode (adding a parenthesized bullet for the former, but not for the latter, in table 1.1). The genres to be contrasted in the present study are marked in bold in table 1.1, and the corpora from which they derive are indicated in the second column. (Other existing and hypothetical genres are interspersed among these, *inter alia* to illustrate a number of linguistically understudied genres.)

The genres in table 1.1 are ordered from top to bottom by their principal degree of synchronicity. Genres of similar synchronicity are only tentatively ordered relative to each other (the written genres in one-way communication, of course, defy ranking altogether). The table nevertheless serves to illustrate an important point. Before conversational writing, synchronous communication relied almost exclusively on the acoustic channel, i.e. on auditory reception (indicated as A in table 1.1). Auditory (A) and visual (V) reception were then largely on a par with the distinction between speech and writing. Table 1.1 illustrates how conversational writing challenges this division; not only is conversational writing synchronous, like spoken conversation, but it also challenges conversation with a more synchronous genre, one amenable to extended supersynchronous communication: split-window ICQ chat. As will be seen in this study, Biber's (1988) multidimensional methodology is a highly useful tool for distinguishing speech and writing on dimensions other than oral (cf. A) vs. literate (cf. V). Yet, split-window ICQ chat was not around at the time of the 1988 methodology's conception, nor was written synchronous communication included in Biber's study, cf. Internet relay chat. Taking synchronicity, especially supersynchronicity, into account is therefore imperative for an adequate description of the orality of conversational writing, in the interpretation of the positions of conversational writing on Biber's dimensions (chapters 5 and 6), as well as in the preceding comparison of conversational writing to speech and writing (chapter 4). Table 1.1 serves to conceptualize the parameters to bring into those considerations.

1.4 Notes on terminology

A few notes are in order with regard to the terminology applied in the present study. Table 1.1 in the previous section lists in bold the “genres” to be considered in the study. The term “genre” is used here in analogy with Biber (1988) referring to “categorizations assigned on the basis of external criteria” (1988: 70), that is, criteria related to the author’s or speaker’s communicative purpose. The genres in Biber (1988) are largely adopted from the categories distinguished in the corpora from which they derive (see Appendix I) and constitute “text categories readily distinguished by mature speakers of a language” (Biber 1989: 5). The present study agrees with Biber’s (1988, 1989) definition of genre, finding the CMC modes studied, especially IRC and split-window ICQ chat, equally distinguishable on external criteria as Biber’s genres, i.e. on the basis of their external format and the distinct situational setting of their production, and thus worthy of the designation of genre (although they may also be referred to as modes).

In addition to genre, however, the term “register” is also regularly used in linguistic studies to refer to situationally defined varieties of speech and writing. While some studies exclusively use the term “genre” (e.g. Biber 1988, Biber & Finegan 1989, Swales 1990, Love 2002), others use the term “register” (e.g. Atkinson & Biber 1994, Biber 1995, Biber et al. 1999, Conrad 2001). Some have attempted to draw theoretical distinctions between genres and registers, e.g. Ferguson (1994) who regards “genre” as “[a] message type that recurs regularly in a community” and “register” as “[a] communication situation that recurs regularly in a society” (1994: 20–21), while others have used the terms rather interchangeably, e.g. Biber (1993: 244) who uses both to refer to “situationally defined text categories.” In 1995, Biber notes that there is “no general consensus within sociolinguistics concerning the use of *register* and related terms such as *genre* and *style*” (1995: 8, original italics), reviewing attempted distinctions as “quite abstract and vague” (1995: 9). In his 1995 study, Biber opts for “register” as a general cover term for all aspects of variation in use, admitting that it corresponds closely to his earlier use of “genre” (1995: 10).

Discussing genre categorizations in corpora, Lee (2001) points out that several genres in corpora really denote sub-genres (e.g. the five fiction genres in LOB), rather than situationally defined varieties, but calls for calm among linguists; “we need not be unduly worried about whether we are working with genres, sub-genres, domains, and so forth, as long as we roughly know what categories we are working with and find them useful” (2001: 52). Advocating consistency in any approach, Lee proposes the usefulness of seeing the terms “genre” and “register” as two different angles, or points of view; “genre” being used to talk about

“membership of culturally-recognisable categories” (2001: 46) and “register” being used to talk about “lexico-grammatical and discoursal-semantic patterns associated with situations” (ibid.). Genres, says Lee, are instantiations of registers (as a genre may invoke more than one register), and “so will have the lexico-grammatical and discoursal-semantic configurations of their constitutive registers, in addition to specific generic socio-cultural expectations built in” (2001: 46–47). As if in line with Lee’s reasoning, Biber & Conrad (2009), although they draw distinctions between “register,” “genre” and “style,” opt to focus mostly on the register perspective, as it is seen as valid for the description of all text varieties. Whereas the genre perspective focuses on the “conventional structures used to construct a complete text within the variety” (2009: 2), say Biber & Conrad, the register perspective can be used to analyze “any text sample of any type” (ibid.).

The present work studies conversational writing from both the genre and the register perspective (leaving individual “styles” largely out of account) but employs the term “genre” from both perspectives in the multidimensional analysis. The conversational writing genres have been identified on the basis of external criteria, and the study aims to identify the genres’ lexico-grammatical patterns, i.e. the “conventional structures” that characterize them as varieties. This is done by contrasting the texts of conversational writing with texts from spoken and written genres, for which the conventional structures have been pre-defined (as sets of co-occurring linguistic features) by Biber (1988). Moreover, as the conversational writing genres here in effect represent only one register each, the genre/register distinction is not of central concern. Accordingly, instead of using “register,” like Biber (1995), in a way that is similar to Biber’s earlier use of “genre,” the present author simply opts for using the original term “genre,” as defined in Biber’s early work (1988, 1989, Biber & Finegan 1989), in the first place.

In Halliday’s model of semiotics (e.g. Halliday 1978, Halliday & Hasan 1989) and systemic-functional linguistics (e.g. Halliday & Hasan 1989, Martin 1992, Halliday 2004), on the other hand, the term “register” is the mainstay construct, whereas “genre” is peripheral. A register is a functional variety of language (Halliday 1978, 2004). It is defined on the basis of three variables of context taken together: field, tenor and mode,¹¹ which essentially represent what is going on in the course of the language exchange, who is taking part, and what role the

11 Halliday’s variable of mode is paradigmatically distinct from mode defined as a genre of CMC. The semiotic notions field, tenor and mode will be explained further in section 2.4. Mode in the Hallidayan sense will be referred to as “semiotic mode,” whenever discussed in non-Hallidayan contexts, to set it apart from “mode” used to denote genres of CMC.

language is playing, respectively. (“Register” in the Hallidayan sense will be further explained in section 2.4.) In the systemic-functional framework, “genre” and “register” are said to represent different semiotic planes. “[A] genre is a staged, goal-oriented, purposeful activity in which speakers act as members of [a] culture” (Martin 2001b: 155), that is, in order to fulfill certain communicative purposes. “Register,” by contrast, has more to do with the particular linguistic choices communicators make in a certain genre. The choices result from the contextual variables field, tenor and mode. Genre thus corresponds roughly to “context of culture” and register to “context of situation” (Martin 2001b: 155). In the present study, the systemic-functional concept of “genre” will not be employed. Rather, in the analysis of conversational writing, only the Hallidayan concept of “register” will be used, to discuss linguistic features associated with the field, tenor and mode of the discourse, i.e. the distinct situational setting of the discourse. The configuration of the field, tenor and mode is realized in any given text:

Any piece of text, long or short, spoken or written, will carry with it indications of its context. We only have to hear or read a section of it to know where it comes from. This means that we reconstruct from the text certain aspects of the situation, certain features of the field, the tenor, and the mode. Given the text, we construct the situation from it. (Halliday & Hasan 1989: 38)

In sum, in the present study, whenever “genre” is used, it is with regard to Biber’s framework, and to the extent that “register” is used, it is employed in one of two ways: 1) as interchangeable with Biber’s notion of genre, since, as mentioned, several authors have used “register” and “genre” interchangeably (as will be seen, for instance, in section 2.2 surveying previous research into speech and writing) or 2) in connection with semiotic, systemic-functional interpretations. The context of each discussion will clarify in which meaning the term “register” is employed.

A fair number of abbreviations will be used throughout the study. Several of these were encountered in the sections above and, as seen, they are usually explained upon first encounter – if not, and for repeated reference, readers may consult the list of Abbreviations (front matter). Corpus citations throughout the study contain full names of genres and abbreviations for corpora. The corpus citation convention applied for the conversational writing genres is “corpus genre + text number + (corpus name),” e.g. “Internet relay chat text 3b (UCOW)” and “Split-window ICQ chat text 11 (UCOW),” except for when several short samples are conflated into one example, in which case only the genre(s) and corpus name are given, e.g. as in example (1) above, citing “Internet relay chat and split-window ICQ chat (UCOW).” As indicated among conversations in table 1.1, face-to-face conversations from the Santa Barbara Corpus of Spoken American English

("SBC," for short), from the 1990s (Du Bois et al. 2000), are studied alongside other genres in the investigation here. A subset of the corpus was adapted and annotated for the present study to supplement the British LLC spoken genres (studied in Biber 1988), for updated reference and a more global approach (see section 3.4 for a description of the procedure). To distinguish among face-to-face conversations from the two corpora, they are named "face-to-face conversations LLC" and "face-to-face conversations SBC" in the discussions. The two are treated as separate genres, although, admittedly, they constitute regional more than situational varieties. Their denomination as individual genres is merely applied for convenience and for consistency with the use of the genre perspective in Biber's (1988) methodology.

1.5 Outline of the study

Following this introductory chapter, the present study is organized into chapters of background (chapter 2), material and method (chapter 3), results (chapters 4 and 5), discussion (chapter 6) and conclusion (chapter 7).

The background, chapter 2, starts out by surveying previous literature on speech and writing, to introduce some of the linguistic features that distinguish between texts from the two modalities. Several of the studies mentioned in the survey are important because they are reflected in Biber's (1988) selection of features, those that he used to map out spoken and written genres. The survey also serves as a theoretical backdrop to the discussions in the results chapters. Next, the chapter introduces Biber's and Halliday's frameworks in separate sections. Biber's approach to linguistic variation is quantitative at its outset, but enables paramount qualitative, functional interpretation, whereas Halliday's approach to linguistic variation is essentially qualitative. The choice of Biber's and Halliday's approaches is partly drawn from Yates' (1993) study of ACMC, as some passages of the present study attempt to parallel Yates' study with analogous analyses of SCMC and SSCMC. Biber's (1988) MF/MD methodology is broadly outlined in chapter 2, as is Halliday's theory of metafunctions in language. The chapter then surveys the literature on computer-mediated communication, among other things to present how previous studies have treated conversational writing. Chapter 2 ends with a description of the interfaces for conversational writing, so as to anticipate the UCOW corpus description in chapter 3.

Chapter 3 is the "Material and method" chapter. It describes the compilation and annotation of the UCOW corpus, the sampling and annotation of SBC, and the application of Biber's (1988) MF/MD methodology to the material. The chapter explains the data retrieval procedure and the calculation of the results.

The quantitative investigation of conversational writing takes all of Biber's 67 linguistic features into account, and the results are presented in two ways: in relation to writing, APMC and speech, in chapter 4, and in relation to all of Biber's spoken and written genres, in chapter 5.

Chapters 4 and 5 are the results chapters. Chapter 4 focuses on the salient features in conversational writing, e.g. those taken up in previous studies of CMC, such as modal auxiliaries and paralinguistic features, but also features rarely accounted for in quantitative studies of conversational writing, such as inserts (an umbrella term for e.g. interjections and discourse markers, typically found in conversations) and "emotives." "Emotives" is an umbrella term invented in the present study for emoticons (e.g. :, ;) and sentiment initialisms (e.g. *lol*, meaning "laughing out loud"), both of which add an emotional zest to chatters' utterances. The thrust of chapter 4, however, is to present qualitative analyses of salient quantitative results from the feature counts in the application of Biber's methodology and to contrast measures of lexical diversity (such as type/token ratio, TTR, and lexical density) in the annotated corpora. The most salient linguistic features in conversational writing are those that deviate from the mean of Biber's spoken and written genres by more than two standard deviations, and these, together with other features presented in chapter 4, epitomize the character of conversational writing. Chapter 4 thus constitutes a major step in the description of conversational writing.

Another major step towards the description of conversational writing in relation to speech and writing is taken in chapter 5, which presents the positions of the conversational writing genres (as well as SBC) on Biber's dimensions of linguistic variation. Like chapter 4, the chapter discusses numerous examples from the corpora, to elucidate the nature of conversational writing, but whereas chapter 4 adduces abundant theoretical anchorage to previous research, chapter 5 essentially breaks new ground as regards conversational writing, with fewer references to previous research. Both results chapters, however, contain analyses and discussions, not just the results. Much of the character of conversational writing thus emerges already in the results chapters, even though the penultimate chapter, chapter 6, is dedicated to a crucial, summarizing discussion of all results.

Chapter 6 revisits the hypotheses and research questions posed in section 1.2. The chapter narrows down what answers to these were provided in the study, discusses the findings, and points out what it means for chatted texts to be conversational. Chapter 7, finally, provides a concluding summary of the study and some suggestions for further research.

Chapter 2. Background

2.1 Introductory remarks

Embarking upon a linguistic study of synchronous and supersynchronous computer-mediated communication, which problematizes the concepts of speech and writing, one might first, as a background, address a number of questions pertaining to the media. Firstly, what differences between speech and writing have been found in previous linguistic studies? Secondly, how can genres/registers of speech and writing be described quantitatively and qualitatively? Thirdly, how has CMC been approached linguistically before, and fourthly, how is synchronous and supersynchronous conversational writing carried out? The present chapter attempts to answer these questions by, respectively, surveying previous literature on speech and writing (in section 2.2), elaborating on quantitative and qualitative approaches in Biber's and Halliday's frameworks (in sections 2.3 and 2.4), surveying the linguistic literature on CMC (in section 2.5) and describing the media for conversational writing (in section 2.6). The last section (2.7) then sums up the chapter.

2.2 Survey of the literature on speech and writing

Since the turn of the twentieth century, the nature of the relationship between spoken and written language has attracted considerable interest among linguists. Woolbert (1922) was one of the first to bring scholarly attention to the similarities and differences between speech and writing. His 1922 article begins:

Speaking and writing are alike – and different. Just how like and how different has never been adequately stated. (Woolbert 1922: 271)

Woolbert's study presented only a number of very limited general observations (of the type "the voice of the speaker can always reveal more than the page – or else less." 1922: 284), but the study served as an important catalyst, a call for research in the field. Following Woolbert, empirical research into lexical and syntactic-semantic differences between spoken and written English proliferated and is documented in a great number of publications (e.g. Horn 1926, Voelker 1942, Johnson 1944, Bachman-Mann 1944, Fairbanks 1944, Chotlos 1944, Drieman 1962, Horowitz & Berkowitz 1964, Horowitz & Newman 1964, DeVito 1964, 1965, 1966, 1967a, 1967b, Gibson et al. 1966, Gruner et al. 1967, Blankenship 1962, 1974, Poole & Field 1976, Lakoff 1982, Chafe 1982, 1985, Chafe & Danielewicz

1987, Biber 1986a, 1986b, 1988, 1989, 2006, Hughes 1996, Biber et al. 1999). The present section is an attempt at surveying some of the results of these studies, as well as those of other influential writers touching upon the topic. The section presents a non-exhaustive historical outline of the general developments, an outline intended to serve more as a background to Biber's (1988) choice of linguistic features and the further discussions of these, in ensuing chapters, than as a full account.¹² The further discussions may, of course, also refer only to Biber (1988) or even to previous or more recent research not presented in this survey. A vast number of the 67 features studied by Biber (1988), however, were picked out because previous studies had shown them to be apt to differentiate between speech and writing (cf. 1988: 223–245). As Biber's choice of linguistic features to study in the multidimensional methodology is of central concern in the present work, an account of studies that influenced Biber (1988), and through him the present study, is a paramount consideration. Biber's features and methodology will then be explained further in section 2.3.

Some of the earliest studies used word frequency counts as a primary method for distinguishing between speech and writing. Their authors began by investigating spoken and written texts separately without systematically correlating the findings. Horn (1926), for instance, compiled a "basic writing vocabulary" of the "10,000 words most commonly used in writing," while Voelker (1942) listed the 1,000 most frequent words in the "active speaking vocabulary" (1942: 193). Bachman-Mann (1944) and Fairbanks (1944) also studied spoken and written language data respectively, trying to discern differences in patterns of linguistic behavior, in speech and writing, between schizophrenic patients and speakers of "adequate" language (Fairbanks 1944: 19). Although the 1944 studies, as well as that of Chotlos (1944), were not primarily aimed at elucidating the character of spoken and written language per se, the authors made significant contributions to the field of textual variation studies. The studies were part of a program initiated and directed by Johnson (1944), intended to develop reliable and differentiating measures for linguistic pathological diagnosing, and involved the application of several measurements to compute differences in lexical variation – among them type/token ratio (TTR).¹³ Ever since, the TTR measurement has been keenly applied in quantitative studies of spoken and written language and

12 For more comprehensive reviews of the literature, see e.g. Akinnaso (1982), Tottie et al. (1983), Chafe & Tannen (1987) and Atkinson & Biber (1994).

13 TTR was devised by Johnson (1939, 1944) for comparison of spoken and written texts from experimental subjects. It is a measure of the lexical variety, i.e. the vocabulary richness within a text, which expresses the ratio of different words (types) to total words

has mostly been found useful. Its utility for conversational writing texts will be examined in section 4.3 of the present study, as TTR is one of the features used to differentiate among texts in Biber's (1988) methodology.

While the early twentieth-century lexically oriented linguists concentrated on either oral or written language, those interested in the structure of language largely focused on the oral to the exclusion of writing. In the preceding century, scholars had regarded writing as the true form of language. At the end of the nineteenth century this had begun to change; the German brothers Grimm had recorded and studied speech in its own right and, in Britain, Sweet and Jones developed phonetics as a discipline within linguistics. In the US, early twentieth-century structuralists recorded and described the mostly unwritten Native American languages. Influenced by Sapir and Bloomfield, the structuralists tended to treat writing as a purely derivative phenomenon, as "visual speech symbolism" (Sapir 1933: 19) or "not language, but merely a way of recording language by means of visible marks" (Bloomfield 1933: 21). Assuming this derived character of written language, they found no motivation to compare speech and writing. However, after a substantial body of oral linguistic data had been collected and described by the structuralists, American transformationalists guided by Chomsky (e.g. 1964, 1965) came to dismiss naturally occurring spoken language as too random for systematic study. Instead, in the generative-transformationalist paradigm, grammatical intuitions were to be analyzed. The primary data was neat text samples collected by means of verbal elicitation from subjects – samples generally free of performance errors, dialect or register variation, and cues to the situational context of their production. As the data was elicited and not taken from authentic discourse situations, it resembled typical writing more than speech.

All the while, educational psychologists, sociolinguists and discourse analysts found reason to demonstrate the need and validity of studying naturally occurring data from both spoken and written language. Generally accepting the notion that speech holds primacy over writing in children's development, they drew attention to the problems of children's transition to literacy. Bernstein (1964, 1970) propounded that the "restricted code" spoken by working class children and the "elaborated code" of middle-class students partly explained differences in their educational performance. Labov (e.g. 1969, 1972a, 1972b) introduced the study of language in its social setting, addressing the relation of non-standard dialects to

(tokens) in a text; "[i]f in speaking 100 words (tokens) an individual uses 64 different words (types), [his/her] TTR [is] .64" (Johnson 1944: 1).

education and children's reading performance, and devised methods for teachers to bring out the verbal capacities of "ghetto" children (1969). Several others also pointed to the linguistic incompatibility between home and school, e.g. Greenfield (1972) who discussed the under-achievement of lower-class children who speak "oral speech" as opposed to middle-class children who speak "written speech." Reacting to the Chomskyan concepts of linguistic "competence" vs. "performance," Hymes (1972) introduced and defined "communicative competence," which entailed linguistic inquiry beyond the sentence. Alongside Hymes' (e.g. 1964), and Gumperz' (e.g. 1965), anthropological studies of language in context, linguists and communication scholars increasingly included extensive performance data drawn from spoken and written texts in their analyses.

The 1960s saw an outburst of creative research designs in experimental studies of the differences between speech and writing, and a great number of interesting results. Drieman (1962) drew up an assumedly exemplary methodology for the study of textual variation, applying a few general principles for the collection of spoken and written data from subjects for comparison. The data for both protocols (the spoken and the written) was obtained from a restricted number of subjects, each of which was elicited 1) in one and the same sitting, 2) under conditions that were as identical as possible for all sittings and 3) from subjects given identical topics for both protocols. Drieman took care to analyze the texts in their entirety, advising against the chopping of texts in variation studies: "Only the *entire* oral and the *entire* written communication are comparable" (Drieman 1962: 39, original italics). Drieman's subjects were asked to speak and write about pictures, and the results of the quantitative analysis found the written texts to be shorter than the spoken, but to contain longer words, more attributive adjectives and a more varied vocabulary. Horowitz & Newman (1964) also asked their subjects to speak and write about equivalent topics and found spoken language to be more "productive and prolific" (1964: 643), to contain longer stretches of language per unit of time, more repetition and more irrelevant elaboration. Horowitz & Berkowitz (1964) compared three methods of writing (handwriting, typing and stenotyping) to spoken language (obtained in the Horowitz & Newman study). Subjects were given 30 seconds to think about one of two equivalent topics, "What does a good doctor mean to me?" or "What does a good citizen mean to me?" (Horowitz & Berkowitz 1964: 621), and then asked to write, type or stenotype about the topic. Results showed that the faster the writing method, the more spoken-like were expressions, even though none of the written methods proliferated material at the rate of speech. Speaking was found to be "far more elaborative, wordy, and repetitive" than writing (1964: 624) and even though

the steno-typed material by various measures approximated speech (followed by typing), all written material remained significantly different from the spoken.

In his 1964 dissertation, DeVito studied undergraduates' comprehension of written and oral technical discourse on identical topics (DeVito 1964). The texts were obtained from ten male faculty members' publications and each faculty member's oral description of his publication. As the faculty members (speech professors) were skilled communicators, DeVito found no significant difference between the students' comprehension of the written and the oral discourse. Nevertheless, the study and several follow-up articles (DeVito 1965, 1966, 1967a, 1967b) revealed a number of significant results regarding the discourse itself. DeVito found the written material to contain more difficult words, more grammatically simple sentences, greater "density of ideas" (1965: 128) and higher TTRs, i.e. a more varied vocabulary. The written texts were also found to be more abstract, containing more nouns and adjectives, but fewer verbs and adverbs, especially fewer finite verbs. The spoken texts displayed more self-reference terms, more pseudo-quantifying terms (e.g. *very*, *most*, *quite*), allness terms (e.g. *none*, *all*, *every*), qualification terms (*if*, *but*, *except*) and terms indicating a consciousness of projection (e.g. *apparently*, *seems*, *appears*).

Gibson et al. (1966) compared undergraduate students' spoken and written texts, employing the TTR measure as well as Flesch's readability formulas (the reading-ease score and the human interest score).¹⁴ In sum, the spoken texts were found to contain a simpler vocabulary and were significantly more readable and more interesting: "The spoken language style tends to be characterized by fewer different words, words with fewer syllables, shorter sentences, and more personal words than the written style" (Gibson et al. 1966: 450). Portnoy (1973) also compared oral and written behavior among college undergraduates, obtaining cloze scores¹⁵ for the collected material and finding users of short words "more comprehensible when speaking" and users of long words "more comprehensible when writing" (1973: 151). In a study similar to that of Gibson et al. (1966), O'Donnell et al. (1967) studied third-, fifth- and seventh-graders' spoken and written texts about two short films, analyzing the results syntactically in terms

14 Flesch reading-ease score (FRE) is calculated by a formula that includes average sentence length and average syllables per word (Flesch 1948, Castello 2008). Flesch human interest score (FHI) is calculated by a formula that includes percentages of "personal words" (e.g. personal pronouns referring to humans) and percentages of "personal sentences" (e.g. exclamations) (Flesch 1948: 229).

15 Cloze score is a measure of readability rating readers' ability to correctly predict words left out in texts (Portnoy 1973).

of T-units (i.e. “minimal terminable units,” defined by Hunt 1964)¹⁶ and transformations. Results showed that the length of the T-unit and sentence-combining transformations¹⁷ increase significantly with advance in grade level. The written texts of children in grades five and seven had more sentence-combining transformations, indicating that writing is structurally more complex at these levels. Among third-graders, the study found slightly greater structural complexity in speech, which was explained by third-graders’ general unfamiliarity with writing.

In the next two decades a vast number of publications on empirical, quantitative and qualitative research into textual variation saw the light of day. Most importantly, the 1970s brought the beginning of a diversification of the field – a shift from dichotomous reasoning (speech vs. writing) towards the gradual identification of textual genres. As early as 1960, Carroll had identified a few lexico-grammatical patterns distinguishing dimensions of “style” among a number of written registers (including e.g. novels, essays, scientific papers and letters) without mentioning the words “genre” or “register” (Carroll 1960). In 1969, Crystal and Davy analyzed situated language use (“styles”) in the discourse of conversation, radio commentary, religion, newspaper reporting and legal documents (Crystal & Davy 1969), and although they made a point of avoiding the term “register,” their discussion of linguistic characteristics of sample texts nevertheless pointed out functional differences among the types of situated language. A few years later, a comprehensive article by Blankenship (1974) served as a guiding light among tentative efforts at staking out registers (also termed “styles”). In an earlier article (1962), she had analyzed oral and written styles; this time she concentrated on six individuals and their six sub-modes of discourse (conversation, oral impromptu, written impromptu, oral extemporaneous, written extemporaneous and manuscript). Blankenship used established measures (sentence and word length, TTR, cloze scores) and studied practically all variables documented in earlier studies (such as those in DeVito’s), but also extended the analytic dimension to include e.g. the extent of qualifications and proportions of adjectives and prepositions. The results were complex, and Blankenship

16 Hunt (1964) defines T-units as “the shortest grammatically terminable units into which a connected discourse can be segmented without leaving any fragments as residue” (1964: 34). As explained by O’Donnell (1974), a T-unit consists of “one independent clause and the dependent clauses (if any) syntactically related to it” (1974: 103).

17 A sentence-combining transformation converts “a pair of sentences into a single sentence by embedding one in the other” (O’Donnell et al. 1967: 35), e.g. combining “The man was poor” and “The man bought an automobile” into “The man who was poor, bought an automobile” (ibid.).

discussed them for each individual subject. Other studies diversified the field by identifying even more variables that distinguish speech and writing. O'Donnell (1974) found writing to be syntactically more complex than speech (with more T-units containing dependent clauses) but also to contain more gerunds, participles, attributive adjectives, passive constructions, modals¹⁸ and perfective auxiliaries, and noted that these lexical features partially account for giving written clauses a greater average length. In the spoken texts, O'Donnell found only nominal dependent clauses, infinitives and progressive auxiliaries to be more frequent than in writing.

Like O'Donnell, Poole & Field (1976) found more adjectives and passives in writing than in speech. The latter also found greater sentence length in written discourse, but more complex syntactic structures in terms of embedding in oral communication.¹⁹ In speech, Poole & Field, like DeVito (1966), found more adverbs and personal pronouns than in writing. Around this time, syntactic and lexico-grammatical studies, like Poole & Field's, increasingly presented results that were concordant with earlier studies, at least for some features. Chafe (1982), for instance, corroborated DeVito's (1966) finding that speech has more first person references, and Chafe & Danielewicz (1987) agreed with earlier studies with regard to greater vocabulary variety in writing (e.g. De Vito 1965, Blankenship 1974). Because of the large volume and slightly repetitive character of findings in the more recent decades, a list of syntactic and lexico-grammatical features might better serve the purpose of summing up the results to date of research into differences between writing and speech (cf. Akinnaso 1982: 104, Goody 1987: 263–264, Biber 1988: 47, 223–245, Hughes 1996: 33–34, Biber et al. 1999). Below is a non-exhaustive list, which includes some of the studies presented above, but also points to more recent publications. In the literature on the English language, it is generally agreed that the following syntactic and lexico-grammatical differences distinguish writing from speech:

Writing has

- more structurally complex and elaborate constructions, as indicated by features such as longer sentences or T-units and more nominal constructions, e.g. nominalizations (Drieman 1962, DeVito 1964, 1966, 1967a, O'Donnell

18 In later studies, modals are found to be more common in speech than in writing (cf. Coates 1983, Biber 1988, Biber et al. 1999 and section 4.2 of the present study).

19 Poole & Field's (1976) study is at odds with other studies regarding embedding, as subordination generally has been found to be a trait of writing.

et al. 1967, Ochs 1979, Chafe 1982, 1985, Chafe & Danielewicz 1987, Hughes 1996)

- more explicit informational content, with complete idea units and all assumptions and logical relations encoded in the text (Woolbert 1922, DeVito 1966, Olson 1977, Hughes 1996)
- more deliberately organized and planned discourse (Ochs 1979, Akinnaso 1982, Gumperz et al. 1984, Chafe 1985, Hughes 1996)
- more decontextualized, detached and abstract discourse (Blankenship 1974, Olson 1977, Chafe 1982, Chafe & Danielewicz 1987, Baron 2000)
- more subordinate constructions, like relative clauses (O'Donnell 1974, Kroll 1977, Ochs 1979, Chafe 1982, 1985, Hughes 1996)
- more passive-voice constructions (Blankenship 1962, O'Donnell 1974, Ochs 1979, Chafe 1982, Chafe & Danielewicz 1987, Biber 1986a, Biber et al. 1999)
- more gerunds, participles and attributive adjectives (Drieman 1962, DeVito 1966, O'Donnell 1974, Chafe 1982, Biber 1988)
- higher TTR, indicating greater vocabulary variety (Drieman 1962, Horowitz & Newman 1964, De Vito 1965, Gibson et al. 1966, Blankenship 1974, Chafe & Danielewicz 1987, Biber 1988)
- higher lexical density, indicating a higher ratio of content words (Ure 1971, Hughes 1996, Stubbs 1996, Halliday 1985a, 2004)
- longer words (Zipf 1949, Drieman 1962, DeVito 1965, Gibson et al. 1966, Blankenship 1974)
- orthography (e.g. initial capitals) and punctuation that signal syntactic relations, prosody, pauses, illocutionary force (e.g. questions, exclamations) and emphasis (Akinnaso 1982, Chafe 1985, Halliday 1985a)
- fewer contractions (Biber 1986a, 1988, Chafe & Danielewicz 1987)
- fewer demonstrative pronouns and deictic terms (Ochs 1979, Biber 1986a, Chafe & Danielewicz 1987, Biber et al. 1999)
- fewer discourse particles/markers (Biber 1988, Biber et al. 1999)
- fewer first person pronouns (DeVito 1966, Gruner et al. 1967, Chafe 1982, Biber 1988, Wales 1996, Biber et al. 1999)
- fewer imperatives, interrogatives and interjections (Biber et al. 1999)
- fewer modal auxiliary verbs (Coates 1983, Quirk et al. 1985, Biber 1988, Biber et al. 1999, Biber 2004)
- fewer incidences of negation overall, but more synthetic, than analytic, negation (Tottie 1981, 1983b, 1991, Biber 1988, Biber et al. 1999)
- fewer incidences of the causative adverbial subordinator *because* (Beaman 1984, Altenberg 1984, Tottie 1986, Biber 1988)

- fewer or no false starts, repetitions, digressions and other redundancies that characterize informal spontaneous speech (Woolbert 1922, Horowitz & Newman 1964, O'Donnell 1974, Chafe 1982, Biber et al. 1999)

By inference, the above list pertains to writing and speech in a converse way (i.e. the features more frequent in writing are rare in speech; the features rare in writing are more frequent in speech). The list thus presupposes a dichotomous relationship between writing and speech – an opposition. Accounts of this opposition abound in the literature and summaries of the characteristics are cast in lists of the following kind (cf. Horowitz & Samuels 1987: 9, Coleman 1996: 44, Baron 1998: 137, 2000: 21, Crystal 2001: 26–28, Hård af Segerstad 2002: 46):

Writing is	Speech is
endophoric ²⁰	exophoric ²¹
informational	involved
objective	interpersonal
a monolog	a dialog
durable	ephemeral
scannable	only linearly accessible
planned	spontaneous
highly structured	loosely structured
concerned with past and future	concerned with the present
formal	informal
expository	narrative
argument-oriented	event-oriented
decontextualized	contextualized
abstract	concrete

The view of speech and writing as two separate homogeneous entities was common in early linguistic accounts of speech and writing. In the 1970s, as noted above, this ingrained conception was loosened, and in the 1980s it eventually decisively modulated to the notion of linguistic genres. In these decades, influential anthropologists and linguists increasingly concerned themselves with language in

20 Coleman (1996) associates writing with *endophoric* mentality and language, i.e. language constructed for interpretation without reference to extra-linguistic information. “An ‘endophoric’ sentence provides all the necessary information within itself: e.g., ‘William Caxton was the first printer in England’” (Coleman 1996: 43).

21 Coleman (1996) associates speech with *exophoric* mentality and language, i.e. language constructed for interpretation with reference to extra-linguistic information. “An ‘exophoric’ sentence can be understood only if one knows the context or situation from which it emerges: e.g., ‘No, I don’t’” (Coleman 1996: 43).

its real-world social context, aided by new technological and computational means and methods for collecting and studying data. In their investigations, they occasionally found a mismatch between forms of speech and writing, and some of the general characteristics ascribed to “speech” and “writing” in the dichotomy. On the one hand, some spoken and written texts are very similar to each other (e.g. public speeches and written exposition). On the other hand, some spoken genres differ significantly from each other (e.g. conversation and public speeches) (Biber 1988: 36). Usage-oriented linguists (e.g. Tannen, Chafe, Danielewicz, Biber) therefore argued that no linguistic or situational characterization of writing and speech holds true for all spoken and written genres. Instead, the linguistic properties of speech and writing vary from context to context, and a considerable overlap obtains between the two media. Speech and writing are not static representations, but rather comprise a multitude of genres with varying degrees of “spokenness” (orality) and “writtenness” (literacy), genres that are scattered in overlap along a continuum. Accordingly, despite its written mode, a note passed to a classmate during class might assume a higher degree of orality (i.e. more of the characteristics of speech) than a formal oration (which might emulate the traits of writing).

In 1982, Tannen edited a volume that brought out the continuum view on a wide front in discourse studies (Tannen 1982a). In one of the papers, Chafe introduced a study set out to compare four “styles of language” (genres) of academics: conversations, lectures, letters and academic prose (Chafe 1982: 36). Chafe identified sets of linguistic features associated with two dimensions of language among the genres: “integration vs. fragmentation” and “detachment vs. involvement.” In 1985, Chafe expounded the dimensions further with illuminating examples from corpus data (Chafe 1985), and in 1987, Chafe and Danielewicz completed the account with quantitative data for each of the genres studied (Chafe & Danielewicz 1987).

Around this time, Douglas Biber wrote a dissertation (Biber 1984) and published a number of articles detailing a study of 41 linguistic features in hundreds of spoken and written text samples (Biber 1985, 1986a, 1986b) that used multivariate statistical techniques to identify dimensions of variation among sampled genres. The studies were the first multifeature multidimensional (MF/MD) approaches to the study of textual variation in both speech and writing (previous studies using multivariate techniques had analyzed only written registers, e.g. Carroll 1960, Marckworth & Baker 1974). Biber’s studies used large-scale corpora; they provided a quantitative methodology unprecedented in the field; and they set the stage for groundbreaking results. In the next two years, Biber extended his empirical research to include the full range of spoken and

written genres identified up to then: the six genres of speech in the London-Lund Corpus (LLC; see Svartvik 1990), the 15 genres of writing in the Lancaster-Oslo/Bergen Corpus (LOB; see Johansson et al. 1978) and two of his own genres of letters (see Appendix I for a list of all genres). Biber explored the syntactic and lexical findings of previous research in the field to list 67 features likely to distinguish among the textual genres and annotated the texts for these features. By analyzing co-occurrence patterns between the features, through multivariate techniques, Biber was able to discover and define six dimensions of variation among the genres. The results were published in his landmark 1988 book entitled *Variation across speech and writing*, a book that bore out the continuum view at its very onset, its title being “Variation *across* speech and writing” instead of “...*between* speech and writing.” As the methodology of Biber’s (1988) study is at the heart of the present study, it is further described in a section of its own, section 2.3.

Following Biber, variationists analyzing speech and writing have abandoned simple dichotomous distinctions that categorize varieties as either formal or informal, abstract or concrete, etc. Rather, genres/registers are seen to differ from each other by being more or less formal, more or less abstract, etc., and/or to vary on several dimensions at once. To distinguish among the full range of genres in a language, a quantitative analysis needs not only to take into account a large number of co-occurring lexical and grammatical features and interpret these in functional terms, but also needs to base conclusions on large, balanced corpora of texts for all genres and define the dimensions of variation among the genres.

To the writer’s knowledge, only one linguist after Biber has taken on the laborious task of carrying out a full MF/MD analysis of the English language, namely Lee (forthcoming) on the British National Corpus, although a few have used a full MF/MD methodology to map out genre variation in other languages, e.g. Besnier in Nukulaelae Tuvaluan, Kim in Korean and Hared in Somali (see Biber 1995 for all three) and Biber et al. (2006) in Spanish. Instead of carrying out new full MF/MD analyses, linguists studying variation in the English language have tended to apply Biber’s established dimensions to come to understand new or historical genres, registers or subregisters relative to the range of spoken and written genres in Biber (1988), e.g. Conrad, who explored variation in academic texts, Atkinson, who studied scientific discourse across history, and Helt, who studied British and American spoken English (see Conrad & Biber 2001a for all three), or relative to the dimensions identified in Biber (1988), e.g. Geisler (2002), who investigated register variation in 19th-century English. However, a few linguists have conducted a new MF/MD analysis to explore a restricted domain of discourse to determine its dimensions of variation, e.g. Kytö (2000), who stud-

ied 17th-century notes of spoken language, and Reppen (2001), who studied elementary students' spoken and written language. Biber himself, and Biber and Finegan, have also applied the MF/MD model to new domains, e.g. Biber (1991) to primary school reading materials and Biber & Finegan (1992, 2001) to historical registers. Biber's own most extensive study of genre variation in English after Biber (1988) is Biber (2003, 2006) in which a new MF/MD analysis was carried out to discover the patterns of variation in university language.

Besides inspiring a host of genre-specific linguistic studies, the awareness of genre differences raised by Biber's studies, and the increasing availability of (online) corpora, have also resulted in authors of grammars taking aspects of both spoken and written production into account. Biber et al.'s (1999) *Longman grammar of spoken and written English*, for instance, provides comprehensive grammatical descriptions of English from four genres (conversations, fiction, newspaper language and academic prose), documenting how grammatical features are distributed across the genres. In the present study, recurrent reference will be made to Biber et al. (1999); section 4.6 of the present study, elaborating on inserts, particularly draws on Biber et al.'s chapter entitled "The grammar of conversation" (1999: 1037ff). Now, from this survey of the literature on speech and writing, we move on to a summary of Biber's (1988) MF/MD methodology.

2.3 Biber's (1988) dimensions of textual variation

To compare conversational writing to speech and writing, the present study utilizes the methodology and results provided in Biber's (1988) book *Variation across speech and writing*. As mentioned, Biber (1988) identified six dimensions, sliding scales, of variation across spoken and written English, and situated a wide range of genres on each of them. His 1988 study presents the positions of the 23 genres on the six dimensions (1988: 128–160). The present study uses Biber's established dimensions and the positions of the spoken and written genres to describe the new genres Internet relay chat and split-window ICQ chat. This section briefly introduces Biber's procedure for identifying the dimensions, outlines the six dimensions of variation and describes how Biber's methodology is employed in the present study.

The first step in Biber's multifeature multidimensional (MF/MD) analysis (henceforth simply MD analysis) was to select a database of spoken and written texts that would represent a broad range of possible communicative functions served in English. Biber decided to study six genres of speech from LLC (comprising 141 texts, totaling 290,000 words), 15 genres of writing from LOB (comprising 324 texts, totaling 654,000 words) and two genres of letters (private and professional, together comprising 16,000 words); see Appendix I for a list of all texts. Next,

Biber identified the set of linguistic features to study, the 67 features expected to have functional associations in the range of genres to be studied. Most of the features had been shown in previous research to distinguish spoken and written texts (cf. section 2.2), others were “potentially important” as they had been associated with certain communicative functions in different texts (1988: 72). The features fell into 16 major grammatical categories (including tense and aspect markers, place and time adverbials, pronouns and pro-verbs, etc.). Table 2.1 lists all the features in their respective categories. Among the studies mentioned in section 2.2 that influenced Biber’s choice of features were Drieman (1962), Horowitz & Newman (1964), Gibson et al. (1966), Blankenship (1974) as for TTR, Zipf (1949) for word length, Blankenship (1962) for past tense verbs and passives, Poole & Field (1976), Chafe & Danielewicz (1987) for personal pronouns, DeVito (1967a), Marckworth & Baker (1974) for nominalizations, Chafe (1982), O’Donnell (1974) for gerunds, participles and attributive adjectives, Coates (1983) for modals, Chafe (1985), Biber (1986a) for contractions, Ochs (1979) for demonstrative pronouns, Beaman (1984), Altenberg (1984) and Tottie (1986) for adverbial subordinators, Schiffrin (1982) for discourse particles, and Tottie (1981, 1983b) for negation; see Biber (1988: 223–245) for a full survey of other studies backing up his selection of features.

Table 2.1: *Linguistic features studied in Biber (1988)*

Tense and aspect markers	Subordination features	Lexical classes
1 past tense verbs	21 THAT verb complements	45 conjuncts
2 perfect aspect verbs	22 THAT adj. complements	46 downtoners
3 present tense verbs	23 WH clauses	47 hedges
	24 infinitives	48 amplifiers
Place and time adverbials	25 present participial clauses	49 emphatics
4 place adverbials	26 past participial clauses	50 discourse particles
5 time adverbials	27 past prt. WHIZ deletions	51 demonstratives
	28 present prt. WHIZ deletions	
Pronouns and pro-verbs	29 THAT relatives: subj. position	Modals
6 first person pronouns	30 THAT relatives: obj. position	52 possibility modals
7 second person pronouns	31 WH relatives: subj. position	53 necessity modals
8 third person pronouns	32 WH relatives: obj. position	54 prediction modals
9 pronoun IT	33 WH relatives: pied pipes	
10 demonstrative pronouns	34 sentence relatives	Specialized verb classes
11 indefinite pronouns	35 adv. subordinator – cause	55 public verbs
12 DO as pro-verb	36 adv. sub. – concession	56 private verbs
	37 adv. sub. – condition	57 suasive verbs
Questions	38 adv. sub. – other	58 SEEM/APPEAR
13 direct WH-questions		

	Prep. phrases, adjectives and adverbs	Reduced forms and dispref. structures
Nominal forms	39 prepositional phrases	59 contractions
14 nominalizations	40 attributive adjectives	60 THAT deletion
15 gerunds	41 predicative adjectives	61 stranded prepositions
16 nouns	42 adverbs	62 split infinitives
		63 split auxiliaries
Passives	Lexical specificity	Coordination
17 agentless passives	43 type/token ratio	64 phrasal coordination
18 BY passives	44 word length	65 non-phrasal coordination
Stative forms		Negation
19 BE as main verb		66 synthetic negation
20 existential THERE		67 analytic negation

Biber's selection of a large set of features was motivated by the emerging view that no single linguistic parameter in itself can capture the full range of differences and similarities among spoken and written genres. Rather, studying linguistic variation with a macroscopic approach requires the analysis of numerous features in numerous spoken and written texts. Previous research had begun to suggest that sets of features occur together (co-occur) in systematic ways in different texts (e.g. Ervin-Tripp 1972, Brown & Fraser 1979, Chafe 1982). Chafe's (1982) discussion of "integration vs. fragmentation" and "detachment vs. involvement," for instance, proposed limited but specific sets of co-occurring features, e.g. that integration is marked by features that package information in texts, such as nominalizations, participles, attributive adjectives and sequences of prepositional phrases, whereas fragmentation shows up as idea units (sentences) introduced with coordinating conjunctions, or strung together by pauses instead of connectives. Chafe had analyzed texts functionally in order to identify the sets of related features. Biber reversed this approach; rather than proposing dimensions of variation on an a priori functional basis, he set out to first identify groups of co-occurring features and subsequently interpreted these in functional terms.

Biber developed and used computational tools to identify, tag and count the occurrence of each linguistic feature in the texts. After all the linguistic features had been counted and normalized to occurrences per 1,000 words, Biber used a multivariate statistical technique called factor analysis to determine which features co-occurred with a high frequency in texts. The sets of co-occurring features he

then called dimensions of variation. Table 2.2 summarizes the groups of co-occurring features associated with each dimension (adapted from Biber 1988: 102–103).

Table 2.2: Summary of co-occurring features on each dimension (Biber 1988: 102–103)

Dimension 1		Dimension 3	
private verbs	0.96	WH relatives: object position	0.63
THAT deletion	0.91	WH relatives: pied pipes	0.61
contractions	0.90	WH relatives: subject position	0.45
present tense verbs	0.86	phrasal coordination	0.36
second person pronouns	0.86	nominalizations	0.36
DO as pro-verb	0.82	time adverbials	-0.60
analytic negation	0.78	place adverbials	-0.49
demonstrative pronouns	0.76	adverbs	-0.46
emphatics	0.74		
first person pronouns	0.74		
pronoun IT	0.71	Dimension 4	
BE as main verb	0.71	infinitives	0.76
adverbial subordinator – cause	0.66	prediction modals	0.54
discourse particles	0.66	suasive verbs	0.49
indefinite pronouns	0.62	adv. subordinator –condition	0.47
hedges	0.58	necessity modals	0.46
amplifiers	0.56	split auxiliaries	0.44
sentence relatives	0.55		
direct WH-questions	0.52	Dimension 5	
possibility modals	0.50	conjuncts	0.48
non-phrasal coordination	0.48	agentless passives	0.43
WH clauses	0.47	past participial clauses	0.42
stranded prepositions	0.43	BY passives	0.41
nouns	-0.80	past participial WHIZ deletions	0.40
word length	-0.58	adverbial subordinator –other	0.39
prepositional phrases	-0.54		
type/token ratio	-0.54	Dimension 6	
attributive adjectives	-0.47	THAT verb complements	0.56
		demonstratives	0.55
Dimension 2		THAT relatives object position	0.46
past tense verbs	0.90	THAT adjective complements	0.36
third person pronouns	0.73		
perfect aspect verbs	0.48	Dimension 7	
public verbs	0.43	SEEM/APPEAR	0.35
synthetic negation	0.40		
present participial clauses	0.39		

Having identified the dimensions through factor analysis, Biber proceeded to interpret the factors functionally to determine what situational, social and communicative functions the co-occurring features represent. In doing so, he considered not just the likely reasons for linguistic features co-occurring, but also the reasons for sets of features showing complementary distributional patterns. Two of the dimensions consist of complementary sets of features, positive and negative (Dimensions 1 and 3), meaning that when features in one set co-occur frequently in a text, the features in the other set are markedly less frequent in that text, and vice versa (see table 2.2). The other dimensions consist of sets of features that either co-occur systematically with a high frequency, or are systematically infrequent in texts. The features in table 2.2 all displayed salient loadings in the factor analysis, meaning that they are all representative of the underlying dimensions.²² Their respective weight is indicated as a positive or negative number, but the positive or negative sign does not influence the importance of a loading. Attributive adjectives (-0.47) thus have a larger loading on Dimension 1 than do stranded prepositions (0.43). The positive and negative signs simply group together the features that are in complementary distribution in texts.²³

For example, consider Dimension 1 in table 2.2. The features above the dashed line (“positive”) tend to co-occur in texts so that texts with a high frequency of private verbs (e.g. *believe, know, mean, think*) also are likely to display high frequencies of e.g. subordinator-THAT deletion, contractions and first person pronouns (e.g. *I don't think* \emptyset *I am*), etc. The features below the dashed line (“negative”) also tend to co-occur in texts, so that texts with a high frequency of nouns, for instance, are likely to have frequent prepositional phrases and attributive adjectives, and such texts often contain long words and display a high type/token ratio. In addition, the positive and negative groups tend to occur in complementary distribution, meaning that texts with an abundance of positive features (private verbs, contractions, etc.) usually contain markedly few occurrences

-
- 22 Seven features with a weight below 0.35 were dropped from Biber's further analysis (i.e. in his calculation of dimension scores) and, for the sake of simplicity, these are not included in table 2.2. Biber admits to these also being salient, but opts for a conservative cut-off point to ward off an otherwise unwieldy number of features loading on most dimensions. Moreover, to assure the experimental independence of dimensions, features with salient loadings on more than one dimension were included only in the dimension on which they have the highest loading (Biber 1988: 93). For a full account of Biber's factor analysis and dimension score calculations, see Biber (1988: 61–97).
- 23 The weights themselves are not included in the calculation of dimension scores (to be described later in this section as well as in section 3.5).

of the negative features (nouns, prepositional phrases, etc.), and vice versa. On Dimensions 2, 4, 5 and 6, on the other hand, the features simply co-occur in systematic ways, so that on Dimension 2, for instance, past tense verbs tend to be accompanied by e.g. third person pronouns, perfect aspect verbs, etc. in texts, or else these features are markedly infrequent altogether.²⁴

Biber's functional interpretation of the dimensions sought to identify the underlying functional, social and communicative purposes associated with each dimension. His interpretation was based on the assumption that linguistic features co-occur in texts because they reflect shared functions. While the co-occurrence patterns had been derived quantitatively, the functional analysis entailed meticulous qualitative analysis of texts and genres, i.e. the assessment of the communicative functions most widely shared by the sets of co-occurring features, as well as analyses of differences and similarities in the genres and the corpus data. Biber's functional analysis resulted in the following interpretive labels for the six dimensions:²⁵

- Dimension 1: Informational versus Involved Production
- Dimension 2: Narrative versus Non-Narrative Concerns
- Dimension 3: Explicit/Elaborated versus Situation-Dependent Reference
- Dimension 4: Overt Expression of Persuasion/Argumentation
- Dimension 5: Abstract/Impersonal versus Non-Abstract/Non-Impersonal Information
- Dimension 6: On-Line Informational Elaboration

To exemplify, Biber's assessment of the features with negative loadings on Dimension 1 (below the dashed line in table 2.2) yielded the interpretation that these indicate an "informational" focus in texts, i.e. the careful integration of information involving precise lexical choice. Analyzing the co-occurrence patterns of these features in texts, Biber found, for instance, written expository prose to represent such informational production. Sample (1), an excerpt from academic prose, illustrates the co-occurrence of "negative" linguistic features on

24 Biber (1988) identified a seventh dimension, as indicated in table 2.2, but found its factorial structure too weak for functional interpretation and therefore excluded it from further analysis. Most later work by Biber also leaves the sixth dimension out of account (e.g. Biber 1989, 2008). The present study leaves out the seventh dimension, but considers the positions of conversational writing on the sixth dimension, as even tentative results may be worthwhile to explore.

25 The labels here reflect denominations in Biber (1988) and minor denominative elaborations provided in subsequent work (e.g. Biber 1995, Biber et al. 1998, Conrad & Biber 2001b). The first dimension has been cast in reversed order in the present study (originally labeled "Involved versus Informational Production" in Biber 1988).

Dimension 1. The sample is illustrative of informational production as it involves dense integration of information: frequent nouns, long words, an abundance of attributive adjectives modifying the nouns (e.g. *physical mobility, interdependent social factors, extra-familial kin, economic resources, social mobility*), frequent prepositional phrases and sequences of prepositional phrases (e.g. *of a number of interdependent social factors*).

- (1) Degree of physical mobility is only one of a number of interdependent social factors which act directly or indirectly to influence the size of an individual's kinship universe. These factors are also related to the amount of contact the individual has with his extra-familial kin and to the differentiations he makes among them; the most important are occupation, economic resources, ownership of property and degree of social mobility.

Academic prose LOB J: text 30

By contrast, Biber associated the set of features on the “involved” end of Dimension 1 (above the dashed line in table 2.2) with involved production, i.e. with interactive, more involved content. The function most widely shared by the features is the communication of interactive or affective content, and the features reflect on-line production circumstances. First and second person pronouns, direct WH-questions, emphatics and amplifiers, for instance, reflect interpersonal interaction and the involved communication of personal feelings and concerns. Reduced surface forms (e.g. contractions, subordinator-THAT deletion, stranded prepositions) are also markers of such involved production, as well as, for instance, features associated with more uncertain presentation of information (e.g. DO as pro-verb, demonstrative and indefinite pronouns). Among the genres Biber studied, face-to-face and telephone conversations display high co-occurrence of features with positive loadings on Dimension 1. Sample (2) is typical of involved production, a face-to-face conversation with, for instance, frequent first and second person pronouns, direct WH-questions (including those initiated with *how*) and contractions (e.g. *you're, won't, aren't, you'd*).

- (2) B: come in . come in – ah good morning
A: good morning
B: you're Mrs Finney
A: yes I am
B: how are you – my name's Hart and this is Mr Mortlake
C: how are you
A: how do you do .
B: won't you sit down
A: thank you
B: mm – well you are . proposing . taking on . quite something Mrs Finney aren't you

A: yes I am
B: mm
A: I should like to anyhow
B: you know what you'd be going into
A: yes I do

Face-to-face conversations LLC 3: text 1a

In addition, a sample of informational production like (1), with abundant “negative” features on Dimension 1, typically displays a marked paucity of “positive” features (e.g. no first and second person pronouns, no direct WH-questions, no contractions, etc.). Conversely, a sample of involved production like (2) typically displays a marked paucity of “negative” features (e.g. few nouns, few long words, few prepositional phrases, etc.).

On Dimension 3, the “positive” features are all associated with explicit/elaborated reference, whereas the “negative” features are typical of discourse with abundant situation-dependent reference. On dimensions with only “positive” features, such as Dimensions 2 and 5, the presence of the co-occurring features on one end of the dimension is simply countered by the absence of the same features on the other end; the presence of features on both dimensions marking texts as belonging to the first part of the interpretive label and the absence of features marking texts as belonging to the second part of the label. For Dimensions 4 and 6, however, the presence of the co-occurring features marks texts as belonging to the entire interpretative label, whereas the absence of the features refers texts to an opposite end, as having no overt expression on persuasion/argumentation, or no on-line informational elaboration, respectively.

After all dimensions had been interpreted functionally, the final step in Biber’s (1988) MD analysis was to compute dimension scores for the written and spoken genres studied, to situate the genres relative to each other in linguistic space. Dimension scores were computed by summing, for each text, the frequencies of the co-occurring features. Before summing the features, all frequencies were standardized to a mean of 0.0 and a standard deviation of 1.0. The corpus mean, i.e. mean frequencies for each feature in the full range of written and spoken texts, constituted the zero point for the comparison of all genres, and the standard deviation of the features in the full corpus constituted the unit, 1.0, to be measured. Accordingly, as the corpus mean for e.g. past tense verbs was 40.1, with a standard deviation of 30.4, a text with 113 past tense verbs was given the standardized frequency 2.4 for past tense verbs. That is, if the frequency of past tense verbs in the text is 113, and $40.1 + (30.4x) = 113$, it means that $x = 2.4$, i.e. that the score is 2.4 standard deviations higher than the mean. The standardized frequencies of co-occurring features on each dimension were then summed,

and on Dimensions 1 and 3 the sum of the “negative” features were subtracted from the sum of the “positive” features, in order to obtain a dimension score for each text on each dimension. The standardization procedure ensured the comparability of texts across genres, preventing the features that occur very frequently, in terms of normalized frequencies, from having an inordinate influence on the resulting dimension scores. The average dimension score for all texts in a genre was then taken to be that genre’s dimension score. In the present study, the same procedure of standardization and dimension score calculation will be undertaken for the conversational writing genres and for the SBC subset (section 3.5).

As an example illustrating the calculation procedure, Biber (1988: 94–95) considers the genre “general fiction” on Dimension 2. The dimension score for each text in the genre is calculated by summing the standardized frequencies of the co-occurring features. For one of the texts, LOB K: text 6, the calculation involves the summing the standardized scores 2.4 past tense verbs, 4.2 third person pronouns, 4.1 perfect aspect verbs, 1.5 public verbs, 1.4 instances of synthetic negation and 2.3 present participial clauses (Biber 1988: 94–95). The resulting dimension score for the text is thus 15.9 (as $2.4+4.2+4.1+1.5+1.4+2.3=15.9$). The dimension score for the general fiction genre is then found by computing the average dimension score for all texts in the genre. On Dimension 2, general fiction has one of the highest dimension scores among all genres, positioning the genre well into the narrative end of the dimension, or more correctly: the high dimension score of general fiction reveals that the texts in the genre are produced by authors with narrative concerns. The fiction genres (general fiction, mystery fiction, science fiction, adventure fiction and romantic fiction) all range on the narrative end of the dimension, typically displaying sequential descriptions of past events involving third person animate participants, whereas e.g. official documents and academic prose range well into the non-narrative end of the dimension, similar to each other only in their lack of narrative concerns. On Dimension 2, face-to-face and telephone conversations rank in intermediate positions, the latter being slightly more narrative than the former.

Once the dimension scores had been computed for all genres, Biber was able to plot all genres on each of the six dimensions. The dimension plots, in turn, allowed further linguistic characterization of individual genres, the comparison of genres and more conclusive interpretations of the communicative functions underlying the dimensions. Most importantly, the multiple dimension plots proved that no single dimension of variation is adequate in itself to account for the range of similarities and differences, and that there is no absolute difference

between spoken and written language; rather, spoken and written genres show considerable overlap across all dimensions.

As outlined above, a complete MD analysis, like Biber's (1988), involves eight methodological steps. These can be summarized as follows (cf. Biber 2008: 825–826).

1. Design a corpus based on previous research and analysis. Collect, transcribe and input texts into the computer. (Pre-existing corpora can be used.)
2. Identify linguistic features to include, together with functional associations
3. Develop software for tagging the relevant linguistic features
4. Tag the entire corpus
5. (Develop additional software to) compute frequency counts of all linguistic features
6. Analyze co-occurrence patterns using factor analysis
7. Interpret factors functionally as underlying dimensions of variation
8. Compute dimension scores for texts/genres on each dimension, compare with mean dimension scores for other texts/genres.

There are two different kinds of MD study following Biber (1988): those that have conducted full MD analyses (steps 1–8 above) and those that apply Biber's dimensions to new areas of research. The latter differ methodologically from the former in that they leave out steps 6 and 7, that is, they do not require a separate factor analysis as they use the previously defined dimensions. (Examples of both kinds of study were given in section 2.2.) The present study is of the latter kind, implementing steps 1–5 and 8. It involves the collection and annotation of a corpus of conversational writing, UCOW, and the annotation of a subset of face-to-face conversations from SBC. The texts are annotated for Biber's 65 linguistic features (TTR and word length not requiring annotation); feature counts are normalized and standardized; dimension scores are computed for the genres (Internet relay chat, split-window ICQ chat, face-to-face conversations SBC); the genres are positioned on Biber's dimensions and, finally, compared with the dimension scores of Biber's 23 genres.

The present study, however, differs from other MD analyses in that it devotes considerable space to the process of computing frequency counts (step 5 above). As mentioned, the standardization of frequencies involves relating the frequencies of linguistic features to their mean frequencies in Biber's full corpus of spoken and written genres in English. The present study exploits the standard deviations of features in Biber's full corpus to investigate and find out what features in the conversational writing corpus deviate by more than two standard deviations ($|s.d.| > 2.0$) from Biber's full corpus. Such features are particularly frequent,

or infrequent, in conversational writing as compared to speech and writing in general, and can be seen to epitomize the linguistic character of conversational writing in a statistically interesting way. These salient features of conversational writing, explored in chapter 4, are sought among all of Biber's 67 features (cf. table 2.1 above) and not just among those to be included in the dimension score calculation (cf. table 2.2). Before computing the dimension scores, the present study also considers other salient features of conversational writing, those studied in previous accounts of CMC discourse (e.g. modal auxiliaries, paralinguistic features, emoticons and abbreviations) as well as previously understudied aspects of conversational writing, such as its lexical density and inserts (all in chapter 4). This is done in order to bring into view the full range of conspicuous traits in conversational writing before the account zooms in on the features co-occurring on Biber's dimensions (cf. table 2.2). The dimension scores of the genres of conversational writing are then presented and discussed in chapter 5.

The present section has outlined how Biber's and others' multidimensional studies set out from quantitative analyses of co-occurrence patterns among linguistic features, and arrive at functional, qualitative interpretations of underlying dimensions of variation. Yet, some researchers claim that these and previous studies with a quantitative orientation fail to adequately address the important differences between speech and writing. The next section explores some essentially non-quantitative approaches to linguistic variation, of which some, particularly those of M. A. K. Halliday involving social semiotics and functional grammar, will be brought into the present study to complement the MD approach in order to ensure an all-round assessment of conversational writing.

2.4 Halliday's and others' essentially qualitative approaches

Whereas Biber's approach to textual variation is quantitative at its outset, but also applies qualitative, functional interpretation of results, Halliday's approach to textual variation (1985a, 1987) is essentially non-quantitative, except with regard to the calculation of lexical density (explained shortly). Several other linguists in the past few decades have also opted for non-quantitative methods for analyzing speech and writing. Early non-quantitative studies include those of Lakoff (1982) on the mingling of speech in writing and writing in speech, Tannen (1982b) on what oral and literate strategies grow out of communicative goals and context in oral and written narratives, and Tannen (1985) on how differences between speech and writing can be accounted for in terms of their relative focus on either involvement or information, properties listed among those in the dichotomous list in section 2.2. Several other properties of speech and writing listed in the dichotomy in

section 2.2 also stem from qualitative interpretation of early syntactico-semantic findings. A substantial number of qualitative studies of speech and writing have been carried out within the field of discourse analysis (see e.g. Schiffrin 1994, Schiffrin et al. 2001) and, as will be explored below, within social semiotics (e.g. Halliday 1978, Halliday & Hasan 1989, Hodge & Kress 1988) and systemic-functional linguistics, a.k.a. functional grammar (e.g. Halliday & Hasan 1989, Martin 1992, 2001a, 2001b, Halliday 2004). The approaches of Halliday and other functional linguists are applied in parts of the present study, as they enable, for instance, the informed analysis of cohesion and lexical density, as well as the qualitative identification of registers via a set of linguistic metafunctions. The present section serves to introduce the utility and the basic concepts of the Hallidayan functional linguistic approaches.

Critical of earlier quantitative studies' focus on taxonomic differentiations, Akinnaso (1982) proposed the study of spoken and written texts from the viewpoint of thematic cohesion. Cohesion was introduced by Halliday & Hasan (1976) as one of the two text-forming components of the linguistic system, making text cohere within itself and with the context of situation (the other one being intonation). Cohesive resources in language are, for instance, reference, substitution, ellipsis, conjunction and lexical cohesion (e.g. repetition of lexical items). Halliday & Hasan (1976: 23) called the relationship between a cohesive item and the item it refers to a cohesive tie and found that the patterns of cohesive ties "effectively define a text":

The concept of COHESION can therefore be usefully supplemented by that of REGISTER, since the two together effectively define a TEXT. A text is a passage of discourse which is coherent in these two regards: it is coherent with respect to the context of situation, and therefore consistent in register; and it is coherent with respect to itself, and therefore cohesive. (Halliday & Hasan 1976: 23, original emphasis)

Halliday & Hasan's work suggested that by investigating the patterns of cohesive ties it is possible to detect underlying differences between speech and writing. Cohesion is part of the "text-forming component in the linguistic system" (Halliday & Hasan 1976: 27), which Halliday later came to call the textual metafunction (explained shortly). Gumperz et al. (e.g. 1984) pursued the study of cohesion and found, among other things, that cohesion in spoken discourse is accomplished through paralinguistic and prosodic cues, whereas in written discourse cohesion must be lexicalized. Cohesion is further explored in section 4.5 of the present study, in the discussion of paralinguistic features in chat.

In several publications (1979, 1985a, 1987), Halliday elaborated on the cohesive, paralinguistic and prosodic devices available in speech and challenged

the prevailing view of writing as being structurally more complex than speech. Above all, his discussion of lexical density challenged how variationists view and measure complexity in speech and writing.²⁶ Halliday found that spoken language is characterized by complex sentence structures with low lexical density (i.e. more clauses, but fewer lexical words per clause), whereas written language has simple sentence structures with high lexical density (i.e. more lexical words per clause, but fewer clauses). His conclusions were not drawn from systematic large-scale quantitative investigation of spoken and written texts, but from isolated examples. Nevertheless, his assertions have been validated, at least partially, in several other studies (e.g. Beaman 1984, Yates 1993, Stubbs 1996). One of Halliday's major contributions to the study of variation in speech and writing was the concomitant finding of greater grammatical intricacy in spoken language, for whereas writing is lexically dense, speech is lexically sparse – and therefore grammatically dense, or grammatically “intricate”:

The complexity of the written language is static and dense. That of the spoken language is dynamic and intricate. [In spoken language,] [g]rammatical intricacy takes the place of lexical density. (Halliday 1985a: 87)

Halliday's measurement of lexical density will be applied to the conversational writing texts in section 4.3 of the present study, and discussed at length there.

In developing his functional grammar, Halliday sought to understand the variety of language usages. Functional grammar is essentially an oral grammar that Halliday suggests ultimately contributes to the understanding of written communication. It covers far too many aspects to be summarized here (see Halliday 1985b, 2004, Martin 1992), but three of its underlying concepts, the “metafunctions,” are central to the present investigation, as they enable the qualitative distinction of registers, and therefore deserve mention. According to Halliday, “[l]anguage is as it is because of what it has to do” (1978: 19), that is “because of the functions in which it has evolved in the human species” (2004: 31). Language has at least three metafunctions:²⁷ 1) “ideational,” i.e. it can represent ideas

26 The lexical density of a text is the proportion of lexical items (content words) to the total discourse (Halliday 1985a, 1987).

27 Halliday uses the term “metafunctions” to set the concepts apart from “functions” as “there is a long tradition of talking about the functions of language in contexts where “function” simply means purpose or way of using language, and has no significance for the analysis of language itself” (Halliday 2004: 31). Metafunctions are “intrinsic to language: that is to say, the entire architecture of language is arranged along functional lines” and the term “metafunction” was adopted in systemic-functional theory “to suggest that function was an integral component within the overall theory” (ibid.).

and relationships of meaning, 2) “interpersonal,” i.e. it serves as a medium of exchange between people, enacting social relationships, and 3) “textual,” i.e. it functions to structure, organize and hold itself together.

Halliday describes language as social semiotic.²⁸ The metafunctions are components of the semantic system in language, “the modes of meaning that are present in every use of language in every social context” (1978: 112). Any given text is thus a product of all three metafunctions. Social semiotics provides a sociological view of semantics, an interface between the social system and the linguistic system. The social context in which a text comes to life is a not just a situation, it is a situation type. The semiotic structure of a situation type can be represented as a complex of three elements: the “field,” i.e. the social action in which the text is embedded, the “tenor,” i.e. the role relationships between the participants and the “mode,” i.e. the channel selected for the communication (including the medium, spoken or written). The three elements together form a conceptual framework for describing the semiotic environment in which people exchange meanings. Detailed specification of the context in terms of its semiotic field, tenor and mode can enable the prediction of a register, that is, the meaning potential typically associated with a given situation type. In his work, Halliday elucidates the systematic correspondence between the semiotic structure of the situation type (the situational elements field, tenor and mode) and the metafunctions. Each metafunction is determined or activated by a particular aspect of the situation; the ideational is activated by features of the field, the interpersonal by features of the tenor and the textual by features of the mode. Table 2.3 outlines the systematic correspondence between the metafunctions and the semiotic structures.

The field, tenor and mode together determine the functional variety, i.e. the register, of the language being used (cf. section 1.4). Language varies with the functions it is being made to serve: what people are doing while speaking or writing, who they are (in terms of statuses and roles) and what exactly the language is being used to achieve (Halliday 1985a). These three variables (what is

28 Semiotics, or “semiology,” was defined by Saussure as a “science that studies the life of signs within society” (Saussure 1966: 16). Social semiotics, a branch of semiotics, is the study of signs and messages in their social and cultural context. Halliday (1978) introduced social semiotics into linguistics to enable the exploration of language as a system of meaning-potential on a higher level than in the tristratal system of semantics, grammar and phonology – a general semiotic level. Each of the three systems (semantics, grammar and phonology) is a system of potential, but constitutes only the realization of a higher-level system, which Halliday defines as “a behavioural system or more generally as a social semiotic” (1978: 39).

going on, who are taking part, and what role the language is playing) respectively indicate what Halliday refers to as field, tenor and mode, and for ease of interpretation they too are inserted into table 2.3, a summary point of reference for the functional analyses in chapters 4 through 6.

Table 2.3: Halliday's three metafunctions in language and related concepts

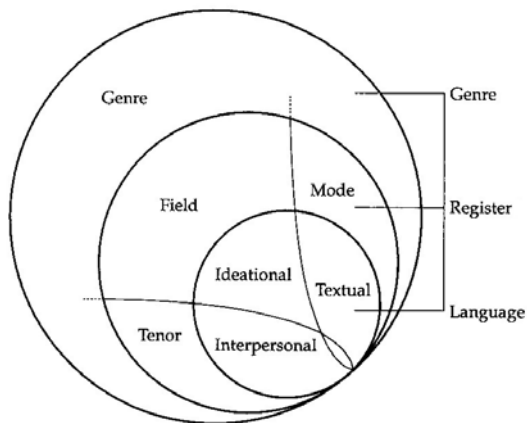
metafunction	semiotic	indicates...	clause as...
ideational	field	what is going on	representation
interpersonal	tenor	who are taking part	exchange
textual	mode	what role the language is playing	message

Halliday's notion of discourse (language) is "the exchange of meaning in interpersonal contexts of one kind of another" (1978: 2). Maintaining that language does not consist of sentences, but rather of text, or discourse, he proceeds to analyze discourse on the clausal level, seeing the clause as "the most significant grammatical unit" (1985b: 101) for representing meaning. Halliday distinguishes three lines of meaning in the clause, i.e. the clause is the product of three simultaneous semantic processes: it functions simultaneously as a representation (in the ideational metafunction), an exchange (in the interpersonal metafunction) and a message (in the textual metafunction) (1985b, 2004). The three metafunctional lines of meaning are realized grammatically in the clause as, for instance, transitivity (in the ideational line), mood and residue (in the interpersonal line) and cohesion (in the textual line). How some of these metafunctional lines may be discerned within the structure of texts will be shown in connection with examples in the present study, when it comes to the functional interpretation of computer-mediated texts. Table 2.3 summarizes the concepts to be brought into consideration. Chapter 4, for instance, discusses several of the lexico-grammatical carriers of meaning, e.g. modality and personal pronouns, which realize interpersonal aspects of the communication. Although only the basic concepts of Halliday's functional grammar are employed in the study, the theoretical framework is believed to provide elucidating clues to the nature of conversational writing as a genre (or register, in Halliday's terms). In the discussion of prevalent linguistic features found in the computer-mediated discourse, the present study will also consult other social semiotic studies (e.g. Fowler & Kress 1979, Hodge & Kress 1988) that provide insights with regard to the parameters and relationships involved in the communication.

Building upon, and complementing, Halliday's and Hasan's work in systemic-functional linguistics, Martin (2001a, 2001b) models language and its connotative semiotics using co-tangential circles; see figure 2.1. The figure visualizes

the stratified model of context in systemic-functional interpretations, in which language is seen to function as “the phonology register, and both register and language function as the phonology of genre” (Martin 2001b: 156). To fully interpret the meaning of a text (language), we take all aspects of context into account, contexts both of situation (register) and of culture (genre). Register is thus “a pattern of linguistic choices, and genre a pattern of register choices” (Martin 2001a: 46).

Figure 2.1: *Metafunctions in relation to register and genre in semiotics (adapted from Martin 2001a: 46).*²⁹



As mentioned in section 1.4, the systemic-functional notion of genre will not be expanded upon in the present study, but the notion of register (field, tenor, mode) and its instantiation as language (the field, tenor and mode phased together in a text) will be touched upon. The present section serves as a background to these considerations, but Halliday’s and others’ semiotic approaches will also be further explained and discussed in connection with relevant textual examples, in chapters 4 and 5.

Having surveyed the previous literature on speech and writing (in section 2.2) and elaborated on quantitative and qualitative approaches in Biber’s and Halliday’s frameworks (in sections 2.3 and 2.4, respectively), the present chapter now turns to an account of linguistic approaches to computer-mediated discourse – a survey of the linguistic literature on computer-mediated communication.

29 Permission to use the figure was obtained from the author and the publisher.

2.5 Survey of the literature on CMC

A chronological survey of linguistic research into CMC presupposes the reader's basic conversance with the development of CMC, particularly with regard to the emergence and current relevance of various modes for the communication. Such a survey therefore necessitates an initial brief description of Internet history, preferably non-technical, upon which the survey of linguistic literature on CMC might follow and make greater sense. Accordingly, this section begins with a non-technical account of the basic background concepts, before homing in on the linguistic studies.

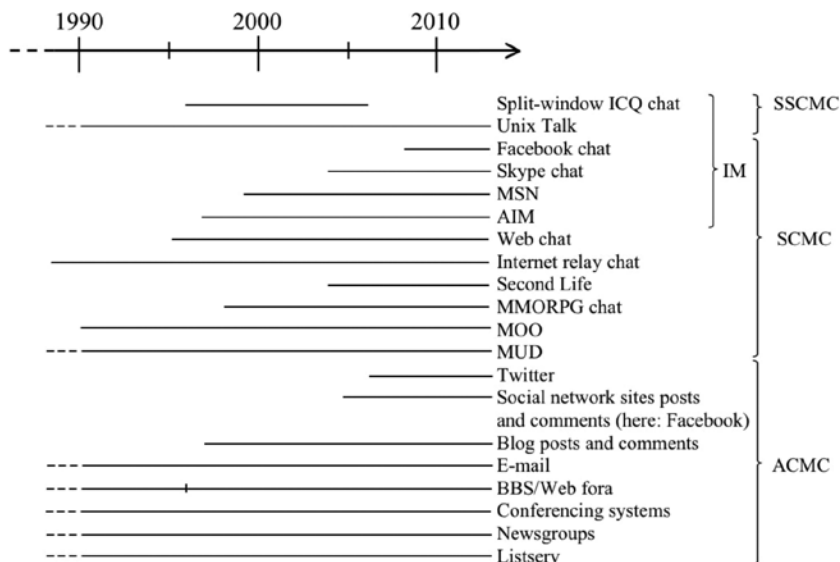
Contrary to popular belief, the advent of human computer-mediated communication dates back to a time before the Internet. CMC originated in the late 1960s when ARPANET officials (in the Advanced Research Projects Agency Network, funded by the US Department of Defense) first managed to enable communication between computers in geographically separate areas. The first CMC message was sent in 1969, comprising three letters ("LOG" between UCLA and Stanford University; see Gromov 1995), and the first e-mail was transmitted in 1972 (Hafner & Lyon 1996). Experimental at first, computer networks remained a means for limited interpersonal communication primarily among computer scientists in the 1970s, who transferred e-mail and, curiously, invented text-based multi-user adventure games, MUDs,³⁰ as early as 1979. The late 1970s also saw the first dial-up BBS, bulletin board system, for storing and sharing data, bulletins and messages. In the 1980s, much scientific effort was invested into developing functional tools for CMC. A few modes, like Unix Talk and VAX Phone, predominantly remained tools for communication among computer professionals (Schulze 1999). Other modes, such as e-mail, early computer conferencing systems (besides BBS), newsgroups and listservs, soon caught on among academic and business users, mostly in American elite universities and organizations (Herring 2001, Baron 2008). The latter development followed upon the ARPANET turning into the Internet in the early 1980s and alongside the development of client-server protocols.

Internet relay chat was invented by Jarkko Oikarinen, a Finnish undergraduate, in 1988, enabling synchronous social communication, one-to-many, outside of game-play (MUD/MOO). In the 1990s, with the rise of commercial Internet service providers, the Internet was rapidly popularized among the general public, and an unprecedented kind of textual mass communication evolved, facilitated by the development of the World Wide Web and a gradual increase in the versatility

30 For a brief explanation of modes and abbreviations, see footnote to figure 1.1 (chapter 1).

of transmissible information formats. For a graphic summary of approximate dates for the emergence of Internet-wide CMC modes, see figure 2.2. In 1996, web fora began to supplant BBSs, although other conferencing systems persisted, and in 1997, web logs (blogs) appeared. The emergence of web chat eludes definition as to a specific year, but, like web fora and blogs, it evolved after the popular introduction of web browsers in the mid 1990s. Unlike web fora (which replaced BBSs), however, web chats did not replace Internet relay chat; rather, IRC persists to this day. The 1990s, nonetheless, saw the emergence of several stand-alone modes for conversational writing, among them the first commercial instant messaging (IM) applications AIM and MSN, as well as ICQ, and the development of MMORPGs incorporating chat.³¹ In the first decade of the third millennium, several major text-based CMC modes appeared, including the IM application Skype chat and the virtual world Second Life (both for SCMC), and the microblogging service Twitter (for ACMC). From 2006, new ICQ versions no longer included the split-window option (which had enabled SSCMC) and ICQ became mainstream IM software (SCMC).

Figure 2.2: Approximate emergence of modes for written CMC.



31 “Stand-alone” means that the mode is accessed in a separate piece of software, in this case outside of the web browser.

Finally, the most penetrating modes for CMC in the past decade have been those in social network sites such as My Space, emerging in 2003, and Facebook, from 2004, involving ACMC in posts and comments, as well as SCMC, as in Facebook chat, launched in 2008, and those for media sharing, e.g. YouTube, launched in 2005, involving limited written ACMC. Today, the Internet is a communication channel for more than three billion connected users (De Argaez 2015), instantly sharing messages, images, audio content, documents, videoclips, software and any other conceivable computer-retrievable information, nearly half of which are users of Facebook (Protalinski 2015), and although Sinophone Internet users are beginning to outnumber Anglophone, English remains the dominant language in, for instance, web site content (De Argaez 2015, W3Techs 2015).

With the obvious exception of Internet telephony and video telephony, linguistic communication over the Internet, to this day, is largely written (cf. Herring 2011b). A few text-based modes have superseded others, just as web fora supplanted BBSs, and some have grown larger than others, e.g. Facebook chat overtaking Internet relay chat by far, in number of users, but the incentive for human communication remains the same – emerging CMC modes simply facilitate our synchronous and asynchronous textual transactions in ever-reconfigured ways. The modes investigated in the present study may seem slightly dated, but really represent conversational writing equally well as would more recently emerged real-time chat modes. Leading CMC scholar Susan Herring, widely recognized as the founder of the field of CMC discourse studies, proposes in Herring (2004b: 33) that “[d]espite the availability of increasingly sophisticated multimedia protocols, CMC remains predominantly grounded in ‘old’ textual practices,” even when different protocols are united in one browser-accessible format. In line with this view, Herring (2013a) cautions communication scholars against mistaking reconfigured phenomena for new forms of computer-mediated discourse. Some recently emerged modes may appear different on the surface, but really have “traceable online antecedents” (2013a: 10). Herring goes on to exemplify how Facebook status update utterances show syntactic, semantic and pragmatic similarities to messages in IRC, MUDs and MOOs (as presented in Werry 1996, Cherny 1994, 1999) and how retweeting (re-posting a message on Twitter) is a modern form of the older practice in textual CMC of “quoting” in asynchronous messages (as shown by Severinson Eklundh 2010). In a different publication, Herring (2013c) discusses the grammar of electronic communications, exemplifying richly from several modes of text-based CMC

as well as SMS.³² Although she claims that “e-grammar” varies across modes, her account of English CMC typography, orthography, morphology and syntax shows considerable similarities across early and more recent modes, suggesting that many early e-grammar innovations have carried over from mode to mode, e.g. from chat to SMSs (e.g. nonstandard typography, such as smileys and the occasional substitution of words or part of words with numbers or letters to save keystrokes).

Another trace of an online antecedent is observed in Herring (2013b: 250), in which it is mentioned that the IRC protocol basically was “borrowed” to create applications such as AIM, web chat and MMORPG chat. It is thus reasonable to deduce that the two modes investigated in the present study, IRC and split-window ICQ chat, recorded in 2002 and 2004 respectively, represent conversational writing equally well as would any IM application, for instance Facebook chat. IM applications particularly share an important situational variable with ICQ, in that they predominantly involve private chat between individuals acquainted in their offline lives. In fact, ICQ is an instant messaging program; its distinct position in the present study is motivated only by its supersynchronicity variable, the only medium variable it did not share with other IM software in the course of ICQ’s decade-long featuring of split-window chat. As if to further endorse the continued relevance of the modes investigated in the present study, there is a passage in Herring (2013a) maintaining that:³³

There is a need to trace relevant antecedents to gain perspective where familiar online discourse phenomena are concerned, in order to do conscientious research. This, in turn, requires some familiarity with earlier CMDA research. Alternatively, familiar phenomena may simply be passed over by researchers in favor of newer, more ‘exotic’ CMD phenomena. (Herring 2013a: 10)

Furthermore, in a commentary to Thurlow & Mroczek’s (2011) co-edited volume on “Digital Discourse,” in which several contributors tend to be dismissive of past research to justify their own approach, Herring (2011a) admonishes that:

32 Herring (2013c) includes SMS among CMC modes.

33 CMDA means computer-mediated discourse analysis and CMD, accordingly, computer-mediated discourse.

Critique is valuable, but in a young field such as computer-mediated discourse studies, which has yet to achieve a widely recognized critical mass, it should build upon, rather than seek to replace, what has already been done. (Herring 2011a: 345)

With Herring's admonishments in mind, this section now turns to a survey of some of the influential publications in the young field of computer-mediated discourse studies.

The scholarly study of the linguistic nature of CMC began in the 1980s, when some scholars became exposed to the first interactive modes. Five noticeable publications on early CMC discourse appeared, the first four on English: Baron (1984), speculating on the effects CMC may have on language change; Murray (1985, 1988), describing CMC discourse in e-mail and a messaging system at IBM; Spitzer (1986), focusing on writing styles in computer conferences; and Severinson Eklundh (1986), detailing a study of letters in the Swedish COM conferencing system. In the early 1990s, language scholars were increasingly exposed to, and intrigued by, the discourse in the emergent media. As early as 1991, Reid, although not a linguist, discussed the deconstruction of social boundaries and the construction of alternative communities in IRC, presenting the social discourse of the mode (Reid 1991). Also in 1991, Ferrara et al. (1991) took on synchronous "interactive written discourse" as an emergent register, finding structural properties similar to e.g. note-taking in the discourse, such as the omission of unstressed pronouns, articles and finite forms of the copula, as well as the shortening of words through abbreviations. Two noticeable publications on text-based virtual reality discourse also appeared in the early 1990s: Reid (1994) on MUDs as sites for social interaction and cultural formation, and Cherny (1994) on discernible gender differences in MOO.

Focusing on ACMC, Yates (1993, 1996) presented a comprehensive study of a large computer conferencing corpus collected from the CoSy system at the Open University, UK. Yates compared his corpus with LLC and LOB, applying Halliday's model of semiotics in the analysis of the ACMC data. His results showed, for instance, that the "field" of the interaction is the text itself and that such a context-free field might be a reason for high levels of modality. In the ACMC discourse, he found a significantly higher use of modal auxiliaries in ACMC than in either speech or writing. Yates (1996) explains the high levels of modality thus:

Not only must the text carry the social situation, it must also carry the participants' relationship to the situation, their perception of the relationships between the knowledge and objects under discussion. (Yates 1996: 46)

Yates argues that the lean semiotic field also has implications for the semiotic “tenor” of the communication, the interpersonal metafunction, as the presentation of self is limited to the CMC text itself; high levels of first and second person pronouns in the ACMC discourse simply result from users’ recurring presentations of themselves. The “[semiotic] mode” of the ACMC, finally, he describes as “neither simply speech-like nor simply written-like” (1996: 46, as mentioned in section 1.1). Yates approaches the textual aspects of ACMC by considering the TTR of the texts, as well as Halliday’s measurement of lexical density, concluding that although ACMC “bears similarities in its textual aspects (e.g., type/token ratio and lexical density) to written discourse, it differs greatly in others, namely pronoun and modal auxiliary use” (1996: 46). In his 1993 full account of the study, Yates explains that he counted the frequencies of ten of Biber’s (1988) features, e.g. pronouns, TTR and modals, but that the study of these was essentially driven by theoretical interest and claims. His application of Biber’s methodology is limited; one result merely draws upon the high first and second person pronoun use and a low third person pronoun use in ACMC, which Yates suggests indicates “that [A]CMC is a subjectively involved and non-narrative form of communication” (1993: 118). The study does not position the ACMC genre on Biber’s dimensions.

As touched upon in section 1.5, Yates’ study serves as an important catalyst to the project described in the present study. The present work partly attempts to parallel Yates’, although with regard to SCMC and SSCMC. It is inspired by Yates’ application of the Hallidayan model of semiotics and Biber’s multi-feature approach, but reverses the significance attributed to these in Yates’ study; in the present study, the full extent of Biber’s (1988) MD methodology, i.e. all 67 features, are considered and the genres positioned on Biber’s dimensions, whereas a more limited Hallidayan semiotic analysis is conducted. Moreover, the layout of chapter 4 here is partly conditioned by adherence to Yates’ findings with regard to the Hallidayan concepts of field, tenor and mode, enabling comparability with Yates’ study. That is to say, after some introductory remarks, chapter 4 opens with a discussion of modal auxiliary use in conversational writing (ideational) and proceeds with an account of personal pronoun use (interpersonal). Next, chapter 4 discusses word length, TTR and lexical density, all in order to explore the textual aspects of the communication, before zooming in on the most salient features of conversational writing.

Another study germane to the present study is the one presented in Collot (1991) and Collot & Belmore (1996). Collot’s (1991) is the only investigation, to the

present author's knowledge, to have positioned a genre of CMC on all of Biber's (1988) dimensions, in Collot's case asynchronous BBS communication. Collot collected and annotated an ACMC "electronic language" corpus with two components, texts composed online and those composed offline, and positioned the two components (genres) on the dimensions. The results of Collot's feature counts are valuable points of reference for the comparison of SCMC and SSCMC to ACMC. The present study uses the online component of Collot's corpus, essentially equivalent to e-mail communication, to represent ACMC in the presentation of feature count data in chapter 4. That is, in the treatment of salient features in conversational writing, graphs in chapter 4 indicate average figures for SCMC and SSCMC as well as for ACMC (Collot's online corpus), writing (Biber's written genres) and speech (Biber's spoken genres + the part of SBC annotated for the present study). Collot's corpus will be further described and exemplified in chapter 4, and the position of its online component briefly commented upon in section 5.1 and discussed in chapter 6. Collot's (1991) feature count data were chosen over Yates' (1993) to represent ACMC in the present study, simply because the former cover a greater range of features. In the discussion of lexical density, however, Yates' figure will be adduced (part of section 4.3), as Collot (1991) did not study the lexical density of her texts.

Yates' and Collot's studies were both presented as chapters in Susan Herring's (1996a) book *Computer-mediated communication: Linguistic, social and cross-cultural perspectives* (Yates 1996, Collot & Belmore 1996), a ground-breaking collection of essays that helped to stake out the direction of CMC research in at least two disciplines, linguistics and sociology. With methodological discussions and empirical results, the book combined perspectives on several issues and laid the groundwork for the linguistic inquiry into CMC. Noticing how linguists generally had "been slow to consider computer-mediated language a legitimate object of inquiry" (1996a: 3), Herring set out to promote exemplary linguistic studies in her book, to motivate further research. Out of Herring's (1996a) five chapters with linguistic perspectives, four have a bearing upon the present study; besides Yates' (1996) and Collot & Belmore's (1996), also Werry's (1996) on the discursive properties of IRC, commented upon in section 4.5 on paralinguistic features, and *passim*, and Herring's (1996b) on gender differences in listserv messages, which is relevant to a brief discussion of gender differences in ICQ emotives usage (emoticons and sentiment initialisms) in section 4.6.

Herring's (1996a) call for research had the desired effect; linguistic CMC research gained impetus towards the end of the 1990s and has continued to

evolve alongside the emergence of new, and reconfigured, CMC modes. Not least, a number of significant studies have been published in the scholarly journals *Journal of Computer-Mediated Communication* (est. 1995) and *Language@Internet* (est. 2004), of which Herring has been editor-in-chief for several years, and a few “handbooks” for studies of online language have appeared, e.g. Crystal (2001, 2011a), Baron (2008) and Herring et al. (2013). Nevertheless, while some aspects, such as innovative orthography and neologisms, have been diligently explored (e.g. by Jonsson 1998, Schulze 1999, Crystal 2001, 2004a, Baron 2008, Waldner 2009, Rowe 2011), other aspects are largely understudied, leading many scholars to concur that the field of CMC discourse studies is still in its infancy. The remainder of this section divides the survey of linguistic CMC studies into three domains: studies involving ACMC; those investigating SCMC, briefly elaborating on two conversational writing analyses relevant to the present study; and finally, studies of SSCMC.

Studies of ACMC discourse have covered various aspects of most asynchronous modes; representative publications include Baym (1996) and Severinson Eklundh (2010) on newsgroups, Davis & Brewer (1997) on computer conferencing, LeBlanc (2005) and Biber & Conrad (2009: 190–198) on web fora, Baron (1998, 2000), Zitzen (2004), Anglemark (2009), Cho (2010), Georgakopoulou (2004, 2011b) and Rowe (2011) on e-mail, Nilsson (2003), Scoble & Israel (2006), Anglemark (2009) and Peterson (2011) on blogs, Lee (2011) on Facebook status updates, and Petrović et al. (2010) and Pak & Paroubek (2010) on Twitter. Related to the ACMC discourse field is the study of text messaging, SMS, still sparingly explored by linguists, even though significant contributions are made in Hård af Segerstad (2002), Ling (2005) and Ling & Baron (2007, 2013).

When it comes to SCMC, the “older” modes MUD/MOO and IRC, as yet, have received more attention than the “newer” (cf. figure 2.2). Language use in the text-based MUDs/MOOs has been studied from various perspectives by e.g. Turkle (1995), Cherny (1994, 1999) and Herring (2013b), whereas chats in graphic virtual worlds have been less explored, although see e.g. Örnberg (2003) on “linguistic presence” in three virtual worlds (On-live Traveler, ActiveWorlds, Anarchy Online), Herring et al. (2009) on the chat in an online first-person shooter game, and Newson (2011) on chat in the MMORPG World of Warcraft, for exceptions. Similarly, the “older” mode IRC has received more scholarly attention than “newer” IM modes. After Reid (1991) and Werry (1996), mentioned above, linguistic studies of IRC and other synchronous IRC-like online chat include Ko (1994, 1996), Jonsson (1998), Schulze (1999), Mar (2000),

Ooi (2002), Freiermuth (2003), Lin (2007), Forsyth (2007), Forsyth & Martell (2007), Waldner (2009) and Herring (2013b). By contrast, studies of IM have a shorter history; empirical milestones include Hård af Segerstad's (2002) study of a Swedish university IM system called WebWho, Baron's (2004, 2010) partly gender-differentiated studies of AIM conversations among college-age students, Squires' (2007) investigation of gendered use of apostrophes in AIM (females used more) and Tagliamonte & Denis' (2006, 2008) comprehensive study of IM among Canadian teens – of which all, except Hård af Segerstad's, are on English. Most of the IRC and IM studies mentioned will be referred to and/or explained *passim* in the present study, that is, they will be brought in whenever relevant to discussions of data and results. Two of the studies, however, deserve to be introduced here as they pertain to the methodology of the present study: Ko (1994, 1996) and Freiermuth (2003), both corpus-based analyses of computer chat.

Ko (1994) compiled a minimal 2,000 word corpus of synchronous classroom chat between students (from a Daedalus InterChange system) and annotated the text for 28 of Biber's (1988) features, those co-occurring on Dimension 1 (see table 2.2). Ko compared the feature counts from his chat corpus with Biber's counts for face-to-face and telephone conversations from LLC (to represent speech), and academic prose and official documents from LOB (to represent writing), but instead of computing a dimension score for the classroom chat corpus, he divided the features into three distributional patterns. Into the first pattern fell features with frequencies intermediate between the frequencies of speech and writing; the second pattern involved features more frequent in chat than in either speech and writing, and the third pattern consisted of features less frequent in chat than in either speech or writing. The features in the first pattern showed a distribution in chat noticeably more akin to speech than to writing (e.g. an abundance of first and second person pronouns). The second and third patterns interestingly distinguish Ko's chatted text from both speech and writing. Six features were most frequent in the chatted text: WH-questions, indefinite pronouns, BE as main verb, WH-clauses, discourse particles and analytic negation. Conversely, six features were least frequent in the chatted text: nouns, prepositions, attributive adjectives, hedges and sentence relatives – the TTR of the chatted text also being the lowest of all three corpora. In 1996, Ko published a slightly modified version of the study, this time with speech represented only by LLC face-to-face conversations, and writing only by LOB official documents. The 1996 version presents the same three-fold distributional pattern and the

same features in respective group, except for the feature second person pronouns (which this time is among the features most frequent in chat).

Ko's (1994, 1996) findings may be indicative of distributions in synchronous CMC, but his minimal corpus size, comprising one single 2,000-word text, is problematic. Biber (1990) asserts that samples of *ten* texts are required to reliably represent a genre, and that each sample should contain a minimum of 1,000 words to make frequency counts stable across samples (see also Biber & Finegan 1991). Consequently, the corpora compiled and annotated for the present study each comprise ten texts or more; see section 3.1 of chapter 3 (Material and method) for details. Ko's indicative findings may, nevertheless, be worthy of further consideration in connection with results obtained in the present study and will be referred to, whenever relevant.

Freiermuth (2003) compiled three corpora, of 3,000 words each, for comparison between speech, writing and synchronous chat from one and the same content domain: political discussion. The spoken corpus was transcribed from a TV-show called *Politically Incorrect*, the written corpus was sampled from the editorial section of the *Standard Times* newspaper, and the chatted corpus was recorded from an America Online political chat channel called *From the Left*. Freiermuth did not use Biber's (1988) methodology, but rather annotated the texts for the grammatical and functional features defined by Chafe & Danielewicz (1987) as apt to distinguish between spoken and written genres. The features can be broadly grouped into five categories: vocabulary variety (e.g. TTR), vocabulary register (literary vs. colloquial vocabulary, contractions), syntactic integration (e.g. prepositional phrases and sequences of these, attributive adjectives and participles), sentence-level conjoining (e.g. clausal coordination) and markers of involvement vs. detachment (such as first person pronouns, phrases like *you know* and responses to questions, which mark involvement, and passives, which mark detachment). Most of the features in Freiermuth's chatted texts showed a frequency distribution intermediate between speech and writing. Only two features were more frequent than in either speech or writing, viz. questions and, surprisingly, passives, whereas several features were more rare in the textual chats, e.g. prepositions, participles and *you knows*. To the extent that Chafe & Danielewicz' feature definitions coincide with Biber's (1988), Freiermuth's chat corpus results will be commented upon in the present study, even though Freiermuth's (2003) chat corpus, like Ko's (1996), is on the small side. However, throughout chapter 4, the views of Chafe & Danielewicz (1987), as well as those of Chafe (1982, 1985), will be brought in on a fairly wide front to elucidate discussions.

Last, but not least, this survey of the literature on CMC turns to linguistic studies of SSCMC. Despite hunting high and low for these, the present author has managed to detect only one extensive such study, Anderson et al. (2010), although several mention split-window ICQ chat or other supersynchronous protocols like Unix Talk and VAX Phone in passing, e.g. Jonsson (1998), Condon & Čech (2001), Herring (2002, 2007), Hård af Segerstad (2002) and Baron (2008, 2010). A few studies of linguistic significance have also appeared on the fringe of the discipline, for instance those in psychology exploring the effects of SSCMC on turn-taking (McGrath 1990, Woodburn et al. 1991, Van der Wege & Clark 1997 and Babineaux forthcoming).

Unix Talk was the earliest supersynchronous protocol, available in the 1970s, soon followed by VAX Phone, in which the communication window splits horizontally into two or three sections, depending upon the number of interlocutors,³⁴ and in which the transmission of text occurs keystroke by keystroke. ICQ chat, launched in 1996, built upon these functions for its split-window mode (Herring 2002). While Talk and Phone have mostly been used by computer professionals with access to Unix and VAX operating systems, and less so today than before, ICQ was widely popular among the general public in English-speaking countries several years into the third millennium, reaching over 100 million users in 2001 (DeCoursey 2001), and continues to thrive in certain countries, for instance Germany and Russia. Today, ICQ is no longer available for written SSCMC but for SCMC (as well as voice and video calls) and ACMC, on computers and cell phones. “Split-window ICQ chat” in the present study, as implied by the designation, denotes *only* the split-window mode that allows supersynchronous written communication, the function that set ICQ apart from the other modes studied. It is the communication carried out in split-window modes that is understudied linguistically; as mentioned, the only extensive linguistic account of SSCMC found is Anderson et al. (2010). The SSCMC studies mentioned in this section are all on split-window communication, although none specifically on ICQ.

Anderson et al. (2010) investigate interaction management in three-person VAX Phone written conversations, finding that users appropriate and adapt “many techniques from face-to-face conversations for the local management of conversations, including turn taking, turn allocation, and explicit interruption management” (2010: 1) but also violate these; rather than follow the face-to-face conversation principle of “no gap, no overlap” (Sacks et al. 1974, Anderson et al. 2010: 9), whereby most face-to-face conversationalists allow gaps for no more

34 Unix Talk permits chat between two participants only.

than three seconds and avoid overlapping each other, the VAX Phone chatters accomplish their turn exchange by the use of “overlapping intermittent talk followed by lengthy strategic pauses” (Anderson et al. 2010: 1). By employing intricate notation and timing of texts, Anderson et al. find simultaneous talk to occur in 30 percent of the turns in their recorded data, but also find frequent gaps. Earlier psychological studies found less overlap in SSCMC than do Anderson et al., although significantly higher incidence of overlap in SSCMC than in face-to-face conversations. Employing slightly different measurements, Van der Wege & Clark (1997) report approximately 3% overlapping words in SSCMC vs. 2% in face-to-face conversations (at $p < .001$) and Condon & Čech (2001) report 22% overlapping utterances in SSCMC vs. 7% in face-to-face conversations (the latter citing Babineaux forthcoming for the SSCMC figure). McGrath (1990) simply posits that “simultaneous input in a true chat mode,” (cf. SSCMC), “by-passes the turn taking idea [...] by violating the natural communication pattern of one and only one speaker at a time” (1990: 51).

The present study is concerned with finding out whether the SSCMC of split-window ICQ chat affords users greater face-to-face-likeness (orality) than does Internet relay chat, or whether the conversational discourse in SSCMC surpasses face-to-face conversations on any dimension, but approaches the issue from the lexico-grammatical, i.e. text-linguistic, point of view, rather than from the interaction management point of view, even though Anderson et al.’s (2010) and the other studies mentioned, of course, may well inform discussions along the way. Needless to say, it is now high time for a presentation of the media for conversational writing.

2.6 Description of the media for conversational writing

Recall from section 1.2, especially figure 1.2, that the categories speech, writing, APMC, SCMC and SSCMC have the working label “media” in the present study, suggesting that SCMC is one medium and that SSCMC is another medium. Common for all modes of SCMC (listed in figure 1.1) is that the communication is carried out turn by turn, with no overlap, whereas in all modes of SSCMC interlocutors’ turns may be realized simultaneously, with up to complete overlap. The present study investigates one mode of communication to represent SCMC, viz. Internet relay chat, and one mode to represent SSCMC, viz. split-window ICQ chat, seeing that these two modes may be considered prototypical of their respective media, just as, for instance, face-to-face conversations may be regarded as prototypical of speech, and as, for instance, academic prose has been suggested to be stereotypical writing (cf. Biber 1988: 161–162). Genres of SCMC

and SSCMC are likely to display varying degrees of prototypicality, or rather, different positions along Biber's six dimensions of textual variation, just like the genres of speech and writing, but for the working purposes of the present study it is meaningful to regard only these two modes. IRC and split-window ICQ chat are, after all, the first conversational writing genres to be positioned on Biber's (1988) dimensions.³⁵

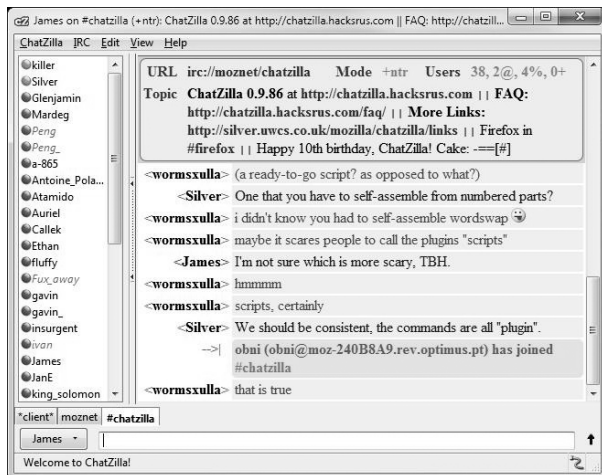
SCMC is carried out in a variety of software and protocols. In mainstream IM (e.g. Facebook chat, Skype chat, MSN Messenger, AIM and current versions of ICQ) users compose their personal "buddy list" of people with whom they are potentially interested in communicating; on Facebook, the chat list of "friends" is automatically generated. Either way, the list indicates the online status of friends. Communication with online friends then occurs as SCMC, turn by turn, while messages to offline friends are delivered as ACMC upon the recipient's re-entry. In the other modes of SCMC, by contrast, no "buddy list" is composed; rather, the communication in these modes takes place with whoever is available on the site (in web chat), in the public chat room/channel (in IRC) or in the virtual world (e.g. Second Life, MMORPG or MUD/MOO) and ACMC is generally not possible. Several IM programs also offer chat with random participants or in public chat rooms/channels. Conversely, in IRC, private chat can take place in a window separate from the channel, either via a special command or by opening a person-to-person connection, "client-to-client" (Pioch 1997, Mar 2000, Herring 2002). The Internet relay chats recorded for the present study, however, exclusively derive from public chat channels with numerous participants.

To connect to IRC, a person uses a chat client, a piece of software, much like connecting to the web necessitates the use of a web browser. Chat clients come in a variety of commercial and non-commercial versions, all with the same basic functions; the user logs on to a server, opts for a nickname and selects a channel, upon which the client displays a list of logged on participants and the chatting begins. Figure 2.3 illustrates SCMC carried out in the IRC channel #chatzilla, with the list of participants' nicknames displayed in the left column, a typical chat client layout. In public channels, messages are displayed to everyone in the channel, in the server's temporal order of receipt, with the producer's nickname

35 Chatted texts from virtual environments lexico-grammatically may constitute one group of SCMC, IM another, and web chat/IRC a third group. Future research will help to define the various SCMC modes; in the present study the modes are kept separate mainly for descriptive clarity. Their diversity apart, all SCMC modes share one and the same kind of turn-by-turn transmission, a characteristic decisively distinguishing them from SSCMC, which is transmitted keystroke by keystroke.

automatically appended before the message. Messages, or rather, turns, are typed in the bottom field and transmitted in their entirety when the user hits the enter key. This means of transmission, *hitting the enter key*, is what distinguishes all modes of SCMC from SSCMC, for in SSCMC, by contrast, users need not hit enter to transmit their turn.

Figure 2.3: Screenshot of Internet relay chat window (SCMC).



In SSCMC, that is, in split-window chat, such as split-window ICQ, Unix Talk or VAX Phone, chatters' communication is transmitted *keystroke by keystroke*, with backspacing, deletion and redrafting immediately visible on both, or all three, participants' screens. Chat clients for SSCMC, unlike those for SCMC, do not come in a great variety, but are limited to the versions released by the communication modes' commercial originators, the ICQ, Unix and VAX companies.³⁶ In ICQ, just as in other mainstream IM programs, users designate their own "buddy list," which indicates friends' online status. In contemporary ICQ chat, chatting with an online friend means synchronous communication, SCMC, whereas in the first decade of ICQ's existence, it meant supersynchronous communication, SSCMC. Messages to offline friends, in either version of ICQ, are also

36 It is possible, or even likely, that split-window ICQ chat, Unix Talk, VAX Phone and similar SSCMC systems represent one and the same textual genre of CMC. For the clarity of discussions, however, split-window ICQ chat is kept apart from the other supersynchronous systems in the present study.

possible (ACMC). In Unix Talk and VAX Phone, on the other hand, no “buddy list” is composed; rather, communication can take place only between individuals logged on to the same server or similar operating system, and the communication is solely supersynchronous. As mentioned, Talk, Phone and similar programs run on Unix, VAX, or Unix- or VAX-like operating systems, generally only at computer professionals’ command. ICQ, by contrast, runs on operating systems in widespread public use, even during ICQ’s decade of enabling SSCMC. Figure 2.4 demonstrates typical split-window interaction, part of UCOW’s split-window ICQ chat text 4.

Figure 2.4: Screenshot of split-window ICQ chat (SSCMC).

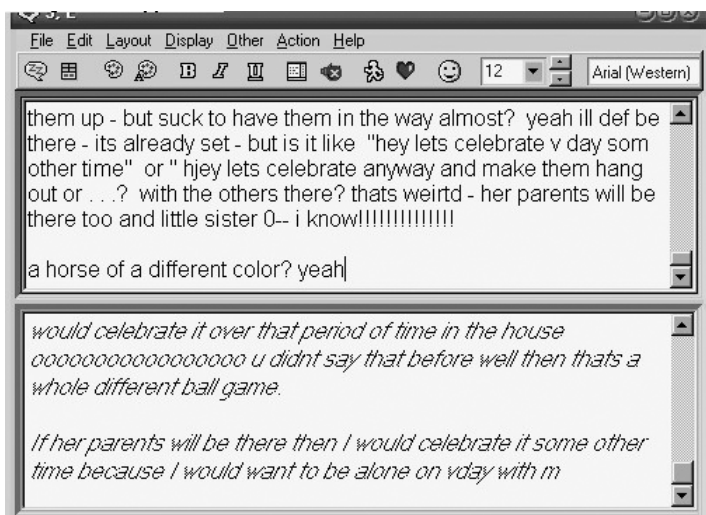


Figure 2.4 shows how the chat window is horizontally split into two parts, one for each interlocutor. A video clip of the same passage shows extended overlapping turns, several instances of hesitation, false starts, self-correction, backspacing and redrafting, that is, it shows language under production in a way similar to how the same features would be rendered evident in the acoustic medium of speech. To preserve the communication, the software is equipped with a logging device. The textual logs fail to capture the redrafting of messages; instead, turns are recorded upon their completion, when the “speaker” pauses. The logs nevertheless provide ample material for lexico-grammatical analysis, as will be seen in the ensuing chapters. The textual logs of the Internet relay chat sessions and the split-window ICQ chat sessions make up the corpus material to be analyzed in the present study. The

next chapter describes how the UCOW corpus was created, that is, how the chats were recorded, purged, annotated and adapted to enable the application of Biber's (1988) methodology, i.e. the frequency calculations, normalization and eventual computation of standardized scores and dimension scores. Before moving on to the "Material and method" chapter, however, a brief summary of the present chapter is in order.

2.7 Chapter summary

The purpose of this background chapter has been to answer a number of questions readers may have on the threshold of a text-linguistic study of CMC setting out to problematize the concepts of speech and writing. Four major questions were addressed. Firstly, what differences between speech and writing have been found in previous linguistic studies? The chapter began by surveying some influential linguistic studies of speech and writing from the early 20th century onwards, exemplifying empirical syntactic and lexico-grammatical findings that in early studies were ascribed to either speech or writing and in more recent work are seen to distinguish among textual genres. The second question addressed was how genres/registers of speech and writing can be assessed quantitatively and qualitatively. Two complementary approaches were introduced and outlined as methods suitable for the present study: the quantitative/qualitative study of dimensions of textual variation, employing Biber's (1988) methodology, and the essentially qualitative semiotic, or systemic-functional, approach to register variation devised by e.g. Halliday (1978, 2004), Halliday & Hasan (1989), Martin (1992, 2001a, 2001b), involving the identification of the field, tenor and mode of a communicative situation for the adequate description of registers. The chapter then moved on to consider the third question, that of how CMC has been approached linguistically before, by surveying the literature on CMC, correlating the emergence of modes with relevant linguistic studies, but also tracing antecedents of current modes worthy of study. Several studies of SCMC and SSCMC with a bearing on the present investigation were elaborated upon, evidencing, among other things, the scarcity of text-linguistic SSCMC studies. The final question addressed was how SCMC and SSCMC, as instantiated in Internet relay chat and split-window ICQ chat, are carried out. Typical interfaces for each medium were presented to illustrate the basic difference between SCMC and SSCMC: the turn-by-turn vs. keystroke-by-keystroke means of transmission. UCOW consists of one synchronous and one supersynchronous component, both instances of conversational writing. The present study intends to relate the components to each other, and to speech and writing, in the endeavor to provide a detailed

description of conversational writing. Chapter 3 presents the compilation, annotation and adaptation of the conversational writing corpus, and an SBC subset, and the computations involved for obtaining average figures for comparisons across the media, as well as for positioning the genres on Biber's dimensions of textual variation; in other words, chapter 3 embarks on the empirical investigation of conversational writing.

Chapter 3. Material and method

3.1 Introductory remarks

In the present study, two genres of conversational writing are contrasted with spoken and written genres to investigate the relationships between conversational writing, speech and writing. A corpus of conversational writing was collected for the purpose, and a corpus of face-to-face conversations was sampled to create a complementary corpus of spoken conversations. As the study uses Biber's (1988) methodology, the chatted texts will also be contrasted with the two corpora studied by Biber, i.e. LLC for speech and LOB for writing (see Appendix I). The corpus of conversational writing consists of two components: one of SCMC, viz. Internet relay chat, and one of SSCMC, viz. split-window ICQ chat, both annotated for the purposes of the present study. The complementary, spoken corpus consists of a subset of face-to-face conversations sampled from the Santa Barbara Corpus of Spoken American English, part 1 (Du Bois et al. 2000). The sub-corpus, henceforth called the "SBC subset," was essentially brought into in the study to supplement Biber's (1988) genres from LLC with more recent spoken material. Table 3.1 presents briefly, in numbers, the corpora compiled/sampled and annotated, corpora on which most of this chapter focuses (for Biber's 1988 corpora, see Appendix I). (The corpus of ACMC, with which findings will also be contrasted, is introduced in section 4.1.)

Table 3.1: *Size of corpora compiled/sampled and annotated for the present study*

	Corpus	Number of texts	Min. length	Max. length	Average length	Corpus size
Conversational writing (UCOW)	Internet relay chat	10	961	999	984	9,841 words
	Split-window ICQ chat	12	459	1150	772	9,261 words
Face-to-face conversations	SBC subset	14	673	720	712	9,962 words

The next two sections of this chapter, 3.2 and 3.3, describe the collection, adaptation and annotation of the conversational writing corpus, UCOW (Uppsala Conversational Writing Corpus). Each of the two UCOW components, Internet relay chat and split-window ICQ chat, constitutes a corpus in itself, meaning that it may be referred to both as a corpus and as a UCOW component throughout this study. The ensuing section, 3.4, describes the sampling and annotation of the

SBC subset. The chapter proceeds, in section 3.5, to explain how Biber's (1988) MD methodology was applied to the data, focusing on the frequency standardizations and dimension score calculations. Section 3.6 then turns to the two corpora studied in Biber (1988), summarized in Appendix I. The section describes how average figures for writing and speech were obtained from these, for speech in combination with the SBC subset data, and how the results will be presented and analyzed in this study. The final section, 3.7, sums up the chapter.

3.2 Creating and annotating a corpus of Internet relay chat

The corpus of Internet relay chat was recorded in 2002 from five public Internet relay chat channels: #20_something, #30_something, #Chat_world, #Family and #USA. Each channel was logged for several hours, yielding log files of several hundred thousand words in total for the five channels. From these log files, two sample texts were drawn from each chat channel (texts 1a and 1b from #20_something, texts 2a and 2b from #30_something, etc.). Each sample text consisted of thousands of words but was subject to a purging procedure aimed at sifting out only the linguistic messages explicitly keyed in by the human conversationalists. The raw samples were all long enough to ensure that the resulting texts, containing the full log of user-generated messages, would comprise approximately one thousand words each; cf. table 3.1. Exemplified in (1) are the first few minutes of an unpurged sample from the channel #Family, which ultimately contributed the first twelve lines to Internet relay chat text 4a in UCOW. Sample (1) includes a session start message, time stamps, bracketed nickname turn indicators and server-generated messages such as join- and quit-information (indicated by ***), of which all except the bracketed nickname turn indicators were purged to sift out the twelve user-generated turns. The twelve turns of current interest are then re-presented as (2).³⁷

- (1) Session Start: Mon Mar 25 18:01:47 2002
[18:01] *** Now talking in #family
[18:01] <River> woohoo,
[18:01] <Genie500> Laughing Out Loud
[18:01] <River> my hair is almost as long as yours
[18:02] <Genie500> now ya know who to look for honking across the street
[18:02] <River> yep
[18:02] <Genie500> really?? lol

37 Examples of material excluded from (2) and other passages of text, e.g. channel operator interference, action commands and foreign language turns, are found in Appendix IV.

[18:02] *** edi-tr has joined #family
 [18:02] <River> well just in the back
 [18:02] <Genie500> Laughing Out Loud
 [18:02] *** edi-tr has quit IRC (Killed (NickServ (Nickname Enforcement)))
 [18:03] <Genie500> and what color is yours??
 [18:03] *** EmeL_ has joined #family
 [18:03] *** Guest_984 has joined #Family
 [18:03] *** edi- has joined #family
 [18:03] <lookingforagirl> blue
 [18:03] *** blue-ice has joined #Family
 [18:04] <Genie500> oh river just a sec I gotta turn something off for you to
 send okay
 [18:04] <River> this one is from 95 without the glasses .

- (2) <River> woohoo,
 <Genie500> Laughing Out Loud
 <River> my hair is almost as long as yours
 <Genie500> now ya know who to look for honking across the street
 <River> yep
 <Genie500> really?? lol
 <River> well just in the back
 <Genie500> Laughing Out Loud
 <Genie500> and what color is yours??
 <lookingforagirl> blue
 <Genie500> oh river just a sec I gotta turn something off for you to
 send okay
 <River> this one is from 95 without the glasses .

Internet relay chat text 4a (UCOW)

Example (2) represents the default format in which examples from the Internet relay chat texts are presented in the ensuing chapters of this study.³⁸ It retains the bracketed nickname turn indicators to mark turn boundaries, although, needless

38 Nicknames are pseudonyms in themselves, and not real-life identities, and have traditionally been retained in published research on texts deriving from public CMC domains, such as public IRC channels, with non-sensitive content (cf. Werry 1996, Danet et al. 1998, Schulze 1999, Ooi 2002, Waldner 2009). As their real-life identities are disguised by such pseudonyms, the IRC chatters were not asked for their informed consent to be recorded in the present study. This is in line with e.g. Rafaeli et al. (1998), Liu (1999), Cameron (2001) and Waldner (2009), the second of whom notes that public IRC conversations are “acts deliberately intended for public consumption” (Liu 1999: no page number available) and may be regarded as exempt from practices of obtaining informed consent. Moreover, it was felt that intruding in conversations to obtain chatters’ consent was not only impracticable, but also would have disrupted the natural

to say, the indicators were not subject to linguistic annotation or feature counts. Rather, the texts for annotation contain only the linguistic messages explicitly keyed in by participants; see example (3). It is in a state such as (3) that the corpus is reflected in table 3.1; the Internet relay chat component of UCOW contains a total of 9,841 words of such linguistic messages exchanged between interlocutors, drawn from 10 texts averaging 984 words each.

- (3) woohoo,
 Laughing Out Loud
 my hair is almost as long as yours
 now ya know who to look for honking across the street
 yep
 really?? lol
 well just in the back
 Laughing Out Loud
 and what color is yours??
 blue
 oh river just a sec I gotta turn something off for you to send okay
 this one is from 95 without the glasses.

Internet relay chat text 4a (UCOW)

Once the texts had been distilled to the format exemplified in (3), the linguistic annotation began. The purpose of the annotation procedure was to mark up all occurrences of Biber's (1988) 65 linguistic features likely to distinguish among spoken and written texts; see table 2.1 (the features type/token ratio and word length do not require markup). Attempts were made to run a few texts through automatic part-of-speech tagging software, e.g. the CLAWS tagger, for an annotational starting point. Because of the irregular orthography of the chatted texts, however, the automatic taggers repeatedly failed or achieved insufficient accuracy, which made manual annotation the only remaining option.³⁹

The manual annotation proceeded in a series of time-consuming steps, each involving meticulous consideration.⁴⁰ First, the texts were annotated for parts

flow of conversations to the point of destroying the data (cf. Sveningsson 2001, Waldner 2009).

39 Developing software for the purpose of tagging the UCOW texts was beyond the time allotted for the study. Manual annotation was instead carried out straight onto the texts in a word processor, which eventually conveniently enabled feature counts by way of the program's "find all" option.

40 The demanding task of manual annotation explains the relatively limited size of the corpora compiled and sampled for the present study (see table 3.1). However, although 10,000-word corpora may appear small, few corpus linguistic studies have been based

of speech roughly in accordance with the tagset devised in the Penn Treebank project (Santorini 1990, Marcus et al. 1993). The part-of-speech tagging provided useful basic classification, as some of the Penn Treebank tags directly correspond to features in Biber (1988), e.g. VBD for past tense verbs and NN for nouns. In the annotation process, however, a great number of tags had to be modified to denote Biber features, and new tags invented ad hoc, to eventually cover all of the 65 features. Example (4) shows the resulting annotation of the example considered here, in which a great number of custom-made tags mark up the linguistic items.

- (4) woohoo,/IJ
 Laughing/VX Out/ADV Loud/ADV
 my/PRP1s hair/NN is/VBP/-BE almost/46/47 as long/-PADJ/+ADJ as
 yours/+PRP2s
 now/TADV ya/PRP2s know/VBP/PRV who/23 to/TO look/VBI for/PP
 honking/-28 across/PADV the street/NN
 yep/IJ
 really?/?/IJ/ADV lol/EM
 well/IJ/ADV/DP just/49 in/PP the back/NN
 Laughing/VX Out/ADV Loud/ADV
 and/CONJ/65 what/?WHQ color/NN is/VBP/BE yours?/?+PRP2s
 blue/-PADJ/+ADJ
 oh/IJ river/NPna just/49 a sec/NN I/PRP1s gotta/+NM/?CT turn/VBPi
 something/PN off/ADV for/PP you/PRP2s to/TO send/VBI okay/-PADJ/+ADJ
 this/DET one/NN is/VBP/BE from/PP 95/CD without/PP the glasses/NN .
 Internet relay chat text 4a (UCOW)

Table 3.2 provides a key for interpreting the tags in (4). The annotation was tailored to comply closely with the algorithms provided in Biber (1988: 211–245) for the detection of linguistic features. Matching the algorithms in the annotation was a *sine qua non* for ensuring the material maximum comparability with the spoken and written texts in Biber’s study. As an example, verbs were tagged as infinitives only when following infinitive marker *to* and an optional adverbial element (that is, whenever identified by the algorithm “to+(adv)+vb”), as *look* and *send* in example (4), both tagged /VBI (Biber’s feature no. 24). Occasionally,

on such large manually, and single-handedly, lexico-grammatically annotated datasets. Three of Biber’s (1988) (automatically annotated) genres are of a similar size, viz. science fiction, personal letters and professional letters (see Appendix I), which is seen as unproblematic as Biber (1990) convincingly demonstrates and establishes that ten 1,000-word text samples (i.e. a total of 10,000 words) suffice for the adequate lexico-grammatical representation of a genre.

the close compliance with Biber's algorithms meant that an item's affiliation with a particular linguistic category had to be ignored. Consider, for instance, the occurrences of *be* as main verb in example (4), one tagged /-BE and two tagged /BE. Biber's algorithm for *be* as main verb (1988: 229) detects only the instances in which the verb is followed by a determiner, possessive pronoun, address title, preposition or adjective. The algorithm thus would exclude the first instance of *be* as main verb in (4), as it is followed by a downtoner adverb, but properly detect the other two instances, which are followed by a possessive pronoun and a preposition, respectively. Accordingly, in the manual annotation, the first item had to be marked as deviant, /-BE, and only items tagged /BE eventually counted as instances of Biber's feature no. 19, as inferable from table 3.2.

Table 3.2: *Tags used in the annotation of the first twelve turns in Internet relay chat text 4a (UCOW)*

Tag	Description	Feature no.	Explanation
+ADJ	adjective not identified as predicative	40	
ADV	adverb	42	
BE	BE as main verb	19	
-BE	not identified as BE as main verb		see Biber's algorithm for 19
CD	cardinal number		not a Biber feature
CONJ	conjunction		not a Biber feature
?CT	not identified as contraction		see Biber's algorithm for 59
DET	demonstrative	51	
DP	discourse particle	50	
EM	emotive		genre-specific feature
IJ	interjection and/or insert		genre-specific feature
+NM	not identified as necessity modal		see Biber's algorithm for 53
NN	noun	16	
NPna	proper noun, nickname address term		genre-specific feature
-PADJ	not identified as predicative adjective		see Biber's algorithm for 41
PADV	place adverbial	4	

Tag	Description	Feature no.	Explanation
PN	indefinite pronoun	11	
PP	preposition	39	
PRP1s	first person singular pronoun	6	
PRP2s	second person singular pronoun	7	
+PRP2s	not identified as second person singular pronoun		see Biber's algorithm for 7
PRV	private verb	56	
TADV	time adverbial	5	
TO	infinitive marker		not a Biber feature
VBI	infinitive	24	
VBP	present tense verb	3	
VBPi	present tense verb, base form	3	
VX	progressive verb		not a Biber feature
?WHQ	not identified as WH-question		see Biber's algorithm for 13
23	WH clause	23	
-28	not identified as present part. WHIZ deletion		see Biber's algorithm for 28
46	downtoner	46	
47	hedge	47	
49	emphatic	49	
65	non-phrasal coordination	65	

Table 3.2 indicates which items in example (4) count as instances of Biber's (1988) features (by feature numbers in the third column in accordance with table 2.1 of the present study) and explains why others do not. Items not counted as instances of Biber features diverged for one of two reasons: either they did not conform to detection by Biber's algorithms (like/-BE) or they simply constitute a feature not studied by Biber (e.g. /CD, cardinal numbers); see the explanations column in table 3.2. Moreover, three tags were developed in the present study to denote instances of genre-specific features, none of which was studied in Biber (1988): the tag for nicknames used as address terms, /NPna, typical of IRC; the tag for inserts, /IJ, for e.g. interjections, frequent in conversational genres; and the tag for emotives, /EM, marking up emoticons ("smileys") and sentiment initialisms, typical of chatted texts. Nicknames used as address terms (e.g. *river* in

example 4) are not regarded as nouns in this study (unlike other proper nouns), to avoid skewing the noun count of the IRC texts, but their frequency of occurrence nevertheless affects the relative incidence of other features in IRC, as will be seen in section 5.2.1. Inserts and emotives will be expounded upon in section 4.6, in which the annotation of inserts is explained further.

Even though table 3.2 lists only the tags used for the example text given here, it is indicative of the procedure for detecting Biber items, the items to be included in the linguistic feature counts. Among the more than one hundred tags used for the annotation, several were interim, and a slightly more limited set was eventually subjected to feature counts – the tags that mark up Biber features. The final step of the annotation process involved summing up the latter items. Example (5) illustrates the incidence of Biber features, as such items are here numbered in accordance with the third column of table 3.2 (based on the feature list in table 2.1).

- (5) woohoo,/IJ
 Laughing Out/42 Loud/42
 my/6 hair/16 is/3 almost/46/47 as long/40 as yours
 now/5 ya/7 know/3/56 who/23 to look/24 for/39 honking across/4 the street/16
 yep/IJ
 really?/?/IJ/42 lol/EM
 well/IJ/42/50 just/49 in/39 the back/16
 Laughing Out/42 Loud/42
 and/65 what color/16 is/3/19 yours??
 blue/40
 oh/IJ river/NPna just/49 a sec/16 I/6 gotta turn/3 something/11 off/42 for/39
 you/7 to send/24 okay/40
 this/51 one/16 is/3/19 from/39 95 without/39 the glasses/16.

Internet relay chat text 4a (UCOW)

The sample in (5) thus contains, for instance, seven adverbs (feature no. 42), two first person pronouns (feature no. 6), seven nouns (feature no. 16), etc. As mentioned above, the genre-specific tags /NPna, /IJ and /EM are of further interest (the first in section 5.2.1, and the latter two in section 4.6), but are not numbered in (5), as they are not among Biber's (1988) features.

As seen in table 3.1, the UCOW Internet relay chat texts contain 984 words on average. To make feature counts comparable across texts and genres, Biber's (1988) methodology prescribes the normalization of counts to occurrences per 1,000 words. Internet relay chat text 4a contains 975 words, which means that the occurrences had to be multiplied by 1,000/975 to attain normalized frequencies. The raw frequency for adverbs in text 4a, 82, for instance, was normalized to 84.1 (as $82 \times 1,000/975$ is 84.1). The resulting normalized frequencies for all

the features in every Internet relay chat text are found in Appendix II table 5, and the normalized frequencies for the whole Internet relay chat corpus are found in Appendix II table 1 (based on the sum of average frequencies in the individual texts, divided by the total number of texts). The tables in Appendix II thus sum up results of fundamental importance for the Internet relay chat corpus, as well as those for the other corpora studied, results upon which this study is based.⁴¹

Two of Biber's (1988) features, type-token ratio (TTR) and word length, are used to measure the lexical diversity and specificity of texts. These two features did not require annotation, but necessitated some other processing of the texts. For the purpose of TTR and word length calculations, the texts in the present study were purged of all regular punctuation except apostrophes within words, emoticons and simple imagery,⁴² rendering texts the appearance exemplified in (6). In compliance with Biber (1988: 238), TTR was calculated "by counting the number of different lexical items that occur in the first 400 words of each text, and then dividing by four." Three pieces of lexical analysis software were used for computing the TTR, viz. KWIC 5.0, AntConc 3.2.4 and WordSmith Tools 5.0.0.334, which all yielded congruent results. For the average word length count, only WordSmith Tools was used; the full texts were used as input and no normalization was needed. Section 4.3 explains the procedures further and discusses the results. The figures for TTR and word length in the texts studied are given among the features in the tables of Appendix II (features 43 and 44), as numbers equally central to the present study as those of the annotated features.

- (6) woohoo
Laughing Out Loud
my hair is almost as long as yours
now ya know who to look for honking across the street
yep
really lol
well just in the back
Laughing Out Loud
and what color is yours
blue
oh river just a sec I gotta turn something off for you to send okay
this one is from 95 without the glasses

Internet relay chat text 4a (UCOW)

41 The raw frequencies are provided in Appendix III, but no reference is made to them in the study.

42 "Simple imagery" is explained further in section 4.3.

Besides TTR and word length, section 4.3 of the present study also considers the lexical density of texts. Lexical density was introduced in section 2.4 as Halliday's (1985a) only quantitative means for distinguishing between spoken and written texts, indicating a low lexical density for spoken and a high lexical density for written texts. The lexical density calculation in the present study was carried out on texts in a state such as in example (3) above, i.e. with the punctuation retained. Single stranded punctuation marks were not counted, but emoticons (e.g. :, ;) were counted as words (non-lexical). Lexical density measures the ratio of lexical words (i.e. content words) to the total number of words. The measurement is insensitive to text length, which means that it was computed on the full corpus, and was not normalized. The full corpus size, however, was increased slightly before the lexical density calculation, to compensate for abbreviations and contractions, the former typical of conversational writing. Abbreviations such as *idk* meaning "I don't know" and *nm* meaning "not much" were thus tokenized, that is, considered as if their constituents were spelled out, i.e. *idk* as four words (*I, do, n't, know*), *nm* as two, except for sentiment initialisms, e.g. *lol*, *rofl* and *lmao*, which were considered as uniform words.⁴³ The lexical density count further considered accidentally conjoined words (such as *guesysea*) as separate tokens, and accidentally separated words (such as *out side*) as single tokens. To account for these irregularities, a total of 165 tokens were added to the Internet relay chat corpus size, most of the tokens deriving from the tokenization of abbreviations and contractions. Section 4.3 details which of all the tokens were then taken to be lexical words. As mentioned, lexical density is a measurement of the ratio of lexical words to the total number of words, the total number of words in the Internet relay corpus being 9,841+165, i.e. 10,006, for the lexical density calculation only. Section 4.3 further explains how the measurement of lexical density was applied to the material and discusses the findings.

43 Determinant for the treatment of chat abbreviations was their propositional content. Whereas the tokenized abbreviations typically convey propositional content, the sentiment initialisms, just like emoticons, typically are non-propositional and rather may be regarded as "textual indicators of illocutionary force" (cf. Dresner & Herring 2010: 260). Moreover, the first two sentiment initialisms, *lol* and *rofl*, in effect have become lexicalized in the English language; see further section 4.6. *Lol* means "laughing out loud," *rofl* "rolling on the floor laughing" and *lmao* "laughing my ass off" (Crystal 2004a). Sentiments spelled out in the original text, e.g. *Laughing out loud* in examples (2)–(5), of course, remained several tokens in the lexical density calculation.

3.3 Creating and annotating a corpus of split-window ICQ chat

Collecting a corpus of split-window ICQ chat between individuals in private conversations demanded a greater effort on the part of the present researcher than did the recording of the Internet relay chat discourse, which is readily available in public chat channels (cf. section 3.2). The split-window ICQ chat component of UCOW was collected in 2004 by logging conversations between high school seniors in the northeastern USA. The present author involved two high schools in the project, which yielded 23 informants' conversations, and two high school students were recorded in their home. All in all, 12 texts of split-window ICQ chat were compiled, as indicated in table 3.1. Out of these texts, eleven were conversations between dyads and one was a conversation among three people. Eighteen students were male and seven were female. Most conversations took place between males or in mixed-sex dyads; only one conversation involved two females.

About a week before the recording, the subjects were informed about the study, both orally and in writing, and asked to bring home an informed consent form for their guardians' review and signature, if the student was underage (below the age of 18). All students interested in participating brought home the form and brought it back signed, regardless of their age. For the recording in the home setting, informed consent was obtained orally from the subjects' parents. The recording event took place during one lesson in each high school, and for the equivalent period of time in the home setting. A computer classroom was set up for the purpose in the high schools, and a home office for the latter event. The students formed dyads, and a triad, on their own before entering the classroom and were assigned computers as distant from their conversational partner as possible upon entering the classroom. Computers had been pre-set with the required software, the ICQ chat Pro 2003b program as well as the HyperCam screen capturing software, the latter intended to capture the split-window ICQ chat action as a video file. After all students had been seated, they were introduced to the software and allowed ten minutes to practice. As they were all apt and avid online chatters from before (most with more than ten hours of chatting experience, typically in chat rooms or on AIM, although none had used ICQ), they immediately caught on and managed the ICQ program. The following four instructions were given in a sheet taped to the physical desktop beside each computer:

- Do not move, close, cover up or disrupt the chat window while recording
- Immediately close any popup-windows
- Do not follow links or advertisements
- Would you like to save this Chat session? Click OK and save to desktop

Because of the limited time allowed in the classroom, the time for recording conversations was restricted to about 20 minutes. Students were informed that the content of their discourse would not be subject to assessment and were explicitly encouraged to converse freely on any topic of their choice. Nevertheless, before the students were instructed to initiate the recording, they were given a few topics to resort to in the event that their conversations might run dry. The suggested topics were written on the board: “Plans for the weekend,” “Plans for the summer,” “Plans for next year”; in the home setting the same topics were suggested orally.⁴⁴ As it turned out, only a few utterances in four of the texts may have sprung from these suggestions. As will be seen in textual examples throughout this study, most split-window conversations revolve around other topics and appear remarkably uninhibited and diverse, considering that the discourse was recorded in a situation of elicitation. Example (7), the first 15 turns of split-window ICQ chat text 8, indicates typical split-window ICQ discourse produced, with very few traces of the situation of elicitation (such traces are discernible in lines 6–10, i.e. only in four out of 71 lines in the full text). In the recordings, the interlocutors typically carried on eagerly with the conversation initiated while practicing and were all noticeably unperturbed by the experimental setting, as evident in lines 11 ff. of example (7). Upon finishing their recording, the students saved their conversations both as a screen-captured video clip and as a textual log file. Before leaving the room, they were further instructed to retrieve a minor remuneration for their participation from underneath the taped instruction sheet, which they did gladly (as no remuneration had been mentioned before).

- (7) <9> YES
 <9> !!!!!!!
 <I>
 <I> hey baby
 <9> we suck at this
 <I> well there ya go... uits time to record our 20 minutes sessions
 <9> did u press record
 <9> yep
 <I> yeah did u?
 <I> ok good
 <I> so question...
 <9> who said i hooked up with her

44 Future plans was one of several productive topics given to informants in Renouf's (1986) elicitation of spoken English.

- <I> if u dont wanna be with laurie anymore, why did u just hook up with her on saturday???
- <9> we were both lying there and i kissed her but i wouldnt say we hooked up
- <I> i asked her yesterday when th elast time u hooked up and she told me satruday. but dont tell her that im telling u this.

Split-window ICQ chat text 8 (UCOW)

The corpus of split-window ICQ chat was then collected from the classroom computers. Due to the varying quality of the video clips, it was decided that the textual log files would constitute the material for lexico-grammatical analysis (the corpus is made up of the entire collection of logs, i.e. the chats were not sampled). Unlike the IRC logs, the raw textual log files of the split-window chats are readily legible; see example (7), in effect derived straight from such a file.

As seen in (7), participants' turns are preceded by bracketed turn indicators. Just as in IRC, these contain a nickname, even though, for practical reasons, the nicknames in the split-window ICQ recordings were pre-set on computers and not invented by the participants. Unlike in IRC, the split-window chatters were able to personalize their messages with variable fonts, font size and font color, which they did, even though none of this is reproduced in the text samples in this study. This variability in ICQ, nevertheless, is seen as one of the paralinguistic devices available to chatters and is further discussed in section 4.5. The ICQ program also offers a set of graphic emoticons and pre-programmed textual action tropes. Participants did not use the graphic emoticons, but experimented somewhat with the action tropes. A trope is realized as a line of text, which is not explicitly keyed in by the participants, but assigns an action to his/her nickname (e.g. *9 picks a flower and hands it to you*), imitating an action command in IRC. Just as for the IRC material, however, the action lines of the split-window ICQ chat component were removed before the annotation began, along with the bracketed nickname turn indicators and a few foreign language turns (see Appendix IV). No other purging was needed to adapt the ICQ material for the lexico-grammatical annotation and analysis. The average text of the analyzed split-window ICQ chat component is 772 words long, as indicated in table 3.1, and the whole corpus contains 9,261 words.

The annotation of the split-window ICQ texts followed the same procedure as did the annotation of the IRC texts (cf. section 3.2), i.e. the same tags were applied, and Biber's features were eventually identified and counted and the frequencies normalized. The normalized frequencies of the 65 features, as well as the TTR and word length of the individual split-window ICQ chat texts, are found in Appendix II table 6, and the figures for the whole split-window ICQ corpus are found in Appendix II table 2. For the lexical density calculation, as described for

IRC in section 3.2, the size of the split-window ICQ corpus was increased slightly to compensate for items contained in abbreviations (e.g. for *ic*, meaning “I see”) and contractions, as well for accidentally conjoined words (e.g. *lastnight*) minus accidentally split-up words (e.g. *any way*). A total of 51 tokens were added to the corpus size, thus making the denominator for the split-window ICQ corpus in the lexical density computation 9,312 words. The lexical density of split-window ICQ chat, just as that for IRC, will be further discussed section 4.3.

Examples from the split-window ICQ chats will be given in a format such as that in (7) throughout this study, i.e. with bracketed turn indicators retained. Informants’ textual references to personal names, locations, etc., have been carefully masked in order to preserve informants’ anonymity; accordingly, *laurie* in (7) is fictitious. The annotation and the TTR, word length and lexical density calculations, however, were carried out on the original texts with original verbatim references.

3.4 The Santa Barbara Corpus subset

The present section briefly introduces LLC’s conversational genres and compares them to the UCOW genres, in order to explain the motives for studying SBC as a supplementary corpus to LLC. The section further describes the sampling of SBC and the annotation of the SBC subset texts, concluding with a remark on how the SBC subset results are treated in the study.

In section 2.3 of the present study, Biber’s (1988) MD methodology for studying textual variation was introduced. As mentioned there, Biber (1988) discovered six dimensions of variation among written and spoken texts and positioned six genres of speech and 17 genres of writing on each of them. The present study intends to position the two UCOW genres of conversational writing on the same dimensions. By using the established positions of Biber’s genres on the dimensions, especially those of oral conversations (face-to-face and telephone conversations), it should be possible to determine the level of orality in conversational writing, i.e. its similarity to oral conversations. The positions of the conversational writing genres will also help to address, for instance, the two hypotheses stated in section 1.2, suggesting different levels of orality in SCMC and SSCMC. Throughout this study, samples from conversational writing will be contrasted with textual samples of spoken conversations, as well as with samples from other genres, to exemplify for instance the distribution of lexico-grammatical features. The spoken genres in Biber’s (1988) study derive from LLC and the written genres from LOB, as well as two collections of letters (see Appendix I for a list of genres studied in Biber 1988). As the conversational genres are of particular interest

in the present study, the comparability of UCOW to LLC conversations is an important concern.

The conversations in LLC are face-to-face and telephone conversations recorded among speakers of British English in the 1960s and 1970s, most of whom were academics (Greenbaum & Svartvik 1990). UCOW, as described in the two sections above, was recorded among random chatters in various chat channels, and among high school seniors in private conversations, in the 2000s. The ages of the IRC chatters are unknown, and their varieties of English are unpredictable (ranging from the “global” English of EFL speakers, to the subtle regional variants of native speakers). The split-window ICQ chatters, however, are a fairly homogeneous group of adolescent American English speakers (most from middle-class suburban neighborhoods). To improve the quality of comparisons between UCOW and oral conversations it was thus desirable to study a corpus of spoken American English, alongside LLC, preferably one recently collected. The corpus opted for, to fulfill these requirements, was Part 1 of the Santa Barbara Corpus of Spoken American English, here SBC, recorded in the late 1980s to mid-1990s (Du Bois et al. 2000).⁴⁵ In addition to being regionally and temporally closer to UCOW, SBC also represents the spoken conversations of “a wide variety of people of different regional origins, ages [inter alia teenagers], occupations, genders, and ethnic and social backgrounds”⁴⁶ and is thus possibly socially more diversified than LLC.

SBC part 1 was released in 2000 and consists of 14 face-to-face conversations. To limit the burden of annotation in the present study, it was decided that the SBC conversations be sampled to obtain a corpus of a size similar to the UCOW components, i.e. approximately 10,000 words. Consequently, the first 712 words (on average) from each of the 14 conversations were sampled, to form an SBC part 1 subset corpus; see table 3.1. This SBC subset was first purged of its timestamps and stripped of its original prosodic mark-up, before it was annotated with the tags used in the present project; cf. section 3.2.⁴⁷

Compared to the collection, adaptation and annotation of the two UCOW components (sections 3.2 and 3.3), the sampling, adaptation and annotation of the SBC subset was a fairly straightforward task. Unlike the UCOW texts, the

45 The Santa Barbara Corpus of Spoken American English is currently part of the International Corpus of English (ICE) project, directed by Gerald Nelson; see <<http://ice-corpora.net/ice/index.htm>> (accessed 2015-10-13).

46 Cf. the Santa Barbara Corpus web site <<http://www.linguistics.ucsb.edu/research/santa-barbara-corpus>> (accessed 2015-10-13).

47 The SBC texts were obtained prior to their part-of-speech mark-up in the ICE project.

texts of the SBC subset have regular spelling and punctuation and lack abbreviations denoting several words, which make them comparatively easy to annotate. The few foreign language turns found (one of which is exemplified in Appendix IV) were removed from the subset, but no other purging of the raw texts was needed. Example (8), the first ten turns of “face-to-face conversations SBC” text 2, shows how SBC subset texts will be presented in this study. Speaker names indicating turns, e.g. “Jamie” and “Harold” in (8), are carried over from the original corpus, but were naturally not subject to annotation.

- (8) Jamie: How can you teach a three-year-old to tap dance.
Harold: I can't imagine teaching a
Jamie: Yeah,
really.
Miles: Who suggested this to em.
Harold: I have no idea.
It was probably my sister-in-law's idea because,
I think they saw that movie.
Jamie: Tap?
Harold: What was the,
Miles: They had
Harold: the movie with that really hot tap dancer.
Jamie: Oh that kid.

Face-to-face conversations SBC text 2

The annotation and summing-up of Biber's (1988) features in the SBC subset followed the same procedure as for the UCOW texts; see sections 3.2 and 3.3. The resulting normalized frequencies of the features in the individual SBC subset texts, as well as the TTR and word length of texts, are found in Appendix II table 7 and the frequencies, TTR and word length of the whole SBC subset are found in Appendix II table 3. The lexical density of the SBC subset was also computed, for the eventual comparison with the UCOW genres (to be explored in section 4.5). Unlike for the UCOW components, however, the lexical density calculation for the SBC subset required no tokenization of the texts (cf. section 3.2) beyond the tokenization of contractions. The regular orthography of the linguist-transcribed spoken texts means that no words have been accidentally split-up or conjoined, and in the spoken texts no words are “hidden” in abbreviations.

Crucially, the SBC subset corpus provides a supplementary point of reference, besides LLC, for comparisons between conversational writing and face-to-face conversations in the present study. The SBC subset supplements LLC in three valuable ways: it is regionally comparable with the split-window ICQ chats, as both represent American English; the SBC subset is also temporally adjacent to

UCOW, as the two were recorded in successive decades; and, finally, the SBC subset possibly represents the English of a socially more diverse set of speakers, including adolescents, making it slightly better suited than LLC for comparisons with UCOW. This is not to say that LLC is ruled out in the analyses to come. On the contrary, LLC feature counts will be referred to on a regular basis, as they are integral to Biber's (1988) investigation. Textual examples of face-to-face conversations, however, will mostly be drawn from the SBC subset. Moreover, "face-to-face conversations LLC" and "face-to-face conversations SBC" will be positioned as separate genres on Biber's dimensions (in chapter 5).

Spoken English, nonetheless, consists of more than face-to-face conversations. Besides face-to-face-conversations, LLC contains texts from telephone conversations, interviews, broadcasts, spontaneous speeches and prepared speeches; see Appendix I. Of the six genres in LLC, the two conversational genres (face-to-face conversations and telephone conversations) are of primary concern in the forthcoming comparisons with conversational writing. The vast majority of LLC's telephone conversations were recorded in the 1970s (only five are from the 1960s), which may put this genre in less urgent need of supplementing with updated corpus data than the LLC face-to-face genre (which has only slightly more texts from the 1970s than from the 1960s). Even so, a newer telephone conversations corpus admittedly would have been desirable. Annotating such a corpus, however, was beyond the time scope of the study. Consequently, no corpus of telephone conversations, beyond LLC, was sampled or otherwise brought into the present study.

One final remark needs to be made here in connection with LLC and the SBC subset. As mentioned in chapter 1, the present study intends to contrast the distribution of salient features in conversational writing with the distribution of the same features in speech, writing and APMC. When it comes to the medium of "speech," the SBC subset will be merged with the LLC genres to constitute a uniform point of reference. Accordingly, in chapter 4, the medium of speech is represented by LLC's six genres of speech (cf. Appendix I) combined with the SBC subset face-to-face genre. Just how average figures for this combined set of "speech" were obtained will be further described in section 3.6. This chapter now turns to a description of how Biber's (1988) MD methodology was applied to the feature count data from UCOW and the SBC subset to obtain standardized scores and, eventually, dimension scores for the genres under study.

3.5 Standardization and dimension score computation

This section follows up on section 2.3, in which Biber's (1988) methodology for computing dimension scores for the genres of speech and writing was described. As outlined there, a complete MD analysis involves eight methodological steps. Of the eight steps outlined in section 2.3, the present study has, by this point, implemented steps 1–5. Three corpora have been “designed” in the present study, namely the two UCOW components “Internet relay chat” and “split-window ICQ chat,” as well as the “SBC subset” (step 1). The linguistic features chosen for the study are the same as those identified in Biber's (1988) original study, i.e. the 67 features listed in table 2.1 (step 2). The three corpora have been manually tagged for their occurrences of all of these features, and the average TTRs and word lengths of all texts have been computed (steps 3 and 4), and finally, frequency counts have been computed and compiled into tables in Appendix II, along with TTRs and word lengths (step 5). As the present study relies on Biber's pre-defined dimensions, it leaves out steps 6 and 7 of a complete MD analysis. This means that only step 8, the final step, remains, in which dimension scores are computed for the texts/genres on each dimension. As mentioned in section 2.3, however, the present investigation lingers in step 5 for a while, devoting considerable space to the discussion of salient results obtained. The present section explains what this means, before homing in on the dimension score calculations.

To understand the calculations to be surveyed here, readers are advised to first review the tables in Appendix II, which are central to most results to be presented in this study. Appendix II contains seven tables. Tables 1–3 sum up the normalized feature counts, i.e. the descriptive statistics, for each of the corpora annotated in the present study: table 1 for Internet relay chat, table 2 for split-window ICQ chat and table 3 for the SBC subset. Tables 5–7, furthermore, present the equivalent normalized frequency counts for each text in the three corpora: table 5 for the IRC texts, table 6 for the split-window chats and table 7 for the SBC subset texts. The tables mentioned have all been introduced in the sections above. Now, readers are recommended to turn to Appendix II table 4.

Appendix II table 4 constitutes the zero point (i.e. the baseline) for the standardization of frequencies and for the dimension score calculations, in Biber's (1988) study as well as in the present one. The table, drawn from Biber (1988: 77–78), gives the normalized frequencies of the features in Biber's *full* corpus of speech and writing, that is, the average figures for *all* of the LLC and LOB texts, as well as the two collections of letters (cf. Appendix I). The table thus forms the backdrop against which the individual texts and the individual genres in Biber (1988) were measured, as well as those studied in other MD analyses following

Biber (1988) (e.g. in Conrad & Biber 2001, as exemplified in section 2.2), and now it constitutes the baseline for calculations in the present study. To obtain standardized scores for individual features, and eventually dimension scores for individual texts and genres, the features, texts, and genres are all contrasted with Appendix II table 4.

As touched upon in section 2.3, the normalized frequencies are first contrasted with Biber's full corpus mean, i.e. the left-most column in table 4, and then with Biber's full corpus standard deviations, the rightmost column in table 4, to obtain standardized scores, henceforth "standard scores," for the features. More specifically, a standard score for a feature is obtained by performing the following calculation.

$$\text{standard score} = \frac{(\text{frequency in text} - \text{mean frequency in Biber's full corpus})}{(\text{standard deviation in Biber's full corpus})}$$

For example, consider first person pronouns (feature 6) in Internet relay chat text 1a (Appendix II table 5). The text has a normalized frequency of 48.0 first person pronouns. By applying the above calculation, the standard score arrived at for first person pronouns in IRC text 1a is 0.8; that is, the frequency of first person pronouns in IRC text 1a is 0.8 standard deviations higher than Biber's (1988) full corpus mean. Once the standard scores for first person pronouns have been computed for all the IRC texts, the mean of these, viz. 1.1, may be taken to be the standard score for first person pronouns in the whole genre of "Internet relay chat."

Next, consider the same feature in split-window ICQ chat text 1 (Appendix II table 6). First person pronouns appear to be more common in split-window chats than in IRC. Accordingly, by the above calculation, the normalized frequency of 110.5 first person pronouns yields the standard score 3.2 for the feature in split-window ICQ text 1. The average standard score arrived at for first person pronouns in the whole genre of split-window ICQ chat, once computed, is 2.4. In other words, first person pronouns are more than two standard deviations more frequent in split-window ICQ chat than in Biber's corpus as a whole. Features with absolute standard scores above 2.0 deviate markedly from Biber's mean for speech and writing, i.e. they are markedly frequent. First person pronouns may, consequently, be regarded as one of the most salient features in conversational writing, by virtue of their high frequency in split-window ICQ chat (despite their not being as salient in IRC).

Chapter 4 of the present study explores the features that deviate from Biber's mean by more than two standard deviations ($|\text{s.d.}| > 2.0$) in either, or both, of the

conversational writing genres.⁴⁸ These features are seen to characterize conversational writing by their high relative frequency, or relative infrequency, in the chats and are the most influential (“most salient”) contributors to the dimension score(s) of the relevant chat genre(s). As mentioned in section 2.3, chapter 4 also considers other salient features of conversational writing, those studied in previous accounts of CMC discourse (e.g. modal auxiliaries, paralinguistic features, emoticons and abbreviations) as well as previously understudied aspects of conversational writing, such as its lexical density and inserts. *That* is how the present study “lingers” in step 5 of the MD methodology (cf. section 2.3), before moving on to present the results of the final step in the MD methodology, the dimensions scores of the new genres, in chapter 5.

The computation of dimension scores for the new genres “Internet relay chat,” “split-window ICQ chat” and “face-to-face conversations SBC” followed the procedure described for Biber’s (1988) genres in section 2.3. A genre’s dimension score is found by averaging the dimension scores for all texts in a genre. This means that dimension scores are first computed for the individual texts. As mentioned in section 2.3, the dimension score of a text is computed by summing the standard scores for the features co-occurring on the dimension, except on Dimensions 1 and 3, on which the sum of the “negative” features’ standard scores is subtracted from the sum of the “positive” features’ standard scores (cf. table 2.2). Section 2.3 exemplified the dimension score calculation for a general fiction text on Dimension 2 (as explained in Biber 1988: 94–95). The present section briefly considers Internet relay chat text 1a on the same dimension, to further illustrate the procedure.

The dimension score for IRC text 1a is calculated by summing the standard scores of the features co-occurring on Dimension 2 (cf. table 2.2). In the chatted text, the features display a much lower incidence than in the general fiction text exemplified in section 2.3, an incidence generally even lower than the mean for Biber’s full corpus. The standard scores for the features on Dimension 2 in IRC text 1a are -0.7 for past tense verbs, -0.8 for third person pronouns, -1.3 for perfect aspect verbs, -1.1 for public verbs, 0.2 for synthetic negation and -0.6 for present participial clauses, the negative numbers indicating that the features are rarer than in Biber’s mean. The resulting dimension score for the text is thus -4.2,⁴⁹ which reflects the sparsity of the features on this dimension. While a high incidence of Dimension 2 features indicates a narrative concern in a text, as in

48 Appendix V lists the features with a |standard score| >2.0 in the genres studied.

49 The value -4.2 is the sum of unrounded standard scores.

the general fiction text exemplified in section 2.3, a low incidence indicates a that a text is unmarked for narrative concern. As it turns out, most IRC texts, like text 1a, display a considerable paucity of the lexico-grammatical markers of narration co-occurring on Dimension 2. The average dimension score for the genre, consequently, turns out to be very low, positioning Internet relay chat on the non-narrative extreme of the dimension scale, opposite to Biber's (1988) fiction texts. The dimension scores of the conversational writing genres, as well as those of the SBC subset, will be further presented and discussed alongside Biber's (1988) genres, on all dimensions, in chapter 5.

As astute readers may have noticed upon review of the tables in Appendix II, a dimension score for a genre may equally well be computed directly by summing the standardized scores for features in the whole genre (e.g. for IRC, those computed by contrasting Appendix II table 1 with table 4), without considering the dimension scores of the individual texts. This is feasible because the descriptive statistics for the whole genre (cf. Appendix II table 1) in reality simply reflect the average frequencies of the genre's constituent texts (cf. Appendix II table 5). The roundabout way for computing dimension scores (via individual texts) was, nevertheless, taken in the present study for the sake of adherence to Biber's (1988: 94–95) description of the procedure. Moreover, besides presenting the dimension scores for the genres, chapter 5 will also present the spread of dimension scores across the individual texts (e.g. as minimum and maximum values), results that inevitably rely on the computation of dimension scores for individual texts. Lastly, statistical tests also rely on the availability of such scores.

3.6 Average figures for writing and speech, respectively

As mentioned in chapter 1, the primary purpose of the first results chapter (chapter 4) is to document the features that are salient in conversational writing. The chapter expounds on the incidence in SCMC and SSCMC of such features and contrasts this with the distribution of the same features in writing, speech and ACMC, i.e. at the level of medium (cf. section 1.2). The media to be contrasted in chapter 4, as described, are “writing,” “speech,” “ACMC,” “SCMC” and “SSCMC” (even though the latter three, of course, comprise only one prototypical genre each, BBS conferencing, Internet relay chat and split-window ICQ chat). For all the features to be investigated, normalized frequencies will be contrasted, rather than standard scores. The most salient features, i.e. those that in conversational writing deviate by more than two standard deviations ($|s.d.| > 2.0$) from Biber's (1988) mean, will be treated in sections 4.2 (viz. personal pronouns) and 4.4 (the other most salient features); section 4.1 explains why they are presented in

separate sections. Chapter 4 also considers several other Biber (1988) features, as well as features typical of conversational writing that were not studied in Biber (1988). Whenever possible, the quantitative findings are contrasted with the quantitative findings in Collot's (1991) corpus of ACMC, as well as with the findings for writing and speech, respectively. The figures for ACMC are derived directly from Collot (1991), and the figures for SCMC and SSCMC derive from the feature counts in this investigation. The present section is dedicated to describing how average figures for the media "writing" and "speech" were obtained.

In the section above (3.5), Appendix II table 4 was seen to represent the "zero point" for comparisons of all genres, as it summarizes the mean frequencies for the features in Biber's (1988) full corpus of written and spoken texts. Appendix II table 4 (from Biber 1988: 77–78) can thus be regarded as representing the mean frequencies of the features in the English language overall. In the present study, by analogy, the written genres studied in Biber (1988) are taken to represent the medium of writing. Average figures for writing, accordingly, were obtained by considering the average normalized frequencies for the written texts studied by Biber (1988). Biber (1988: 247–263) details the average normalized frequencies for the 17 genres of writing (to save space, Biber's tables are not reproduced here, although Appendix I presents a list of the genres). The following algorithm was employed to compute the normalized frequency in a medium.

$$\text{normalized frequency in medium} = \frac{\sum((\text{normalized frequency in genre}) \times (\text{no. of texts in genre}))}{\sum(\text{no. of texts in genre})}$$

To exemplify the computation, we will consider the occurrence of first person pronouns in "writing." As indicated in Biber (1988: 247–263), the genre "press reportage" has a normalized frequency of 9.5 first person pronouns, "press editorials" 11.2, "press reviews" 7.5, etc. As seen in Appendix I here, Biber's (1988) written corpus consists of 44 texts of press reportage, 27 texts of press editorials, 17 texts of press reviews, etc. The normalized frequencies are occurrences per 1,000 words of running text and, for the current purpose, each text in a genre may be seen to consist of the average normalized frequency for a feature in the genre. To the full "corpus" to represent the normalized frequency in writing, press reportage thus contributes 44×9.5 first person pronouns, as the genre consists of 44 texts, each containing on average 9.5 first person pronouns. The genre of press editorials, by inference, contributes 27×11.2 first person pronouns, as each text contains on average 11.2 first person pronouns. By summing the first person pronouns calculated thus in all the genres of writing, and dividing the sum by

the total number of written texts (i.e. 340; cf. Appendix I), the average normalized figure for first person pronouns in writing is obtained, that is 17.0.⁵⁰

Average normalized frequencies for all of Biber's (1988) features to represent "writing" here were obtained by the same procedure, even for TTR and word length. As mentioned, the normalized frequencies for salient features will be presented and discussed in chapter 4. For TTR, the presentation of average figures in chapter 4 will be complemented with standard deviations. The standard deviation for TTR in writing was computed by considering the standard deviations for TTR in Biber's individual genres, given in connection with the normalized frequencies in Biber (1988: 247–263). The calculation was carried out by applying the following equation, in which x is the TTR in each genre and n is the number of texts involved.

$$\text{TTR standard deviation in medium} = \sqrt{\frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n-1}}$$

Turning now to the medium of "speech," readers may recall (from section 3.4) that the medium of speech in the present study is represented by Biber's six genres of speech, deriving from LLC, combined with the SBC subset face-to-face conversations genre. Average normalized frequencies for the features in Biber's (1988) spoken genres are presented in Biber (1988: 264–269), and Appendix I here lists the number of texts in each of Biber's spoken genres. The normalized frequencies for the same features in the SBC subset of face-to-face conversations are given in Appendix II table 3. In the present study, the average normalized frequencies for Biber features in "speech" were computed by considering the distribution in a combined "corpus" of speech consisting of the 141 texts from LLC studied by Biber (Appendix I), and the 14 texts from the SBC subset, once more applying the algorithm for "normalized frequency in medium" given above.

As an example for "speech," consider again the distribution of first person pronouns. Biber (1988: 264–269) tabulates the mean normalized frequencies for first person pronouns in the LLC genres studied. In "face-to-face conversations" (LLC), there are 57.9; in "telephone conversations," 70.7; "interviews," 50.5; "broadcasts," 11.8; "spontaneous speeches," 60.4; and in "prepared speeches,"

50 This method draws upon Biber's (1988) procedure for computing the frequencies in the full corpus of writing and speech (1988: 77–78), the results of which (see Appendix II table 4 here) appear to average the normalized frequencies in all the individual genres (1988: 246–269), for each genre taking into account its number of texts.

41.8. Appendix II table 3 in the present study correspondingly presents the normalized frequency for first person pronouns (feature no. 6) in the “face-to-face conversations SBC” genre, viz. 61.0. As mentioned, Biber’s LLC spoken genres comprise a total of 141 texts (cf. Appendix I) and the SBC subset consists of 14 texts (cf. table 3.1). In both corpora, each text in a genre may be considered to contain the normalized frequency for the feature in the genre. To the full “corpus” to represent speech in the present study, face-to-face conversations in LLC accordingly contribute 44×57.9 first person pronouns, telephone conversations contribute 27×70.7 , interviews 22×50.5 , etc., and, finally, face-to-face conversations from the SBC subset contribute 14×61.0 first person pronouns. After all these contributions are summed, the sum is divided by the total number of texts (i.e. 155, cf. Appendix I and table 3.1) to obtain the average normalized figure for first person pronouns in speech, that is, 52.8. The same procedure was then used to compute average frequencies in speech for all of Biber’s features, as well as the average figures for speech as regards TTR and word length. The TTR standard deviation for speech was calculated by using the formula “TTR standard deviation in medium” above, as explained for “writing.”

In sum, the normalized frequencies, TTR and word length for writing to be presented in chapter 4 represent Biber’s (1988) 17 genres of writing, and the normalized frequencies, TTR and word length for speech represent Biber’s (1988) six genres of speech, supplemented with the SBC subset face-to-face conversations genre. In the comparisons between conversational writing and speech, however, recurring reference will be made not just to the whole “corpus” of speech (Biber’s six genres + the SBC subset), but also, more importantly, to the conversational genres it contains. Whereas tables and figures in chapter 4 by default present the average figures for “speech” overall, explanations and discussions frequently indicate its constituent average figures for the individual conversational genres, that is, for face-to-face conversations and telephone conversations from LLC, and for face-to-face conversations from the SBC subset. The present study, after all, is concerned not just with the comparison of conversational writing to writing, ACMC and speech, but also, more specifically, with the similarities, or differences, between conversational writing and the spoken conversational genres.

3.7 Chapter summary

Chapter 3 has outlined the methodology for obtaining the data to be investigated in the present study – from the collection of the textual material to the quantitative results. After presenting the size of the three corpora compiled for the study, each corpus was treated in a separate section. First, the recording, adaptation and

annotation of the Internet relay chat corpus was described; second, the collection and annotation of the split-window ICQ chat corpus was explored; and third, the sampling of SBC was motivated and explained, as well as the annotation of the resulting SBC subset corpus of face-to-face conversations. For each corpus, two important tables in Appendix II were highlighted – one summarizing the normalized frequency counts for Biber's 67 features in the corpus and one detailing the normalized frequencies in individual texts. Next, the chapter described the process of standardizing the normalized frequencies. The standard scores then provided the requisite input for computing dimension scores for the genres under investigation, dimension scores that will be presented and discussed alongside textual examples in chapter 5. The penultimate section, finally, explained how average figures for the media writing and speech, respectively, were computed, for comparisons with the CMC media in the ensuing chapter. Chapter 4 is now at hand, which will present the salient features found in conversational writing.

Chapter 4. Salient features in conversational writing

4.1 Introductory remarks

This chapter presents the salient features of conversational writing, both those that have become conspicuous through the standard score calculations using Biber's (1988) methodology, explained in chapter 3, and those that are salient for other reasons. The chapter serves both as a prelude and a complement to the final results of the application of Biber's (1988) methodology to be presented in chapter 5. The principal aim of the present chapter is to point out, describe and discuss the salient features and the functions they serve in conversational writing. Firstly, we will investigate the use of modal auxiliaries and personal pronouns in conversational writing. Modal auxiliaries and personal pronouns are two of the main carriers of interpersonal meaning in language, defined in Halliday's system of semiotics (1985a, 2004), and therefore will be discussed under one and the same heading in the second section of this chapter (4.2). Their distribution in the conversational writing genres reveals a great deal about the modality of the discourse and the presentation of self, enabling informed contrastive analysis of the chatted and spoken texts. The third section of the chapter, 4.3, investigates the lexical properties of conversational writing by contrasting measures of word length, type-token ratio and lexical density in writing, speech and the conversational writing genres. Sections 4.2 and 4.3 largely draw on the choice of features in Yates' (1993) application of Halliday's semiotics to asynchronous CMC, and thus, chiefly, serve to complement the field of CMC variation studies with the analogous documentation of synchronous data. The two sections are kept together to facilitate for readers to compare the results with those in Yates' 1993 study. The fourth section, 4.4, departs from Yates, but stays closely tuned to Biber's (1988) methodology in that it presents the most salient features annotated in the conversational writing corpora. In the present study, ten features altogether have been found to deviate from Biber's (1988) mean for speech and writing by more than two standard deviations. Two of these are first and second person pronouns, which are addressed in section 4.2. The fourth section of this chapter, section 4.4, presents the remaining eight of these features, and what each of them reveals about the kind of communication going on in the chats. The fifth section, 4.5, goes on to survey the paralinguistic cues and extra-linguistic features found in the conversational writing corpora, and the penultimate section of the

chapter, 4.6, presents two salient linguistic features that are not among Biber's list of features, but that nevertheless serve important functions in computer-mediated conversational writing: inserts and emotives. The last section, 4.7, then sums up the results presented in the chapter.

The genres of conversational writing, IRC and split-window ICQ chat, are subsumed into their respective media categories in the present chapter: the media of synchronous and supersynchronous CMC (SCMC and SSCMC), as explained in chapters 1 and 2. The distributions of the linguistic features in SCMC and SSCMC are contrasted with the distributions in three other media: writing, speech and ACMC (asynchronous computer-mediated communication). ACMC is included in this chapter mostly as a reference point, and will receive rather cursory treatment (the focal concern of the study being synchronous and supersynchronous chat), but its inclusion here serves as a useful reminder of the inherent variability of computer-mediated texts. The five media to be compared in this chapter are represented by the following corpora:

Writing:	LOB + private and professional letters (as sampled by Biber 1988; see Appendix I)
Speech:	LLC (as sampled by Biber 1988; see Appendix I) + SBC subset, i.e. first c. 712 words of each text in SBC part 1 (annotated by Jonsson)
ACMC:	"ELC other" corpus of BBS conferencing (recorded and annotated by Collot 1991)
SCMC:	UCOW, the IRC component (recorded and annotated by Jonsson in 2002)
SSCMC:	UCOW, the split-window ICQ component (recorded and annotated by Jonsson in 2004)

The corpus of ACMC, called "ELC other" ("Electronic Language Corpus other"), was collected and annotated by Milena Collot (Collot 1991, Collot & Belmore 1996). It is not available as raw texts, but was annotated with the Biber tags in Collot's original study and represented as feature count data (in Collot 1991), from which the figures are derived for the present comparison. Collot's corpus consists of messages posted to an international bulletin board system, a BBS, located in Canada. It comprises 115,618 words and was collected from nine conferences, their topics ranging from "Chit-Chat" to "Medical" (Collot 1991: 45). The designation "other" implies that the messages were composed online, as opposed to messages composed offline, which were compiled into a separate corpus ("ELC off-line," not to be considered here). Collot was able to positively identify offline messages as they contained a software-generated marker, and those which lacked the marker were assumed to be written online. Collot, however, notes that

“there is always the possibility that certain messages were pre-written using an ordinary word processor or editor” (Collot 1991: 45), which would not add the marker. By labeling the resulting corpus “other” instead of “on-line” she implies that it contains, but is not necessarily limited to, online texts (Collot 1991: 46).

As mentioned in section 2.5, various modes of ACMC have been studied by linguists and communication scholars over the years, including computer conferencing systems (Korsgaard Sorensen 1993, Yates 1993, 1996, Davis & Brewer 1997), listservs (Herring 1996b), newsgroups (Severinson Eklundh 2010), BBSs (Collot 1991, Collot & Belmore 1996), web fora (LeBlanc 2005), e-mail (Yates & Orlikowski 1993, Maynor 1994, Baron 2000, Zitzen 2004, Anglemark 2009, Cho 2010, Georgakopoulou 2011b, Rowe 2011), weblogs (Scoble & Israel 2006, Anglemark 2009, Peterson 2011) and Twitter (Petrovic et al. 2010, Pak & Paroubek 2010). Collot’s study, however, appears to be the only one to have applied Biber’s (1988) full multidimensional analytical tool to ACMC. The readily available frequency counts in her study lend themselves conveniently to comparison with the feature frequencies found for the corpora annotated in the present study, and with those presented in Biber (1988) for LOB and LLC. Comparable frequency counts are particularly amenable to graphic, diagrammatic representation, which is why, in this chapter, Collot’s ACMC corpus will receive its own representation in the figures, even though, owing to the unavailability of comprehensive raw ACMC texts, the ACMC figures for some features will be left uncommented.

4.2 Distribution of modal auxiliary verbs and personal pronouns

In Bybee & Fleishmann’s (1995a) co-edited volume on modality, Guo (1995) briefly, but pertinently, considers English modals, positing that “[i]n English, physical ability can be expressed either by the modal auxiliary *can* or by the adjective *able*, as in *be able to*. Similarly, social permission can be expressed by *can* or *be permitted to*” (Guo 1995: 228, original italics). In each case, the two options are referentially interchangeable. However, the options differ in their grammatical status; modals belong to a closed grammatical class and are thus more grammaticalized than adjectives and verbs, which leads Guo (1995) to further argue that:

This grammatical difference has significant consequences with regard to the meanings expressed. With lexical forms such as *able* or *permitted*, the speaker presents a fact without any personal involvement. We interpret the utterance as ‘I’m stating X to you’. But when modal auxiliaries are used, the resulting utterances are colored by speaker involvement in the form of opinion, affect, or personal dynamics. We interpret such

utterances as 'I'm challenging/objecting to/arguing with you by stating X to you'. (Guo 1995: 228)

Modality thus indicate the speaker's evaluation of his/her proposition, for instance the gradience of likeliness (if the speech event is a proposition) or desirability (if it is a proposal) (Halliday 2004: 116).

Halliday (2004) discusses finite verbs in terms of what they bring in to the clause and their functions in the systems of polarity and modality. Finiteness is expressed through a verbal operator, which is either temporal (realized by tense) or modal (realized through modal auxiliaries). In the system of polarity, the operators appear in positive and negative form (as e.g. *it is/isn't, do that/don't do that, you can/can't do it*), whereas in the system of modality there are intermediate degrees (e.g. *it must/will/may be, you must/should/may do that*, etc.). Polarity is the choice between yes and no, whereas modality construes "the region of uncertainty that lies between 'yes' and 'no'" (Halliday 2004: 147). In this way, the modality system of a language is an important functional component carrying interpersonal meaning (Halliday & Hasan 1989, Halliday 2004). In fact, Guo (1995: 229) proposes that language actually "developed the grammatical category of modal auxiliaries to serve the function of regulating interpersonal relations in social interaction."

Several studies have found modals to be more common in speech than in writing (e.g. Coates 1983, Biber et al. 1999, Kennedy 2002). Bybee & Fleischman (1995b: 8) make the points that "many modal functions surface only in face-to-face interactive discourse," that is, they depend on dialogic "speaker-addressee interaction" (ibid.) and that "modals can be viewed as strategic linguistic tools for the construction of social reality" (ibid.). In a similar vein, Kennedy (2002) notes that modals reflect the role of modality in face-to-face conversations: "to hedge and soften utterances and express subtle differences in degrees of certainty, attitudes, value judgements and the truth conditions of propositional content" (2002: 88, also noted by Andersen 2006: 18). Lexico-grammatically, interpersonal meaning is carried by e.g. markers of mood (indicative or imperative, but also by interrogatives, e.g. WH-interrogatives, as we shall see later), the use of personal pronouns and the choice of modal auxiliaries (Halliday 1978, 1985a, Halliday & Hasan 1989, Halliday 2004). Together, these features reflect the semiotic "tenor" of the communication (Halliday & Hasan 1989), as touched upon in section 2.4; that is, they reflect the personal relationships involved in the communication.

In their model of critical linguistics, Fowler & Kress (1979) consider what they call the grammar of modality, concentrating upon, among other things, the last two linguistic items mentioned: personal pronouns and modal auxiliaries

(see also Hodge & Kress 1988, Yates 1993: 106). Following their example, albeit in the reverse order, we will look first at the distribution of modal auxiliaries, then at the use of personal pronouns, in the genres of conversational writing. The purpose of the investigation is to find out to what extent these two are used in conversational writing, and how their distribution in these genres relates to that in writing and speech, as well as to that in the medium of ACMC. Given the interpersonal nature of conversational writing, and the importance assigned to modal auxiliaries as carriers of interpersonal meaning, we should expect a distribution in conversational writing similar to speech, or, more specifically, similar to traditional conversation (face-to-face and telephone conversations).

The modals included in Biber (1988), and therefore tagged in UCOW and the SBC subset, are the following:⁵¹

Possibility modals: *can, may, might, could* (+ negated forms)

Necessity modals: *ought, should, must* (+ negated forms)

Prediction modals: *will, would, shall* (+ contracted and negated forms)

The distribution of modal auxiliaries in each medium is illustrated in figure 4.1, based on table 4.1. (In the present chapter, all figures and tables are based on average, normalized frequencies per thousand words and derive from Biber 1988: 247–263 for writing, Collot 1991: 69–70 for ACMC, Biber 1988: 264–269 and Appendix II table 3 for speech, Appendix II table 1 for SCMC, and Appendix II table 2 for SSCMC, unless otherwise indicated.) For the results of statistical significance tests among SCMC, SSCMC, writing and speech for the features treated in this chapter, see Appendix VI.⁵² In the figures and tables, the media are ordered according to their basic synchronicity of communication (cf. table 1.1), i.e. from most asynchronous on the left (writing) to supersynchronous (SSCMC) on the right (although the conversational genres of speech, of course, may exceed SCMC in synchronicity). Immediately noticeable in figure 4.1 is the elevation of the ACMC and SSCMC bars. The texts of the two media display identical and remarkably high distributions of modals (totals of 20.5 per thousand words, compared to 12.8 for writing and 15.1 for speech). The frequent use of modals suggests that communication participants in both media are interpersonally

51 The generic term “possibility modals” here designates the modals marking possibility, ability or permission; “necessity modals” designates the modals marking necessity or obligation, and “prediction modals” the modals marking prediction or volition (Biber 1988: 241, Quirk et al. 1985: 219, Coates 1983).

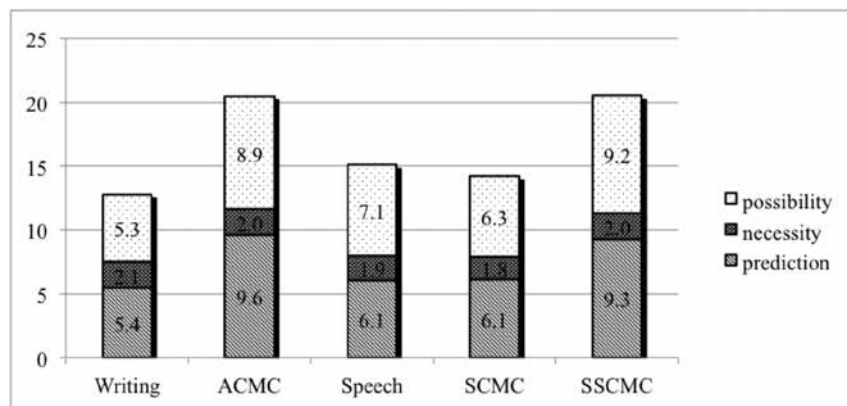
52 No statistical test was carried out on the ACMC frequencies, as the requisite data is not provided in Collot (1991).

involved to a high degree, i.e. they exchange discourse that is “colored by speaker involvement in the form of opinion, affect, or personal dynamics,” to use the words of Guo (1995: 228). With regard to the conversational nature of chatted texts, this seems to be a logical finding, but as for asynchronous texts, it is more unexpected. The overall modal auxiliary use in SCMC (14.2) is also higher than in writing, but slightly lower than in speech, a finding to be discussed later.

Table 4.1: Frequencies of possibility, necessity and prediction modals per 1,000 words (normalized values)

	Writing	ACMC	Speech	SCMC	SSCMC
possibility modals	5.3	8.9	7.1	6.3	9.2
necessity modals	2.1	2.0	1.9	1.8	2.0
prediction modals	5.4	9.6	6.1	6.1	9.3
total	12.8	20.5	15.1	14.2	20.5

Figure 4.1: Distribution of possibility, necessity and prediction modals per 1,000 words (normalized values).



To enable functional comparison of the three CMC media, a brief introduction to the interpersonal aspects of the ACMC corpus is needed. The “ELC other” texts of ACMC (Collot’s material) are unavailable for scrutiny, except for a few examples cited in Collot (1991), but Collot describes the texts as discussions about issues pertaining to the conference topics: e.g. “Medical,” “Finance,” “Sports,” “Current Events,” “Science,” “Cooking,” “Chit-chat” and “Film and Music.” Participants in

The personal relationships under development in the BBS approximate the relationships of previous acquaintances, such as those between the chatters in the split-window ICQ corpus. To form and develop relationships, interlocutors on a BBS, as well as in ICQ, need to stay sensitive to each other's opinions and propositions. By modalizing their utterances, they maintain the ongoing dynamics of social interaction. In example (2) from split-window ICQ chat, part of a discussion of college plans between the two high school classmates recorded, participants' previous acquaintance outside of the medium shines through. The example contains three possibility modals (*could*, *can*), one necessity modal (*should*) and one prediction modal (*'ll*).

- (2) <K> so did you get a scholarship fir tennis or are u just going
<11> yea
<K> did you talk to the coach
<K> o
<11> he said that i have a pretty good chance
<K> is he goign to come see you play
<11> yea this season it starts in march
<11> what school are you going to again
<K> when do u start
<K> [college name] its down in [city name]
<11> are you definetly playing there
<K> yea i went donw and the coach said i *could* start as a true freshmen
<11> well thats good
<K> he came up to talk to my parents and we ate dinner and all kinds of shit
<11> that cool
<K> so yea i *can* sing my letter of intent when ever
<11> thatws cool
<K> you *should* come down
<11> yea defenetly
<K> and i'll come up we *can* chill

Split-window ICQ chat text 10 (UCOW)

While the BBS conference participants seem inclined to form fairly long-standing friendships in the medium, and the participants in ICQ chat are previous acquaintances even outside of the medium, the participants in IRC are casual acquaintances forming fleeting relationships. Messages in conversational writing, unlike asynchronous messages, are produced on the fly; they appear briefly on the screen and then scroll off. However, while split-window ICQ chatters can scroll back and edit their turns, IRC participants' turns, once posted, are uneditable. To get a foothold in the jumble of turns, IRC chatters produce very short

messages; many turns are there simply to signal the user's active presence. To detect conversational threads among the turns, participants must manage to untangle the jumble and the constant flow of server-generated messages (cf. Elser & Charniak 2008). Occasionally, conversations involve more than two participants and last for several minutes, but more often they are dyadic, short-lived and ephemeral. Example (3) from the IRC channel #20_something comprises less than a minute of communication. The example is unformatted, for illustrative purposes, and thus retains time stamps (“[22:14]”) and server-generated messages (lines marked by “***”).⁵³

- (3) [22:14] <^^katy^> wbbb crash
 [22:14] <Princess> i meant shuuuuu. i am hiding from lizard
 [22:14] <^^Crash^^> ty hun hugzz
 [22:14] *** Amike-USA has joined #20_something
 [22:14] <chanel> well you *can* write...but i have a bf...u *should* know that
 [22:14] *** Sweetpea-Soup is now known as ^Sara^
 [22:14] <iowachick> ne one from iowa or illinois?
 [22:14] *** Lara2002-117553 has quit IRC (Connection reset by peer)
 [22:14] <Chaser> chanel babe *can* i have your emailaddress
 [22:14] <^^Crash^^> ty katy babes
 [22:14] *** Pet-Ratty has left #20_something
 [22:14] *** Pablo is now known as Argentino
 [22:14] <chanel> um princess...hate to bust your bubble but chaser is lizard
 [22:14] <^^katy^> np crash..lol
 [22:14] <chanel> you had it chaser
 [22:14] <Princess> oh
 [22:14] <Farkles> princess? tisha?
 [22:14] <Argentino> holaaaaaaaaaaaa
 [22:14] *** stalesgr has quit IRC (<http://ircqnet.icq.com/>)
 [22:14] <^^Crash^^> hi chanel
 [22:14] *** canadiangirl has left #20_something
 [22:14] <chanel> hiya crash;))
 [22:14] <Princess> yes
 [22:14] *** iowachick has quit IRC (<http://ircqnet.icq.com/>)
- Internet relay chat text 1b (UCOW)

53 All subsequent examples of IRC have been purged of time stamps, join and quit messages and other server-generated text. As described in chapter 3, only user-generated conversational text was annotated and included in the linguistic feature counts.

Example (3) illustrates the typically superficial relationships among the IRC participants, a feature that becomes strikingly evident from the full corpus. Short-lived conversations take place between e.g. ^^katy^ and ^^Crash^^, Princess and Farkles, and chanel and Chaser, while iowachick and Argentino simply signal their presence/entrance. Large portions of the IRC corpus, like example (3), consist of greetings and phatic devices whereby users announce their own and others' entrance (*holaaaaaaaaaaaaa*, *hiya* and *wbbb*, i.e. welcome back, where letter *b* is repeated for emphatic endorsement). Politeness terms abound (*ty*, meaning "thank you," *np*, meaning "no problem"), as is often the case in spoken discourse. Considering the rarity of substantial discussion, the high ratio of greetings and phatic devices, and moreover, the impact of altogether verbless turns, it is rather unexpected that modal auxiliaries should find their way into the discourse at all. Judging from figure 4.1, however, modals in SCMC are almost as frequent as in spoken discourse (although no significant difference obtains between the distribution of modals in SCMC, as compared to either writing or speech; see Appendix VI). Nevertheless, it seems that, as messages become more lengthy, or rather, whenever a turn contains a full clause, i.e. a subject and a main verb, the main verb is often preceded by a modal auxiliary, as in *well you can write...but i have a bf...u should know that* (in example 3).

Example (3) contains two instances of the possibility modal *can* and one necessity modal, *should*. On Biber's (1988) dimensions of textual variation, possibility modals load on Dimension 1, as markers of involved production, whereas necessity and prediction modals load on Dimension 4, as markers of overt expression of persuasion. The frequencies illustrated in figure 4.1 indicate higher values than in writing for possibility and prediction modals in all three CMC media, but lower values for necessity modals. The media of ACMC and SSCMC surpass both writing and speech with regard to their distribution of possibility and prediction modals, whereas SCMC displays no significant difference in the distribution of modals compared to either writing or speech (Appendix VI). The division of modals into their semantic categories and their respective distributions in the five media will not be further explored in this section, however, as the network of modal meanings is too complex for the brief analysis intended here. The annotation of their semantic categories was done primarily to enable the positioning of the conversational writing genres on Biber's textual dimensions; see chapter 5, where their respective functions will be explored. Moreover, the modals were not annotated for root and epistemic meanings in all five media (as described in Coates 1983 or Coates 1995), rendering more detailed exploration

of their values impossible. Worthy of notice, nevertheless, is that in Yates' (1993) study of, inter alia, modals in ACMC (from a computer conferencing system), possibility modals were divided into their root and epistemic meanings (by analogy with Coates 1983). Yates' results show that the ACMC discourse makes more frequent use of modals than either speech or writing overall, except for the use of modals of epistemic possibility (*may, might*), which show a distribution similar to writing (higher than speech). In the corpora of SCMC and SSCMC, however, modals of epistemic possibility are found in fewer than every third IRC text, and as less than one instance per ICQ text, which makes them rarer than what Yates found for either writing, speech or ACMC, giving the conversational writing texts a more spoken, than written, character.

Among the genres amalgamated into the mean score for speech in figure 4.1 are the genres of face-to-face conversations, with 15.6 (in LLC) and 16.2 modals (in the SBC subset) per thousand words, and telephone conversations, with 18.3 modals per thousand words (Biber 1988: 264–265). In the conversation genre of the Longman Spoken and Written English Corpus (LSWE), Biber et al. (1999: 486) and Biber (2004) find approximately 22 core modal verbs per thousand words – a rate slightly higher than in any of the corpora studied here. This means, nevertheless, that the figures for ACMC and SSCMC (20.5), as shown in figure 4.1, are more in keeping with the results for core modals in conversation presented in Biber et al. (1999) and Biber (2004), approximately 22, than they are with those for conversations in Biber (1988), 15.6 and 18.3. Biber et al. (1999: 489) find both core modal and semi-modal verbs to be more common in conversation than in any of three written registers (fiction, news and academic prose).⁵⁴ The corpora of ACMC and SSCMC, in figure 4.1, show an overall frequency of core modals that is approximately 60 percent higher than in traditional writing.

Judging from the figures in Collot (1991: 74–75), the BBS conferences (represented as ACMC in figure 4.1) vary in their use of modals; the conferences “Medical,” “Finance” and “Sports” show a greater frequency of modals, with

54 In Biber et al. (1999: 486) semi-modals (*have to, (have) got to, (had) better* and *be going to*) add approximately another five modals to the count for the LSWE conversation genre, yielding a total frequency for modals of approximately 27 per 1,000 words. To relate to this figure, the IRC, ICQ and SBC subset corpora were annotated for semi-modals in a complementary study (unpublished), in which approximately three semi-modals per 1,000 words were found in IRC, four in ICQ and seven in the SBC subset. More precisely, when semi-modals are included, the total normalized count for modals in IRC is 17.0, in ICQ 24.8 and in the SBC subset 23.6.

“Medical” contributing significantly to the high overall frequency, whereas participants in the “Chit Chat” conference use modals to an intermediate degree. The interlocutors in the split-window ICQ chat corpus (SSCMC in figure 4.1) are also interpersonally involved to a high degree, seeing as they are classmates and know each other in the real-life context. As seen in the discussion of example (3), IRC chat (SCMC in figure 4.1) is highly interactive and interpersonal, even though the rate of modals in this fleeting communication is the lowest of all three CMC media. The modal auxiliary use of IRC sits between that of face-to-face conversations from LLC (Biber 1988) and the average for writing, but to the extent that verb phrases do appear in IRC, they seem to contain no fewer modals than those in ICQ. Returning to Guo’s (1995: 229) statement that language actually “developed the grammatical category of modal auxiliaries to serve the function of regulating interpersonal relations in social interaction,” it can be concluded that in all three CMC media, such regulation is going on, even though only the ACMC and SSCMC users employ modal auxiliaries to the extent of speakers in the more recent accounts of conversation (those in Biber et al. 1999 and Biber 2004).

Turning now to a survey of personal pronoun use, it will be seen that the genres of CMC differ from writing and speech in other ways, but in ways that further highlight their functions as media for social interaction.

A subheading in Chafe’s (1982) article on involvement and detachment in literature proclaims that “speakers interact with their audiences, writers do not” (1982: 45, original is in all capital letters). The subheading follows upon Chafe’s characterization of speech and writing into “fragmented” vs. “integrated” discourse and sets the tone for his further delineation of speech and writing into the qualities representing “involvement” vs. “detachment.” Speakers are typically involved with their audience, a trait manifested, *inter alia*, in speakers’ more frequent reference to themselves, i.e. through their frequent use of first person pronouns (henceforth 1PP). Writers, on the other hand, are detached from their audience and more concerned with presenting “logically coherent,” “consistent and defensible” text which “will stand the test of time” (1982: 45). In his corpora, Chafe finds a ratio of approximately thirteen 1PP in speech to one in writing, the actual numbers being 61.5 and 4.6 per thousand words respectively (1982: 46). A few pages later, Chafe admits that his categorical statements regarding speech and writing apply to extremes on the continuum; his figures are from maximally differentiated samples, spontaneous conversation vs. academic prose. (The ratio in Biber (1988) for the equivalent genres,

face-to-face conversations vs. academic prose, is roughly the same: ten 1PP to one (1988: 264, 255).) Unknown to Chafe in 1982, however, was that in the next two decades genres were to appear, with texts in which the ratio at hand is challenged or augmented further. In SSCMC, more specifically in split-window ICQ chat (see figure 4.2, based on table 4.2), for instance, the ratio of 1PP is that of nearly sixteen to one, compared to academic prose (Appendix II table 2 vs. Biber 1988: 255), or more than nineteen to one, compared to Chafe's corpus of writing (Appendix II table 2 vs. Chafe 1982: 46). Moreover, these CMC genres represent *writing*, rather than speech. Or do they? This idiosyncratic, confounding finding is one of many that suggest the definition of SCMC and SSCMC as something other than either writing or speech, hence warranting the term "conversational writing."

The first, second and third person pronouns included in Biber (1988), and therefore tagged in UCOW, are the following:⁵⁵

- First person (1PP): *I, me, my, myself, we, us, our, ourselves*
(+ contracted forms)
- Second person (2PP): *you, your, yourself, yourselves*
(+ contracted forms)
- Third person (3PP): *she, her, herself, he, him, his, himself, they, them, their, themselves*
(+ contracted forms)

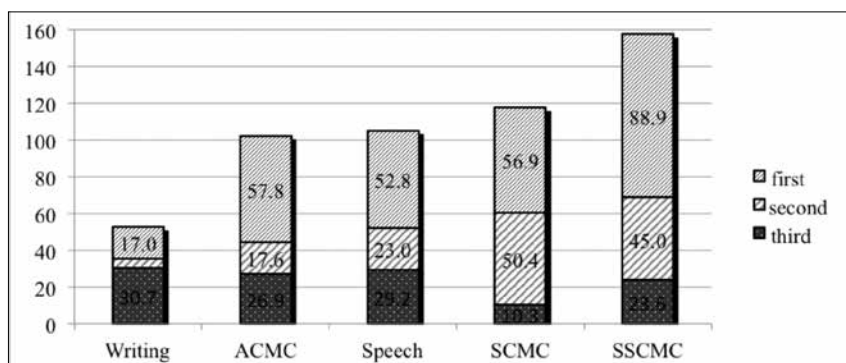
Table 4.2 and figure 4.2 present the distribution of personal pronouns in the media investigated.

55 Biber (1988: 225) subsumes personal, possessive and reflexive pronouns under the heading "personal pronouns." None of the forms for "it" are included, nor are the independent possessive pronouns (*mine, yours, etc.*) (cf. Quirk et al. 1985: 361).

Table 4.2: Frequencies of first, second and third person pronouns per 1,000 words (normalized values)

	Writing	ACMC	Speech	SCMC	SSCMC
first person pronouns	17.0	57.8	52.8	56.9	88.9
second person pronouns	5.0	17.6	23.0	50.4	45.0
third person pronouns	30.7	26.9	29.2	10.3	23.6
total	52.7	102.3	105.0	117.6	157.5

Figure 4.2: Distribution of first, second and third person pronouns per 1,000 words (normalized values).



First and second person pronouns are two of the features that, in either or both of the conversational writing genres, deviate from Biber's mean of writing and speech (Appendix II table 4) by more than two standard deviations ($|s.d.| > 2.0$). They are taken up in this section chiefly because, together with modal auxiliaries, they constitute important carriers of interpersonal meaning in language (Halliday 1985a, 2004, Yates 1993). All in all, ten features deviate thus; section 4.4, below, will explore the other eight: WH-questions, analytic negation, demonstrative and indefinite pronouns, present tense verbs, predicative adjectives, contractions and prepositional phrases. By their sheer relative frequency (or infrequency in the case of prepositional phrases), these features can be said to epitomize the linguistic character of conversational writing. As the word frequency lists (Appendix VII) show the first person singular pronoun (*I*) to be in a distinguished first position (i.e. as the most frequent lexeme) in all three corpora annotated in the present study, and the second person pronoun (*you*) among the top three in all, it seems befitting that our exploration begins with these.

As mentioned, Chafe (1982) and Chafe & Danielewicz (1987) claim that speakers' involvement with their audience is manifested in speakers' frequent reference to themselves. Writers, who rarely see their audience, typically use fewer first and second person pronouns. Chafe (1982) and Chafe & Danielewicz (1987) find that the relationships between speakers/listeners and writer/readers are encoded in language by the varying levels of involvement and detachment in speech and writing. Chafe (1982: 45) argues that the involvement typical of speech arises from the fact that:

It is typically the case that a speaker has face to face contact with the person with whom he or she is speaking. That means, for one thing, that the speaker and listener share a considerable amount of knowledge concerning the environment of the conversation. It also means that the speaker can monitor the effect of what he or she is saying on the listener, and that the listener is able to signal understanding and to ask for clarification. It means furthermore that the speaker is aware of an obligation to communicate what he or she has in mind in a way that reflects the richness of his or her thoughts [...] with the complex details of real experiences [...]. (Chafe 1982: 45)

Chafe (1982: 45) goes on to contrast the experiential involvement typical of speech with the typically detached nature of written discourse:

The situation of the writer is fundamentally different. His or her readers are displaced in time and space, and he or she may not even know in any specific terms who the audience will be. The result is that the writer is less concerned with experiential richness, and more concerned with producing something that will be consistent and defensible when read by different people at different times in different places, something that will stand the test of time. (Chafe 1982: 45)

Fowler & Kress (1979) also find that first person pronouns are rare in writing but regard this as an effect of "appropriate" attendant social practices rather than an effect of the medium (see also Yates 1993: 109). In other words, the "impersonal, generalizing tone of newspapers, textbooks, scientific articles" (Fowler & Kress 1979: 201) calls for a "[r]emoval of the pronoun associated with personal speech" (ibid.). Fowler & Kress, however, take note of varying subjectivity in different genres, observing slightly higher frequencies of first person pronouns in e.g. self-centered articles and eye-witness accounts (1979: 201) than in other writing. In a like manner, Chafe & Danielewicz (1987: 107) investigate first person pronoun use in four genres: conversations, lectures, academic papers and informal letters, finding informal letters to contain the highest number (57 per thousand words) – despite the written mode. Their finding underscores the purported significance of an identified audience, present *or* remote, for the formation of involved discourse, and leads Chafe & Danielewicz to conclude, like Fowler & Kress (1979), that other factors than the medium itself may be at play:

The use of first person pronouns is thus not necessarily a feature which differentiates spoken from written language, but rather a feature which the absence of a direct audience may even foster when the circumstances are right. (Chafe & Danielewicz 1987: 107)

Two circumstances that “foster” involved communication are thus 1) an identifiable, attentive and responsive audience (present or remote) and 2) a medium in which social and cultural practices permit the discussion of self. A third circumstance, as will be seen, is the synchronicity factor. That synchronicity is a predictor of high first and second person pronoun incidence is clearly illustrated in figure 4.2.

Each of the synchronous and supersynchronous genres of conversational writing, in figure 4.2, displays a combined usage of first and second person pronouns (1PP and 2PP) that surpasses either of the asynchronous media, as well as speech. In IRC, 1PP are about as common as in APMC, but 2PP are more frequent than in any of the other media. Furthermore, interlocutors in IRC are the least concerned with third person reference, as seen in the extremely low frequency of third person pronouns (3PP). Example (3) above, from IRC, characteristically contains four 1PP (*i*) and seven 2PP (*ty, you, u, your*), but no instance of a 3PP. “[T]he more first and second person pronouns chatters use, the more involved they are with their fellow interlocutors,” says Freiermuth (2003: 74) in his account of public America Online political chat channel data. In his data, 1PP and 2PP make up about 77 percent of the total personal pronoun use (2003: 128).⁵⁶ In the IRC data in figure 4.2, from the channels #20_something, #30_something, #Chat-World, #Family and #USA, the sum of 1PP and 2PP constitutes more than 90 percent of the personal pronouns (see further figure 4.3). This result seems to indicate an extremely high degree of involvement on behalf of the interlocutors in IRC. The communication indeed tends to center around the self and the second person, as in example (4), but rather than being used to express subjective opinions (as in political chat), the first and second person pronouns in IRC are mostly used for addressing others upon entering and leaving the channel (e.g. *i will be back, c ya, c u*), or for polite speech-act formulae (e.g. *ty*, meaning “thank you,” *yvw*, meaning “you’re very welcome”).

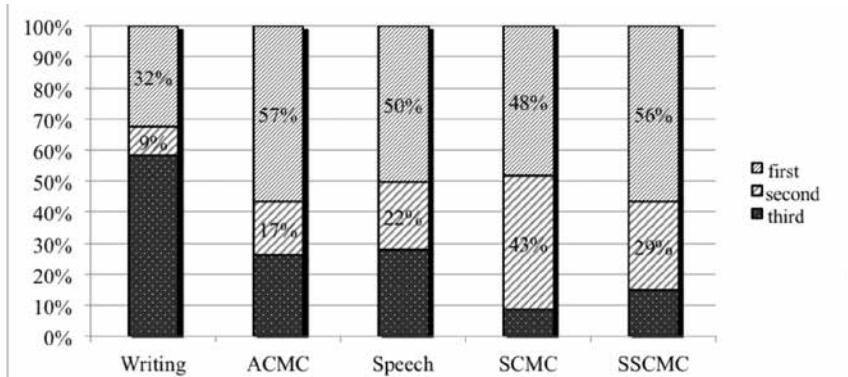
- (4) <Cheeky1> *i will be back*
 <|mad_max|> *ok*
 <|mad_max|> *take care*
 <Cheeky1> *gotta go for 5 minutes*

56 Unlike Biber (1988), Freiermuth (2003) includes the independent possessive pronouns (*mine, yours*, etc.), but not the reflexive pronouns (Freiermuth 2003: 127).

<Cheeky1> u 2 max sweety
 <Cheeky1> c ya in a sec u hunk of spunk
 <Cheeky1> hehehe
 <|mad_max|> c u
 <Cheeky1> cya

Internet relay chat text 3a (UCOW)

Figure 4.3: Proportions for first, second and third person pronouns of total personal pronoun use.



Returning to Chafe's (1982) description of involved discourse, quoted above, the IRC communicators, despite the written medium, "share a considerable amount of knowledge concerning the environment of the conversation" (1982: 45). Fellow IRC participants are identifiable, attentive and responsive, and participants know that the medium is intended for social interaction. The communication in the shared window is immediate and responses appear in seconds, as in face-to-face and telephone conversation. To the extent that responding turns appear, the interlocutor in IRC, as in spoken conversations, "can monitor the effect of what he or she is saying on the listener" (Chafe 1982: 45) and "the listener is able to signal understanding and to ask for clarification" (ibid.). Chafe's description of involved spoken discourse therefore for the most part holds true for IRC. Chafe's (1982) description of writing, however, is not applicable to conversational writing. IRC chatters are not "displaced in time and space" (1982: 45), as their communication is synchronous and appears in a shared virtual context. Through the list of logged-on participants, chatters have a notion of their audience, and their discourse therefore, like speech, appears to be more concerned with experiential richness than with the objective sharing of information.

In the supersynchronous medium of split-window ICQ chat, SSCMC in figures 4.2 and 4.3, interlocutors, to an even greater degree than other chatters, appear to be concerned with expressing subjectivity and personal opinion. First person pronouns abound in the corpus, as in (5) which contains nine 1PP (*i, me*) and three 2PP (*you, u*), but no 3PP.

- (5) <J> how come *you* didnt take bio II?
 <10> last year *i* started to like it after a bio class and *i* enjoyed it a lot
 <10> *i* did *i* had a good teacher
 <10> r *u* taking bio 2 lol
 <J> no.. to tell the truth.. *i* hate bio.. to *me*.. its all like studying things and not much creativity like calculus or physics.. were *you* really have to think to solve problems.. *i* guess *i* 0just like math in general
 <10> *i* just hate doing those long problems

Split-window ICQ chat text 9 (UCOW)

The chatters in the ICQ data, high school classmates, are slightly less concerned than the IRC chatters with the second person (judging from second person pronoun use). In the ICQ chats, second person pronouns are used less in greetings and politeness terms (*see you, thank you, you're welcome*) than they are in IRC, but more as parts of committed questions; see example (5). Used thus, the second person pronouns in ICQ, unlike those in IRC, reveal interlocutors' real-life acquaintance and the genuinely involved character of their communication (*how come you didnt take bio II?*, *r u taking bio 2*). Split-window ICQ chatters, furthermore, use more third person pronouns than do IRC chatters – another result of ICQ chatters' acquaintance outside of the medium and their exposure to the same human referents. (On the other hand, the human referents shared in IRC, i.e. the fellow chat participants, are mostly referred to by their nicknames and not by third person pronouns, to avoid deictic confusion.) In conclusion, split-window ICQ chat, like IRC, is more in harmony with the involved discourse typical of speech as defined by Chafe (1982) than with his definition of writing. Neither chatter has “face to face contact with the person with whom he or she is speaking” but the chatters “share a considerable amount of knowledge concerning the environment of the conversation”; they “can monitor the effect of what [they are] saying on the listener”, and “the listener is able to signal understanding and to ask for clarification” (1982: 45 for all four quotes). In the words of Chafe (1982) this means for split-window ICQ chatters, furthermore, that, like speakers, they are “aware of an obligation to communicate what [they have] in mind in a way that reflects the richness of [their] thoughts [...] with the complex details of real experiences” (Chafe 1982: 45).

Above, we identified three circumstances that foster involved communication: 1) an identifiable, attentive and responsive audience (present or remote),

2) a medium in which social and cultural practices permit the discussion of self, and 3) the synchronicity factor, which enables dialogic communication. Seeing the effect that previous acquaintance has upon ICQ chatters' discourse, a fourth circumstance might be added: 4) close personal acquaintance. Certainly, more factors could be added, but for the present purpose this collection will do. In combination, these factors all contribute to proximity and directness between interlocutors and increase the personal reference among them. In the SCMC corpus, the first three factors are at work and in the SSCMC corpus all four.

What about the asynchronous modes of CMC, then? Judging from figure 4.2, ACMC implements pronominal reference to about the same degree as speech, both as regards the combined use of first and second person pronouns, and as regards overall use. Judging from figures 4.2 and 4.3, ACMC users employ first person pronouns more than speakers and second person pronouns slightly less than speakers. Collot (1991), whose counts underlie the ACMC bars in figures 4.2 and 4.3, notes that first and second person pronouns in her corpus, among other features, "indicate a highly verbal, and personally involved style" (1991: 80) but does not delve further into their use (the ACMC example (1) above contains six 1PP and three 2PP). Yates' (1993) study of asynchronous computer conferencing texts, however, discusses pronominal reference at length, much of which inspired the above account for the synchronous chats. Yates finds first and second person pronouns to constitute 64 percent of all personal pronouns in his ACMC corpus.⁵⁷ In Collot's ACMC corpus "ELC other," the same proportion is 74 percent, slightly more than in Biber's (1988) genres of speech. In Biber's (1988) genres of writing, however, 1PP and 2PP together constitute only 41 percent of all personal pronouns (see figure 4.3). ACMC, despite being a written, asynchronous medium, therefore clearly deviates from the other genres of writing; for another thing, the overall use of personal pronouns in ACMC is nearly twice the number of traditional writing, as measured in normalized frequencies (see table 4.2).

What is it about the ACMC medium that makes for personal, involved communication of this kind? To answer this question, we must look at the written genre that most closely resembles the ACMC genre here: personal letters. Both ACMC and personal letters are produced under at least two of the circumstances that foster involved communication: they are directed at a presumably responding, albeit remote, audience and their attendant social practice is of an interactional

57 Yates (1993, 1996) does not specify whether possessive and/or reflexive pronouns are included in the count of personal pronouns, or exactly which personal pronouns are counted. Collot (1991), however, follows Biber's (1988) feature annotation scheme, which makes her figures ideally suited for comparison with the other media.

kind that permits, expects or condones the discussion/presentation of self. Personal letters are, moreover, exchanged between previous acquaintances. Casual ACMC messages, as seen in example (1) above, assume a similar personal tone as private letters, especially messages exchanged among ACMC users who consciously seek lasting friendships through the BBS. As mentioned before, Chafe & Danielewicz (1987) find first person pronouns to be more slightly more common in informal letters than in conversation. In seeking and maintaining friendship through asynchronous written media such as letters and ACMC, the presentation of self is evidently central, expected and culturally sustained. Among the four factors identified as triggers of involved communication, the synchronicity factor is the only one not at play in the asynchronous discourse.

A few references with regard to traditional asynchronous communication – viz. letters – are in place here. Besnier (1988, 1991, 1995) notes for Nukulaelae Tuvaluan letters (see also Biber 1995, Yates 1993) that they “include phatic communion” and are “heavily affective” toward the addressee (Besnier 1988: 714). He therefore criticizes linguists who regard writing as a medium in which emotional content and self-expressions are minimized (Besnier 1988). Biber’s (1995: 175) multidimensional analysis of Besnier’s letters positions them beyond Nukulaelae Tuvaluan conversations as regards interpersonal reference, noting that they “make frequent reference to the author (‘I’) and receiver (‘you’), even though [the] direct interaction through [the] letters is extended over long periods of time” (Biber 1995: 174). Biber’s (1988, 1995) own collection of personal letters in English does not assume a position beyond English face-to-face or telephone conversations as regards involved production (on his Dimension 1 distinguishing between informational and involved production), but a position second only to conversations, beyond all other genres of writing and speech (1988: 128; see also section 5.2.1 here). First and second person pronouns constitute 61 percent of the total personal pronoun use in Biber’s personal letters (1988: 262). The letters contain 62.0 1PP and 20.2 2PP per thousand words, respectively. In Collot’s ACMC “ELC other” corpus there are 57.8 1PP and 17.6 2PP per thousand words, respectively; see figure 4.2. There is thus a close affinity between personal letters and ACMC, as well as between ACMC and the average for speech, considering their nearly identical overall personal pronoun use; see figure 4.2: 102.3 and 105.0 pronouns per thousand words in respective medium. The spoken genres most akin to ACMC, as regards personal pronominal reference, are spontaneous speeches and interviews (Biber 1988: 268, 266, also noted by Collot 1991 with regard to personal reference as well as to other features).

Turning now to the intermediate bar in figures 4.2 and 4.3 – speech – a few remarks are called for. Numerous linguistic authorities, such as Chafe (1982),

Chafe & Danielewicz (1987), Wales (1996), Biber (1988, 1995) and Biber et al. (1999), have drawn attention to the overall high numbers of personal pronouns observed in speech, as opposed to their numbers in writing (with the exception of personal/informal letters). Explaining message structure in English, Halliday (1985a) declares, in functional grammatical terms, that the Theme in spoken language, “the peg on which the message is to hang,” is often a pronoun, “most typically *I* or *you*” (1985a: 73, original italics). Halliday (2004) expounds:

In everyday conversation the item most often functioning as unmarked Theme (Subject/Theme) in a declarative clause is the first person pronoun *I*. Much of our talk consists of messages concerned with ourselves, and especially with what we think and feel. Next after that come the other personal pronouns *you*, *we*, *he*, *she*, *it*, *they*; and the impersonal pronouns *it* and *there*. (Halliday 2004: 73, original italics)

Wales (1996) notes that the first personal singular pronoun (*I*) occurs most frequently in speech, and that it is the second most common word in the spoken part of the British National Corpus, second only to *the* (1996: 68). Among the spoken genres, Biber (1988) finds personal pronouns most common in telephone conversations (totaling 126.7 per thousand words; 1988: 265), closely followed by face-to-face conversations (totaling 117.9 per thousand words; 1988: 264). Biber et al. (1999) explain that first and second person pronouns, referring to the speaker and the addressee, are “naturally very common in conversation because both participants are in immediate contact, and the interaction typically focuses on matters of immediate concern” (1999: 333). None of the linguists mentioned, however, has investigated conversational writing, such as IRC and split-window ICQ chat. That personal pronouns are “by far most common in conversation” (c. 135 per thousand words in the LSWE corpus; Biber et al. 1999: 333) is a statement that can be qualified. Not only does Biber (1988) find them equally common in personal letters (135.0 per thousand words; 1988: 262), but the investigation of personal pronouns, in the present section, has proved that they are even more common in supersynchronous conversational writing (157.5 per thousand words). Moreover, it is in the conversational writing genres that the ratios of first and second person pronouns to all personal pronouns are the highest. Chafe’s initially striking finding of 61.5 first person pronouns per thousand words in spoken discourse, with which he introduces the concept of “involvement” (Chafe 1982: 46), pales by comparison with the finding of 88.9 first person pronouns in split-window ICQ chat in the present study; see figure 4.2. “Involvement,” instantiated through first person reference, thus epitomizes the character of supersynchronous conversational writing more than it does the character of any other genre.

4.3 Word length, type/token ratio and lexical density

In Biber's (1988) multidimensional study of written and spoken language, "word length" and "type/token ratio" (the ratio between the number of different words, "types," and the total number of words, "tokens,"⁵⁸ per text) are the two features intended to measure the lexical specificity and diversity of texts. They are powerful tools in the study, as differences in lexical specificity and diversity truly are found to correlate with production differences between writing and speaking. Longer words have been found to convey "more specific, specialized meanings than shorter ones" and words tend to "become shorter as they are more frequently used and more general in meaning" (Biber 1988: 238, referring to Zipf 1949). Zipf (1949: 65) finds an "inverse relationship between the lengths of words and their frequency" in language, not just in English, but in several other languages (including Peipingese Chinese, two American Indian languages and the main Western European languages), i.e. that the short words in these languages tend to recur. Zipf (1949), Drieman (1962), DeVito (1965) and Gibson et al. (1966) all consider measures of word length in their studies (as seen in chapter 2), finding longer words more frequent in writing than in speech. The latter three studies, furthermore, employ the measurement of type/token ratio, henceforth TTR, in distinguishing between written and spoken texts, finding higher TTR values in writing. In the present section, the lexical properties of conversational writing are explored, facilitated by the measurements of word length and TTR, as well as by the more revealing measurements for our purposes, those of lexical density.

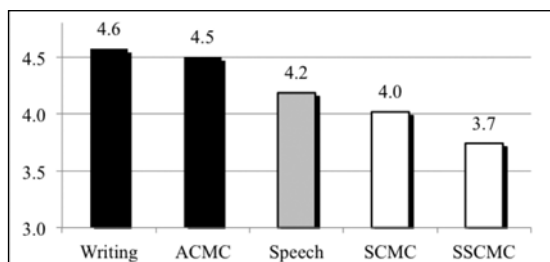
Drieman (1962) and Gibson et al. (1966), by measuring numbers of syllables, and Blankenship (1974), by measuring word length per se (most likely by characters), all find word length to be a distinguishing factor between writing and speech, observing shorter words in speech. The difference in word length is attributed to the different production circumstances of writing and speech, less encoding time in speech and the consequent need for the speaker to select "easy, short, and familiar" words (DeVito 1970: 11). Longer words usually entail higher levels of lexical specificity, and are typically produced under circumstances that permit editing and longer contemplation. When writing, "[w]e can take hours,

58 "Token" by default denotes a string of orthographic keystrokes set apart from other strings by a blank space or a new line. Tokens in conversational writing mostly constitute words semantically, but also for instance initialisms (e.g. *hb* meaning "hurry back"), and emoticons (e.g. :)), not traditionally referred to as words. (Accordingly, even in other sections of the present study, the term "token" is occasionally preferred over "word" when discussing the data.)

if we need to, to find an appropriate word,” say Chafe & Danielewicz (1987: 88), and “...we are free to revise [the words] again and again until they satisfy us” (ibid.). Halliday (1985a) prudently establishes that the distinction between long and short words in reality reflects the continuum from lexis into grammar. The distinction is simply embodied in the spelling system: lexical items typically require a minimum of three letters, whereas grammatical items may comprise only one or two letters. Halliday incidentally points out that most prepositions belong in the grammatical class, “because of words like *at, in, to, on*, which otherwise would have to be spelt *att, inn, too, onn*” (1985a: 63, original italics). The distinction between long (mostly lexical) and short (mostly grammatical) words can thus be seen as fundamental to the difference between writing and speech.

Figure 4.4 indicates the average word length of texts in the five media: writing, asynchronous CMC, speech, synchronous and supersynchronous CMC, respectively. The figure shows a neatly declining scale of word length, from 4.6 orthographic letters per word in writing to 3.7 in split-window ICQ chats. In figure 4.4, as well as in all subsequent diagrams, the written genres are represented in black, spoken genres in gray and conversational writing genres in white. For the p-values from statistical tests of findings in SCMC and SSCMC, as compared to writing and speech, see Appendix VI.

Figure 4.4: Average word length in the five media.



Word length entails “the mean length of the words in a text, in orthographic letters” (Biber 1988: 239). In conversational writing, this is indicated as the number of orthographic keystrokes found between blanks, after texts were purged of all regular punctuation, except apostrophes within words, emoticons and simple imagery.⁵⁹ Example (6) is a part of a text ready for the word-length count, a text

59 See chapter 3 for a description of the purging and adaptation procedure, and section 4.5 for examples of retained imagery.

which exemplifies a few remaining, albeit rare, instances of such imagery (:0), i.e. a smiley, and <===== (=0, a sword). The example illustrates the irregular length of tokens typically found in chats (repeated xxxxxx... etc., meaning “kisses” vs. *c u*, meaning “see you”). To avoid skewing the word length results, extremely long tokens were truncated at 50 keystrokes; five such long tokens existed in the IRC component, and two in the split-window ICQ component.

(6) i dont know who he really is
 yeah women!
 lol
 true
 be careful
 that i am
 hi all
 <===== (=0
 <===== (=0
 any girl wanna chat?
 <===== (=0
 nice sword
 lol
 u have been practising a lot
 he has
 now he is ready
 saba 20 where are you
 alot of work put into that piece of artwork
 to impress the ladies
 lol
 i will be back
 ok
 take care
 gotta go for 5 minutes
 u 2 max sweety
 c ya in a sec u hunk of spunk
 hehehe
 c u
 cya
 xxx
 :0)
 hi

Internet relay chat text 3a (UCOW)

For maximum economy of typing, either for minimum effort or minimum production time, or both, or for mere adherence to genre conventions, IRC

interlocutors abbreviate and contract words and expressions in various ways (e.g. *lol, u, wanna, alot, u 2, c ya, sec, c u, cya* in example (6)). Such abbreviation schemes naturally render short words pervasive. On the other hand, resting a finger on a key for an entire turn, as in *xxxxxx...* etc. in (6), and posting precomposed imagery, such as the sword in (6), are also devices available to chatters, devices which increase the average word length. Nevertheless, from figure 4.4 it is evident that IRC chatters on average operate with shorter words than do speakers. For one thing, speakers cannot abbreviate words, e.g. *see, you*, into their corresponding homophonous letters, *c, u* (or rather, transcribers of speech do not).

Comparing the average *turn* length of IRC (4.3 tokens/turn) with split-window ICQ (7.0 tokens/turn), in connection with the average word length displayed for these genres in figure 4.4, gives the impression that, in conversational writing, longer turns entail shorter words. As seen, ICQ chatters indeed employ shorter words than IRC chatters; the average word length in split-window ICQ is only 3.7 orthographic keystrokes. On the other hand, as seen in example (6), a great number of IRC turns consist of very short messages, e.g. greetings (*hi, cya*), with very short word length. Also, “turn length” in split-window ICQ is a rather artificial concept as it is determined by the logging feature of the software, more than by the actual user. The cut-off point between turns is not always clear-cut in the supersynchronous chats, where simultaneous typing frequently occurs and where users do not hit enter to post their turn. For this reason, turn length is not a reliable construct for comparisons of word length. The results in figure 4.4, nevertheless, underscore that split-window ICQ chatters operate with shorter words than IRC chatters. Example (7), from the split-window ICQ word length count, shows that the ICQ chatters, for the economy of typing, use similar abbreviations as the users in example (6) (*r, u*), as well as apostrophe-less contractions (*thats, im, its, wasnt, didnt*), although the abbreviations are less frequent in split-window ICQ than in the IRC chats.

- (7) what r u doing this weekend
i'm going to be sitting at home watching love movies by myself :(
awwwwwwww
thats cute
im not sure what im doing but it will probably be just as boring
yea its going to suck
well at least you're allowed to go out and stuff
ya thats true
no offence u didn't have to get a speeding ticket
yea thanx

well yea but i wasnt even doing 75 but i just said i did because i didnt want to
fight the case
why u could have won
yea but that means iw ould have to miss a couple days of school just to go to court
Split-window ICQ chat text 1 (UCOW)

The subject matter discussed, of course, could be a factor influencing word length. The topics in the split-window ICQ chats are more tangible and diverse, and the discussions more vivid, than in IRC. On the other hand, both ICQ and IRC chatting are leisure-time activities for casual social interaction, and neither communication requires well-reasoned exposition or highly explicit lexical choices from users. The short word length of conversational writing therefore, more than anything, seems to be determined by the same factor that renders short words in speaking, briefly considered in the beginning of this section: their stronger affiliation with the grammatical rather than lexical classes of words. The lexical density of conversational writing will be further investigated below, but first we must briefly touch upon the classic measurement of TTR.

The type/token-ratio measure is regarded as a useful tool for exploring the vocabulary variety of a given text. To arrive at the TTR, the number of different words (“types”) in a text is divided by the number of words (“tokens”) in that text. Consider the split-window ICQ example below, from example (7), which serves to illustrate the procedure:

“well yea **but i** wasnt even doing 75 **but i** just said i did because i didnt want to fight the case”

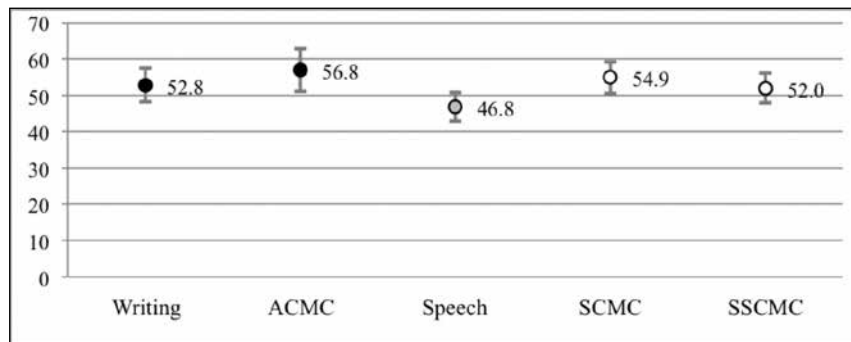
The example contains 22 words (tokens), but there are only 18 different words, as *i* is used four times and *but* twice. The type/token ratio of this sentence is consequently $18/22$, i.e. 0.818, or by convention, expressed as a percentage, 81.8. In order for the TTR to reliably represent the diversity of a text, however, samples must be of substantial length, though not too long as the relation of types to tokens is not linear. Biber (1988) finds the ideal sample size for measuring TTR to be 400 words. In Biber (1988), the ratio is computed “by counting the number of different lexical items that occur in the first 400 words of a text, and then dividing by four” (1988: 238, as explained in section 3.2). All texts in the five media to be compared here have undergone this computation method for TTR, and the results are shown in table 4.3 and figure 4.5, along with the standard deviations among texts.⁶⁰

60 See section 3.6 for an explanation of the procedure for calculating the TTR standard deviation of the texts in the media writing and speech.

Table 4.3: Type/token ratio, with standard deviation

	Writing	ACMC	Speech	SCMC	SSCMC
type/token ratio	52.8	56.8	46.8	54.9	52.0
standard deviation	4.7	5.9	3.9	4.4	4.1

Figure 4.5: Type/token ratio, with standard deviation.



As mentioned in section 2.2 and in the beginning of this section, linguists have consistently found higher TTRs in writing than in speech (Drieman 1962, Gibson et al. 1966, Blankenship 1974, Chafe & Danielewicz 1987, Biber 1988, 1999). Chafe & Danielewicz (1987) explain:

[S]peakers tend to operate with a narrower range of lexical choices than writers. Producing language on the fly, they hardly have time to sift through all the possible choices they might make, and may typically settle on the first words that occur to them. The result is that the vocabulary of spoken language is more limited in variety. (Chafe & Danielewicz 1987: 88)

The linguists mentioned above have all dealt with texts that that have been ideally suited to represent their respective media in written format. The written texts mostly derive from published sources and have thereby undergone careful editorial scrutiny, which has rendered misspellings and other irregularities extremely rare in them. The spoken texts for the most part have been transcribed by linguists, who have devoted considerable time and effort to correctly representing speech, with detailed attention to spelling, regularity and consistency. Writing and speech are consequently reliably represented as regards vocabulary variety in figure 4.5.

The texts of CMC are of a different kind. None of them has undergone careful editorial scrutiny or been transcribed by linguists. Instead, they are taken

straight from their respective media and represent authentic user-generated material. ACMC texts may well contain carefully prepared exposition. Yates (1993) notes for his corpus that the ACMC medium does provide the opportunities for redrafting that according to Chafe & Danielewicz (1987) bring about a greater vocabulary in written texts, but notes that these opportunities may not be taken by all CMC users. Yates observes a TTR for ACMC which, like Collof's (1991) ACMC plot in figure 4.5, is closer to writing than to speech. With regard to the synchronously and supersynchronously mediated texts, however, the TTR representation in figure 4.5 is more problematic. To simplify, here is a speculative, but viable, analogy: if the conversational writing texts were writing, they would be a very first draft, produced under severe time constraints, with no chance for editorial or self-revision; if they were speech, and therefore transcribed by linguists, they would be transformed into a format as regular and consistent as spoken texts, and likely attain a TTR similar to speech, or even face-to-face and telephone conversations.⁶¹ The TTR of the SBC subset (face-to-face conversations), for instance, is 44.2.

Speculation aside, why are the TTRs for conversational writing so high? Texts with high TTR display a great number of types, i.e. different words. The heterogeneity of words in conversational writing is immediately noticeable upon studying the texts. Firstly, besides abbreviations, emoticons and certain imagery, the texts bristle with other irregularities: misspellings (*sence* for *sense*), slips of keys (*wel ive* for “we live”), missed keystrokes (*jus* for “just”), contractions with omitted apostrophes (*dont, im, thats, shes*), letters repeated for effect (*desire for fooooooooooooooooood*), graphic re-representation of letters (*\Velcome*), simplified, phonological spelling (*proolly* for “probably,” *sleepin, cuz* for “because,” *kinda* for “kind of”) and multitudes of renderings of one and the same lexeme (*u, ya, yu, you, yah, yôÛ, jÛÛ, Yôù*, to mention but a few representations of the second person pronoun). Secondly, chatters in IRC, unlike speakers in most spoken conversations, repeatedly address each other by nicknames to designate the recipient of an utterance. Greetings, for instance, are frequently followed by nicknames, which serve to designate the recipient as well as to signal that the new user's presence has been noticed (Werry 1996, Anglemark 2009). Nicknames serving as address terms also facilitate the untangling of threads in the communication.

61 Despite the feasibility of such a task, it goes against the grain for the present researcher to transcribe or manipulate this kind of unique user-generated conversational writing data to attain more comparable figures. Worthy of notice, however, is Ko's (1996) study, which arrives at a TTR of only 33.7 for classroom setting SCMC with seemingly regular and consistent user-generated orthography (rating from examples given).

“Such a high degree of addressivity is imperative on IRC, since the addressee’s attention must be recaptured anew with each utterance,” says Werry (1996: 52). The designatory nicknames add to the number of types in a text, especially since they are frequently changed and as users tend to invent their own pet names out of them (e.g. }melons{{ is addressed *melons*, *mels*, Rich23 *rich*, *rick*, [mad_max|mad_max, mad_max, etc., each variant counted as a separate type in the TTR calculation). Thirdly, chatters frequently emulate spoken communication to add emphasis to utterances. In a spoken language corpus, an “utterance” like laughter follows the same regularized transcription convention throughout the corpus, whereas in conversational writing users invent their own “transcriptions” ad hoc (HAHAHAHAHAHAh, ha ha, haha, hahahahahahaha) – in which, as regards TTR, a thirteen-character laughter token counts as a different “type” than a sixteen-character one. This equally applies to chatters’ alternative transcriptions of word stress (*yeesssssss*, *this suxxxxxxxxxxxxxxx*, *yummm....*). Moreover, several of the emoticons may contain repetition of a character for emphasis (whereby :) counts as one “type,” :) as another, etc.). All in all, this user-generated orthographic heterogeneity results in a multitude of types in the type/token-calculation, rendering inordinately high ratios for the conversational writing genres (also noted by e.g. Freiermuth 2003, Forsyth 2007, Forsyth & Martell 2007, who similarly discover type/token ratios in SCMC that are closer to writing than to speech, and who explain this inter alia by the variable spelling and nickname usage).⁶² The representation of conversational writing in figure 4.5 must, therefore, be taken with a grain of salt, and we must find alternative ways to approach and explore the lexical complexity, or lack of complexity, in conversational writing.

This is when we turn to the two more revealing methods for measuring the lexical properties of conversational writing: the measures of *lexical density* and *lexical density per clause* (cf. Ure 1971, Halliday 1985a, 1987, 2004). Unlike TTR, these measures distinguish between lexically complex (“most likely to be written”) and grammatically complex (“most likely to be spoken”) texts (Halliday 1987: 59). While the TTR measure, rather mechanically, indicates the ratio of new types among the tokens, the lexical density measures take into account the lexical properties of the words. Moreover, the lexical density measures are not

62 By contrast, neither Collot (1991) nor Yates (1993) problematizes the high TTR of their ACMC texts as being the result of irregular spelling or other orthographic anomalies. Judging from corpus examples in both computer conferencing studies, participants’ spelling is consistent and appears to follow the norms of writing. The TTR for Collot’s (1991) ACMC corpus in figure 4.5 thus justifiably indicates a vocabulary variety in asynchronous computer-mediated texts above that of writing.

sensitive to text length (Yates 1993). In the discussion of word length above, Halliday (1985a) was shown to have drawn attention to the distinction between lexical and grammatical items in discourse. The short average word length in conversational writing was suspected to be due to there being more grammatical than lexical words in the discourse. It is now time to find out whether this is the case.

The lexical density of a text is the proportion of lexical items (content words) to the total discourse (Halliday 1985a, 1987). It can be measured in at least two ways: the ratio of lexical items to the total number of running words in a text, or to the total number of clauses, with or without weighting for relative frequency in the language.⁶³ In our consideration of the lexical density of conversational writing, no weighting will be employed. To understand the measurement, consider Halliday's (1985a: 61) classic example, which contrasts a written sentence with its "translation" into a likely spoken equivalent:

Investment in a rail facility implies a long term commitment (L:7; G:3)

If you invest in a rail facility, this implies that you are going to be committed for a long term (L:7; G:13)

The first of these sentences (more typical of writing) contains a ratio of seven lexical items (L:7) to three grammatical (G:3), the lexical items being *Investment*, *rail*, *facility*, *implies*, *long*, *term* and *commitment*. A ratio of seven lexical items to a total of ten words yields a lexical density of 7/10, i.e. a lexical density of 70%. The second sentence (more typical of speech) contains more grammatical items and therefore yields a lexical density of 7/20, i.e. 35%. Relative to each other, written language is lexically dense and spoken language is lexically sparse, or put differently: spoken language is grammatically dense; it displays "grammatical intricacy" (Halliday 1985a: 87, 1987: 62ff, 2004: 655).

To calculate the lexical density of a text, all orthographic items (tokens) must first be identified as either belonging to the closed sets of grammatical items, or to the open-ended classes of lexical items – a fairly cumbersome but, as we shall see, worthwhile task. Halliday (1985a: 61) identifies the grammatical items in English to be "determiners, pronouns, most prepositions, conjunctions, some classes of adverb, and finite verbs." He goes on to give a number of example sentences indicating finite full verbs, such as the third person present tense verb *implies* in the above example, as lexical items. In light of the examples, his

63 In weighted lexical density calculations, low-frequency lexical items are given a higher "score" (or "weight") than high-frequency ones. In unweighted lexical density calculations, all items are treated alike (Halliday 1985a, Yates 1993).

definition of “finite verbs” as being grammatical items must therefore be re-interpreted as “auxiliary verbs.” Furthermore, his example sentences indicate all forms of the verbs *be*, *have* and *do* as grammatical items. In the present study, lexical items were consequently taken to be all non-auxiliary, i.e. full verbs (except *be*, *have*, *do*), as well as nouns (including nominalizations, nominal gerunds and proper nouns),⁶⁴ adverbs (except discourse particles, adverbs *all*, *as*, *here*, *how*, *then*, *there*, *when*, *where*, *why*, *anywhere*, *everywhere*, *nowhere*, *somewhere*, *so*, synthetic negation *no*, *neither*, *nor*, analytic negation *not*) and adjectives, in agreement with the examples given in Halliday (1985a: 61–62).⁶⁵ This means that lexical items were found among Biber’s (1988) features (full verbs among e.g. features 1–3, 17, 18, 24–26, 55–58; nouns among features 14–16; adverbs among e.g. features 4, 5, 42, 45–49, and adjectives among features 40–41; see table 2.1 for the numbered features), but also had to be found outside of this list of features, as for instance the main verbs of progressive verb phrases are not identified by it. The identification of lexical items therefore required a separate round of annotation, beyond the annotation of Biber’s features.

With the identification of the lexical items completed, the calculation of lexical density for each corpus was fairly straightforward. As mentioned, the lexical density measurement simply indicates the ratio (percentage) of lexical items to the total number of running words. The results are presented in table 4.4, along with the lexical densities calculated by Yates (1993) for LOB, to represent writing, and for LLC, to represent speech. No lexical density was calculated for ACMC in Collof (1991); therefore, to represent ACMC in table 4.4 is the figure for Yates’ (1993) computer conferencing corpus. The results are not presented graphically here since previous graphs, and graphs to come, indicate figures for Collof’s (1991) ACMC corpus and incorporating Yates’ (1993) ACMC figure for only this feature would interrupt the consistency across graphs.⁶⁶

64 Nicknames used as address terms (very common in IRC) were not included in the count for proper nouns, to avoid skewing the data, but nicknames used about a third person were included, as well as all other proper nouns.

65 Numerals, infinitive markers, inserts (except e.g. *Shit*, *God*; see table 4.9) and emotives were considered to be “grammatical” words.

66 Note that Yates’ (1993) lexical density figure for ACMC is only indicative here, as it might be that Yates’ computer conferencing corpus deviates lexico-grammatically from Collof’s (1991) corpus of BBS communication. As Yates’ ACMC corpus texts are unavailable, the lexical density of ACMC will not be discussed further here, apart from concisely corroborating, in this footnote, Yates’ (1993: 94) conclusion that ACMC and writing are close on this measure.

Table 4.4: *Unweighted lexical density for five corpora (LOB writing, ACMC and LLC speech from Yates 1993)*

Unweighted lexical density	
LOB writing	50.3
ACMC	49.3
LLC speech	42.3
Face-to-face SBC subset	36.6
SCMC	38.7
SSCMC	39.6

Judging from table 4.4, conversational writing ranks lower than LLC speech, but higher than face-to-face conversations from the SBC subset, as regards lexical density. Ure (1972) conducted a study of the lexical density of 30 written and 34 spoken texts, finding most written texts to have a lexical density of over 40% and most spoken under 40%.⁶⁷ Halliday's (1985a) example sentences contrasting written and spoken versions of the same messages display lexical densities above 45% for the written, and below 45% for the spoken versions. Halliday (1987), moreover, experiments with a passage of formal written English, rewording it in two steps into a "less written" and a "more spoken" version and finds the lexical density to dramatically decrease with increased "spokenness." His formal "written" version has a lexical density of 55%, his "less written" 47% and his "more spoken" version 39%. Even though no explicit dividing line is drawn in Halliday (1985a, 1987), one around 45% seems relevant. Stubbs (1996) finds a large overlap in lexical density among the genres of writing from LOB (with a range of 40 to 65 percent) and those of speech from LLC (with a range of 34 to 58 percent), and therefore no absolute difference between writing and speech, but establishes that the lexical density measurement is a "robust method of distinguishing genres" (1996: 76). In the LSWE corpus, Biber et al. (1999: 61) find "conversations" to have the lowest (41%) and "news" the highest lexical density (63%). Bringing the implicit dividing lines of these studies to bear on the results in table 4.4, we find conversational writing well settled on the spoken side of the continuum.

Interestingly, however, LLC speech and face-to-face conversations (SBC subset) diverge from each other slightly. As mentioned in section 3.4, LLC contains spoken texts of not just dialogs (e.g. face-to-face and telephone conversations),

67 Ure (1971) gives no account of what word classes were included among those with "lexical properties."

but also spoken texts of monologic character (e.g. broadcasts and speeches). In a discussion of LLC's genres, Stubbs (1996) discloses what might be suspected here, namely that the more monologic genres slightly boost the lexical density for LLC overall. Ure (1971) likewise notes higher lexical density among prepared than among unprepared spoken texts. More importantly, Ure (1971) makes a penetrating remark regarding texts with low lexical densities. She finds spoken texts with the lowest lexical densities to exclusively derive from sources where there is verbal response to the speaker, or some perceptible nonverbal response that would make the speaker adjust their language. This kind of response, known as *feedback*, she identifies as "an even more powerful factor in determining lexical density than the spoken/written choice" (1971: 448).

That feedback contributes to lower lexical density is borne out, in table 4.4, also in that conversational writing approximates the SBC subset face-to-face conversations more than does speech overall, or writing for that matter. Whereas the written genres of LOB contain monidirectional texts, the texts of the conversational writing genres, just like face-to-face conversations, are by default bidirectional. Besides feedback, Ure (1971) considers the influence of personal and social relations to have a bearing on lexical density, arguing that when impersonal texts coincide with those without feedback, the lexical density is increased. The face-to-face conversations from the SBC subset and the conversational writing texts, in table 4.4, all contain personal communication, some between previous acquaintances ("familiar" as opposed to "distant" relations, in Ure's 1971: 449 terms), which implies that their lexical density is loosened up.

What features in the face-to-face conversations from the SBC subset, then, account for giving the genre a lexical density below conversational writing? The answer lies not with the lexical items, but rather with a few of the grammatical ones, among which four stand out: face-to-face conversations contain more third person pronouns (as seen in section 4.2), more prepositions (to be explored in section 4.4), more of the impersonal pronoun *it*, and slightly more discourse particles, than does conversational writing. Example (8) from SBC serves to illustrate the abundance of grammatical items in face-to-face conversations, a discussion among friends cooking a meal together.

- (8) Roy: I could eat one of those.
Marilyn: You could?
Pete: Hm.
Roy: Well,
but I won't.
Pete: Then I guess
Roy: I mean,

Marilyn: Okay.
 Pete: Divide it in half.
 Roy: well don't
 Marilyn: Then I'll
 Roy: Y-
 What you oughta do though Mar,
 cook all the fish.
 Marilyn: Hm.
 Roy: Cause
 well,
 we won't use it,
 if you don't cook it.
 Now.
 Marilyn: Well I was gonna make ceviche with the leftovers.
 Roy: Oh alright,
 that sounds good.

Face-to-face conversations SBC text 3

Example (8) contains 65 words, only 17 of which are lexical items (*eat, guess, mean, divide, half, though, Mar, cook, fish, use, cook, make, ceviche, leftovers, alright, sounds, good*), yielding a lexical density of merely 26.2 for the passage. Among the grammatical words, there are three prepositions (*of, in, with*), three pronouns *it*, and as many as five discourse particles (*Well, well, Now*).

Pronoun *it* is often used as prop-*it*⁶⁸ in oral conversations but also substitutes for a range of referents, for “nouns, phrases, or whole clauses” (Biber 1988: 226). Referents in conversations are frequently tangible objects, as the fish in example (8). Chafe & Danielewicz (1987) explain the high frequency of pronoun *it* in spoken conversations thus:

Speakers not only have less time to choose vocabulary, but they also cannot or do not take the time to be as explicit about what they are referring to. A symptom of this kind of vagueness is the use of third person neuter pronouns, usually *it, this, or that*. Typically, the antecedent of a pronoun has been spelled out in an earlier noun phrase. (Chafe & Danielewicz 1987: 90, original italics)

Chafe & Danielewicz note that in conversations the antecedent is typically spelled out first, and then referred to by inference from the textual or situational

68 Prop-*it* is a dummy pronoun used as “‘empty’ or ‘prop’ subject, especially in expressions denoting time, distance, or atmospheric conditions” (Quirk et al. 1985: 348), e.g. “What time is *it*? *It*’s half past five,” but also for instance as “nonreferring” *it* with “vague implications of ‘life in general’, etc.,” e.g. “How’s *it* going?” (Quirk et al. 1985: 349, original italics).

context. In conversational writing, objects are rarely shared, or tangible, and deictic *it* therefore emerges less frequently than in face-to-face conversations. Instead, chatters by necessity refer to objects as nouns, i.e. as lexical items, which in turn contribute to the slightly higher lexical density figure for conversational writing.

The discourse particles annotated in the present study are *well*, *now*, *anyway*, *anyhow* and *anyways* (Biber 1988: 241), the first one by far the most frequent. Discourse particles are used to maintain conversational coherence (Biber 1988, Aijmer 2002). *Well* helps speakers in involved discourse monitor the information flow to the listener, and to ascertain that the communication is functioning smoothly (Chafe 1985, Schiffrin 1985); *now* is closely related to *well*, but also has a discourse-organizing function (Aijmer 2002: 57) and *now*, *anyway*, *anyhow* and *anyways* also function as emphatic topic changers. The moderate incidence of discourse particles in conversational writing implies that users find other ways to monitor the information flow, and to introduce new topics (ways expounded in Zitzen 2004). Another likely reason for their relative rarity (3.3 in IRC and 4.9 in ICQ, compared to 7.7 per thousand words in the SBC subset) is that the behavior of many users is governed by economy of typing. In spoken conversations, *well* frequently occurs in brief sequences of overlapping speech, when both speakers attempt to make their voices heard. Conversational writers do not encounter such situations, as every typed word is assumed to be read, and, consequently, for economy of typing users more often leave out discourse particles and cut straight to their message. Interestingly, however, Ko (1996) finds more discourse particles in his chat corpus than in face-to-face conversations, and attributes this to chatters' "increased need to monitor the flow of information in a situational context where there are multiple participants and no simultaneous feedback cues available to show listenership" (1996: no page number available). In the conversational writing corpora here, discourse particles are about as common as in the medium of speech overall (i.e. in Biber' spoken genres + the SBC subset, which together contain on average 4.2 discourse particles per thousand words).

The messages in IRC, i.e. the *turns*, contain only 4.3 tokens on average, while the turns in the annotated SBC subset contain 8.1 words on average (no equivalent figure for LLC speech was computed or found in the literature). Split-window ICQ turns, with 7.0 tokens on average, also appear to be shorter than in speech. As mentioned, however, comparing the turn length of split-window ICQ with those of IRC or speech is not practicable as, in split-window ICQ, turns are determined by the logging feature of the software, more than by the users. Consequently, for the analysis of textual complexity here, the *turn* is not an altogether reliable

construct. All the same, it must be recognized that the perceived complexity of texts relies not only on the overall lexical density of texts, how closely packaged the information is in general terms. The perceived complexity also depends on the packaging of the information into the constituent grammatical structures of the text. Halliday (1985a) identifies the most relevant of these structures to be the *clause*; “The clause is the grammatical unit in which semantic constructs of different kinds are brought together and integrated into a whole” (Halliday 1985a: 66). The clause is also seen as the most reliable construct upon which to carry out comparative investigations into the genre variation of language. For comparative purposes, the main requirement is consistency, and the clause is recognized as “perhaps the most fundamental category in the whole of linguistics,” as well as “critical to the unity of spoken and written language” (1985a: 67). Therefore, to relate the perceived complexity of texts to the discrepancy of clauses in spoken and written discourse, Halliday introduces the next measure to be considered, lexical density per clause.

The perceived complexity of a text depends not just on the lexical density overall, but also upon the composition of the text’s clauses, especially the length of clauses. The average clause in the annotated face-to-face conversations from SBC is 5.7 words long; in IRC it is only 3.9 and in split-window ICQ 4.6. “Lexical density per clause” indicates the number of lexical items per clause.⁶⁹ Consider again the ICQ turn from example (7), which will help to explain the calculation procedure:

“well yea but i wasnt even doing 75 | but i just said | i did | because i didnt want to fight the case”

The number of lexical items in the above turn is six (*even, just, said, want, fight, case*) and the turn consists of four non-embedded clauses (separated by vertical lines).⁷⁰ There are consequently, on average, 6/4, that is, 1.5 lexical items per clause, in this turn. Now contrast the chatted turn with a sentence from the biography genre of LOB:

69 In this study, Halliday’s (1985a) definition of the clause was observed, i.e. both finite and non-finite clauses were counted, whether independent (in “parataxis”) or dependent (in “hypotaxis”), but not restrictive relative clauses (which Halliday 1985a: 84 calls “embedded”). For further description of what constitutes a clause; see Halliday (1985a: 67ff).

70 In the example, “i did” is an instance of hypotaxis, but not “embedding,” in Halliday’s (1985a: 83) terms.

- (9) The story of the resplendent premiere, the gradual disintegration and eventual catastrophic debacle of this first French production of Don Giovanni can be followed in detail through the reviews in the contemporary press.

Biographies LOB G: text 44

The sentence from LOB contains 18 lexical items in one single clause, yielding an extremely high lexical density per clause: 18.0. Stubbs (1996: 75) finds the particular text from LOB (G: text 44, i.e. the full text) to have among the highest lexical densities of the written texts (58%), but does not carry the investigation further to the clausal level. Halliday (2004), however, explains that the complexity of spoken and written language is two-fold right down to the clausal level: the complexity of spoken language is grammatical, while that of written language is lexical. He describes the different complexities thus:

In spoken language, the ideational content is loosely strung out, but in clausal patterns that can become highly intricate in movement: the complexity is dynamic – we might think of it in choreographic terms. In written language, the clausal patterns are typically rather simple; but the ideational content is densely packed into nominal constructions: here the complexity is more static – perhaps crystalline. (Halliday 2004: 656)

Spoken language becomes complex by being grammatically intricate. Just as in spoken conversations, the ideational content of the split-window ICQ-turn above is “loosely strung out” (cf. Halliday 2004: 656), but the chatter “builds up elaborate clause complexes out of parataxis and hypotaxis” (cf. 2004: 654) (e.g. paratactical *but i just said* and hypotactical *i did* in the turn from example (7)). Written language, on the contrary, typically “becomes complex by being lexically dense: it packs a large number of lexical items into each clause” (2004: 654), even though the clausal pattern overall is rather simple (e.g. only one verb, *can be followed*, in example (9)). Halliday notes that the total number of lexical items in written texts usually just have “fewer clauses to accommodate them” (2004: 655). What typically happens in writing is that the lexical items are incorporated into nominal groups, as in example (9) (e.g. the long subject *The story of the resplendent premiere... Giovanni*). The nominal group is grammar’s primary resource for “packing in lexical items at high density” (Halliday 2004: 655).

Halliday (1985a), however, admits that the term “lexical density” is semantically loaded and repeatedly cautions against thinking of written texts as more complex: the measurement equally could have looked at the same phenomenon from the grammatical end; “[w]e could [say] that the difference between spoken language and written language is one of [grammatical] intricacy, the intricacy with which [spoken] information is organised” (1985a: 62). Halliday (1985a, 1987,

2004) therefore consistently calls spoken language more intricate than written. While spoken language represents phenomena as “processes,” written language represents phenomena as “products” (Halliday 1985a: 81); complex relationships are expressed “clausally” in spoken language and “nominally” in written language (Halliday 2004: 655; see also Castello 2008). Both kinds of complexity, nevertheless, can be accounted for under a single generalization, the notion of lexical density, which measures the different kinds of complexity, grammatical and lexical, that arise “in the deployment of words” (Halliday 1985a: 63). With that, we now turn to the measurement of lexical density within clauses with regard to the corpora annotated in the present study.

In the discussion of lexical density per clause here, only the results for the corpora annotated in the present study are tabulated, as no comparable average results were found for LOB, ACMC or LLC.⁷¹ As the numbers of lexical items had been identified already in the general lexical density calculation, the calculation of the new measure merely required the identification of the total number of clauses in each corpus. The total number of lexical items was then divided by total number of clauses for each corpus. Recall from the discussion of average turn length, in connection with example (6) above, that IRC turns are frequently very short (occasionally consisting of no more than a token, e.g. turns *true*, *ok*, *:0*, *hi*). Similar short turns are found in the SBC subset face-to-face conversations (*So*, *Well*, *No*, *Yeah*), and, although to a lesser extent, in SSCMC (*awwwwwwwww* in split-window ICQ example (7)). No matter how short, a turn was always counted as, at least, one clause. The resulting average numbers of lexical items per clause are presented in table 4.5. As mentioned, this measure is known as “lexical density per clause” (Halliday 1985a, 2004), and it is found in the first column of the table.

Table 4.5: *Unweighted lexical density per clause and related measures*

	Lexical density per clause	Proportion of lexical items per clause	Average clause length
Face-to-face SBC subset	2.1	36.6%	5.7
SCMC	1.5	38.7%	3.9
SSCMC	1.8	39.6%	4.6

71 The calculation of lexical density per clause for LOB and LLC is beyond the scope of the present study. Yates (1993) presents no unweighted lexical density per clause for his ACMC corpus and, as mentioned in section 2.5, Collot (1991) did not study lexical density at all.

From table 4.5 it is clear that SCMC, i.e. IRC, contains the fewest lexical items per clause (1.5), but SSCMC, i.e. split-window ICQ, also has fewer lexical items per clause than face-to-face conversations. On the basis of various samples, Halliday (1985a: 80) notes that “a typical average lexical density [per clause] for spoken English is between 1.5 and 2, whereas the figure for written English settles down somewhere between 3 and 6.” Given Halliday’s well-established measure of lexical density per clause, then, conversational writing is definitely not typical writing, but rather shares an important defining characteristic of speech – a low lexical density.

Chatted and spoken texts are made up of large numbers of interrelated short clauses, whereas traditional writing contains longer integrated clauses. This means that any vital interpretation of the lexical density per clause, in table 4.5, must be accompanied by the consideration of average clause length in each of the three media (tabulated in the third column of table 4.5). Furthermore, to explain the utility of the lexical density per clause measure, a provisional measure is interspersed into table 4.5: the proportion of lexical items to total items in the average clause, termed “proportion of lexical items per clause.” From this measure, found in the second column, we can deduce a one-to-one correspondence with the figures presented in table 4.4 for lexical density overall (e.g. SCMC’s proportion of 38.7% lexical items in the clause, in table 4.5, is reflected in its corresponding overall lexical density of 38.7, in table 4.4).⁷² The provisional measurement is provided here to demonstrate the one-to-one relationship between Halliday’s measures of lexical density and lexical density per clause, that the measures in reality are the same. Halliday’s application of the lexical density measure on the clausal level simply underscores the variability of clause length in different genres. Lexical density per clause is a more sensitive measure of lexical density, one that takes into account the number of clauses in texts of equal length and generates more explicit differences in score.

Comparing numbers of lexical items in the clause is a straightforward task; as seen in table 4.5, a typical spoken clause contains more than two lexical items, whereas a chatted clause contains fewer than two. More intriguingly, as turns in conversational writing most frequently consist of a single clause, the measure of lexical density per clause provides a glimpse into the typical turn of chatted interaction. The measure thus enables us to capture the special properties of

72 The percentages indicated in the “proportion of lexical items per clause” column in table 4.5 are based on the unrounded figures for lexical density per clause divided by unrounded average clause length.

chatted language, and their relationship to face-to-face conversations. From table 4.4 we rated that conversational writing ranges slightly higher than face-to-face conversations as regards lexical complexity overall, whereas in table 4.5 the lexical density per clause measure renders a slightly nuanced picture, one which draws attention to the short average clause length in conversational writing.

Calculating the average clause length in LOB writing and LLC speech, or in any other written or spoken corpus, is unfortunately beyond the scope of this study, even though such a project, for contrastive purposes, is highly recommended and anticipated. Until such figures are obtained, the analysis of the lexical complexity of conversational writing is bound to remain a preliminary one. Nevertheless, given Halliday's finding that a typical average lexical density per clause for written English "settles down somewhere between 3 and 6" (1985a: 80) and the concurrent general findings of lexical densities around 50% for writing (see table 4.4), a reasonable deduction is that writing on average contains more than six words per clause, possibly up to twelve. Based on the discussion of numerous invented examples, Halliday assumes that the lexical density per clause for writing is "likely to be of the order of twice as high as that for speech" (1985a: 80). Chafe & Danielewicz (1987) discuss clause construction in spoken and written language, finding "intonation units" (the majority of which are clauses) to vary in length, from 6.2 words per unit in conversation to 9.3 in academic papers. "[U]nder normal conditions," they explain, "a speaker does not, or cannot, focus attention on more than can be expressed in about six words" (1987: 95). Chafe & Danielewicz (1987) point out that writing frees writers from the constraint of production time that keeps down both the lexical variety of spoken language and the size of spoken intonation units. Although their argument holds true for traditional writing, it is inapplicable to conversational writing; chatters are highly constrained in time; they produce even shorter clauses than speakers, and chatted clauses contain fewer lexical words than spoken clauses. Given Halliday's (1985a) assumption that the lexical density for writing is likely to be twice as high as that for speech, it will be interesting then, in the future, to find out the lexical density per clause relationship between writing and conversational writing. A plausible assumption is that it will be of the order of three times as high in writing as in conversational writing.

In conclusion, the measures applied in the analysis of the lexical diversity and specificity in conversational writing have yielded a number of important findings. Firstly, the average word in the computer chats is shorter than in any other medium. Short words are seen as an effect of the immediacy of the online

medium, the short encoding time and users' economy of typing, but also as an effect of the words belonging to the grammatical classes – findings which in turn accentuate the similarity between conversational writing and spoken conversations. Secondly, the type/token ratio of conversational writing is by definition a high one; the tokens in conversational writing display a striking lexico-orthographic heterogeneity, an abundance of types. This heterogeneity is explained by the particularities and irregularities of the uniquely user-generated material, in sharp contrast to the nature of corpora of traditional edited writing and the consistently transcribed corpora of speech. TTR is ultimately deemed an inadequate tool for determining the nature of conversational writing as regards its relationship to traditional writing and speech. Thirdly, a more reliable tool for measuring the lexical complexity of the chatted texts was the measure of lexical density, which finds the ratio of lexical items to all items in the texts, thereby reflecting the relationship between lexical and grammatical items. Conversational writing presents itself with a lexical density intermediate between the average for speech from LLC and face-to-face conversations from the SBC subset. The measure of lexical density per clause, finally, reveals the character of typical clauses in conversational writing, presenting their fewer lexical items per clause than in the face-to-face spoken texts. Complemented with average clause length, the measure was used to determine the proportion of lexical items in clauses. The preliminary results, pending the calculation of average clause length in writing, give the impression that conversational writing is slightly more lexically dense than face-to-face conversations, even on the clausal level, even though the lexical density per clause measure better than the overall lexical density measure manages to accentuate and reflect the short average turns of conversational writing. Taken together, the results point in the direction that conversational writing is a variant of spoken communication, or more precisely: a means of communication in which users package information in grammatically intricate ways that are definitely more speech- than written-like.

4.4 The most salient features

In section 4.2, first and second person pronouns were mentioned to be the first two features, in either of the conversational writing genres, that deviate by more than two standard deviations ($|s.d.| > 2.0$) from the average of Biber's spoken and written genres (Appendix II table 4 from Biber 1988: 77–78). They were taken up in connection with modal auxiliaries as these grammatical categories together are two of the carriers of interpersonal meaning (Fowler &

Kress 1979, Halliday 1985a, Hodge & Kress 1988). It was mentioned in the section, however, that altogether ten out of Biber's 67 features deviate in such a way, and that the present section is dedicated to the other eight: direct WH-questions, analytic negation, demonstrative and indefinite pronouns, present tense verbs, predicative adjectives, contractions and prepositional phrases. Table 4.6 summarizes the frequencies per thousand words of these salient features. By their sheer frequency in the chatted texts (or infrequency, in the case of prepositional phrases), these features together give an intimation of the linguistic character of conversational writing. Out of the features to be taken up below, the first two (direct WH-questions and analytic negation), like modal auxiliaries and personal pronouns, also form part of the interpersonal system in language: direct WH-questions as markers of mood, and analytic negation (*not*, *n't*) as a marker of polarity within the modality system (Halliday & Hasan 1989). The remaining six features do not conform as clearly as these to any one of Halliday's metafunctions in language, but will be surveyed on their own terms, as their distributions reveal important patterns. The two ensuing sections, 4.5 and 4.6, will give an account of other salient features of conversational writing, features not found through Biber's (1988) methodology, which nevertheless are instrumental for determining the nature of the communication. Once all of these features have been considered, we will be ready to apply the final step of Biber's (1988) methodology, to position conversational writing on Biber's six dimensions of variation (in chapter 5).

Table 4.6: *Frequencies per 1,000 words for the most salient linguistic features (i.e. normalized frequencies). "N.a." means that the figure is not available*

	Writing	ACMC	Speech	SCMC	SSCMC
first person pronouns	17.0	57.8	52.8	56.9	88.9
second person pronouns	5.0	17.6	23.0	50.4	45.0
direct WH-questions	0.1	2.9	0.8	3.5	3.9
analytic negation	6.4	15.1	13.9	13.1	29.7
demonstrative pronouns	2.3	6.5	10.6	6.6	16.4
indefinite pronouns	0.9	4.6	3.1	11.7	6.0
present tense verbs	64.6	67.6	112.3	147.2	168.5
predicative adjectives	4.8	n.a.	4.9	8.4	15.3
contractions	4.6	16.6	36.1	30.8	55.0
prepositional phrases	117.3	116.9	91.1	47.0	42.0

Halliday (1985a, 2004) and Halliday & Hasan (1989) propound the theory of metafunctions in language, because “it helps us to interpret the features that we actually find in the text” (Halliday & Hasan 1989: 35–36). The variables of “field,” “tenor” and “mode” “collectively determine the functional variety, or register, of the language that is being used” (1985a: 44). The interpersonal metafunction, the tenor of the communication, reflects the personal relationships involved and is realized in texts through e.g. modal auxiliary use (the hedging of statements), personal pronouns (the presentation of self), both dealt with in section 4.2, as well as through mood (declarative, imperative or interrogative) and the system of polarity (the use of negation). As it turns out, the last two grammatical categories, just like the first two, contain features that distinguish the five media contrasted in this chapter from each other. With regard to the grammatical category of mood, only the interrogative mood is annotated in the texts, in the form of direct WH-questions (detected as WH-pronoun, e.g. *what, where, when, how, why*, + auxiliary), but its distributional pattern reveals the inherently communicative function of conversational writing. Analytic negation (*not*, including the contracted form) is found in previous research to correlate with spoken, communicative texts, and by analogy conversational writing could be expected to follow and display a similar distribution. The five media contrasted, as before, are writing, ACMC, speech, SCMC and SSCMC. ACMC is included for reference in the diagrams but, as the corpus is unavailable, no ACMC text examples will be given. Rather, the survey of all features below focuses on the distributions of the features in writing, speech and the conversational writing genres. Figures 4.6 and 4.7 present the distribution of interrogative WH-questions and analytic negation in the five media. Figures 4.6–4.13 in the present section all reflect table 4.6, representing occurrences per thousand words (i.e. normalized frequencies). All figures and tables in the present and ensuing sections of this chapter are based on average numbers from Biber 1988: 247–263 for writing, Collot 1991: 69–70 for ACMC, Biber 1988: 264–269 and Appendix II table 3 for speech, Appendix II table 1 for SCMC, and Appendix II table 2 for SSCMC, unless otherwise indicated, and the results of statistical tests between SCMC, SSCMC, writing and speech, as before, are found in Appendix VI.

Figure 4.6: Direct WH-questions.

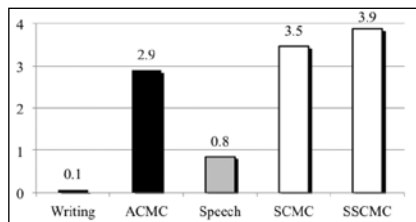
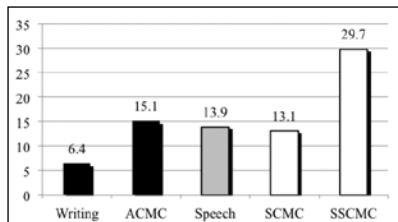


Figure 4.7: Analytic negation.



Questions, both yes/no questions and WH-questions, indicate “a concern with the interpersonal functions and involvement with the addressee” (Biber 1988: 227). Yes/no-questions cannot easily be identified by automatic analysis, and were therefore not included in Biber’s (1988) methodology, but WH-questions, which are more easily identified, were tagged and counted in all of Biber’s genres (in Biber 1988: 247–269), as well as in Collot’s genre of ACMC (in Collot 1991), and, in the present study, in the SBC subset face-to-face conversations (amalgamated into the bar for speech in figure 4.6; see Appendix II table 3 for the frequency in the SBC subset) and the conversational writing genres (SCMC and SSCMC; cf. Appendix II tables 1–2). Biber et al. (1999: 203) point out that “interrogative clauses tend to occur in dialogue situations,” and that “they are frequent only in conversation and (to a lesser extent) in fiction” (ibid.). Judging from figure 4.6, however, Biber et al.’s statement is up for qualification; direct WH-questions are used to an even higher degree in CMC than in spoken conversations (also noted by Ko 1996).⁷³ Figure 4.6 underscores the interpersonal and involved character of computer-mediated discourse: while direct WH-questions are nearly absent in traditional writing, they are slightly more common in speech, and very common in conversational writing. Among the genres amalgamated into the speech bar in figure 4.6, are face-to-face conversations from LLC and from the SBC subset, which contain 0.7 and 2.7 WH-questions per thousand words, respectively, and telephone conversations with 1.1 – genres that contribute to raising the overall figure for speech, but that nevertheless are surpassed by all modes of CMC. Typical WH-questions in IRC are *Where do you come from?*, *What do you do?*, *How are you doing?*, and in ICQ *What are you doing this weekend?*, *How did that go over?*. The slightly different nature of the questions in IRC (more general) and in split-window ICQ (more specific), moreover, reveals the status of the relationships in the two corpora of conversational writing; the IRC chatters are beginning their acquaintance, whereas the ICQ

73 Freiermuth (2003) finds questions overall many times more frequent in chat than in speech and writing, but his results do not specify the occurrence of WH-questions.

chatters inquire into mutually known circumstances – revealing their previous acquaintance with each other. Ko (1996) aptly explains chatters' frequent questions as partly a cohesive strategy; as participants' physical separation obstructs them from coherent and orderly patterns of turn-taking, frequent WH-questions helps them to structure the interaction, "in compensation for the unavailability of other turn-taking cues such as intonation, gesture, and gaze" (1996: no page number available).

Analytic negation (*not*, including the contracted form) and synthetic negation (*no*, *neither*, *nor*) are devices grammaticalized in language for speakers and writers to express negative opposition. Analytic negation typically occurs in conjunction with finite verbs (Biber et al. 1999, Halliday 2004), e.g. *doesn't*, *isn't*, *can't*, and realizes "an essential concomitant of finiteness": polarity, i.e. "the choice between positive and negative" (Halliday 2004: 116). Tottie (1981, 1983b, 1991) finds negation overall to occur twice as often in speech as in writing. Tottie (1991), furthermore, finds "*not*-negation to prevail in spoken language" and "*no*-negation to dominate in written language" (1991: 140, original italics). Biber (1988), like Tottie (1983a), distinguishes between analytic and synthetic negation, finding analytic negation (e.g. *she didn't write any letters that day*) to be more colloquial and fragmented, and synthetic more literary and integrated (e.g. *she wrote no letters that day*). In Biber (1988), accordingly, analytic negation is found to be more than twice as frequent in communicative, spoken interaction, than in written discourse, a finding reflected in figure 4.7. Similarly, in the LSWE corpus, Biber et al. (1999) find negative forms overall to be many times more common in conversation than in writing, with analytic negation most common in conversation and synthetic most common in news. Given the conversational nature of computer chat, analytic negation, as might be expected, turns out to be prevalent in the chats, making the feature deviate markedly in SSCMC from spoken and written language overall. In figure 4.7, SCMC shows a distribution of analytic negation similar to speech, although notably lower than the face-to-face conversations in the SBC subset, which contain 18.9 occurrences per thousand words (Ko 1996, however, finds more analytic negation in his SCMC corpus than in face-to-face conversations). SSCMC, by contrast, contains more than twice as many occurrences of analytic negation as does speech overall.

Upon studying the occurrences of analytic negation in both genres of conversational writing, a few functional distinctions can be made. In IRC, *not* frequently occurs in answers to questions like *How are you?* and *What's up?*: e.g. *not too bad*, *not much*, *not much, u?* and in other generally mitigated, friendly expressions like *don't miss me too much*, *don't mean to sound ungrateful*, and *you have a good day*

now, won't you? The nature of negated expressions in IRC thus reveal the ephemerality, or tentativeness, of relationships formed in the channels. In split-window ICQ, by contrast, analytic negation is often found in connection with adversarial discussions, such as in *DON'T EVEN START WITH ME!!!!!!*, *that wasn't me, it's not funny*, but also in connection with involved, supportive discourse: *so how come you don't talk to mike anymore?*, *I can't take when he is in a bad mood* – which reveal participants' close relationships in real life, outside of the medium. In the split-window ICQ corpus, moreover, turns are occasionally hedged with the abbreviation *idk* (meaning “I don't know”),⁷⁴ a mitigating “marker of uncertainty” (Tsui 1991: 619, Diani 2004: 162) such as in *like idk i'm one of those scarcastic girls...*, *idk he's confusin*, a typically spoken feature not found in the IRC corpus. The ICQ communication thus, more than IRC, serves as an extension of the face-to-face interaction that takes place regularly between interlocutors – involving both adversarial and supportive discourse, as well as mitigation. Tottie (1982, 1983b) attributes the greater frequency of analytic negation in spoken than in written language to the greater frequency of denials, rejections, questions, supports, repetitions and mental verbs in speech. Several of Tottie's (1983b) fundamental categories of negative sentences (e.g. denials, rejections and supports) appear to be more frequent in split-window ICQ than in IRC. In addition, the distribution of what Tottie calls mental verbs (e.g. *know*, *think*, *mean*), largely private verbs in Biber's (1988) methodology, is much higher in split-window ICQ than in IRC (cf. Appendix II tables 1 and 2). Split-window ICQ contains more affective discussions than IRC, with expressions of denial, rejection, support and opinion that ICQ chatters recurrently modalize by means of negative polarity.

The modality system of language, manifested inter alia in modal auxiliary use, choice of mood, negation, and the insertion of a mitigator like *idk*, is available to speakers for encoding attitude towards a statement or the content of an utterance (Hodge and Kress 1988, Yates 1996, Halliday 2004). Hodge & Kress (1988) explain the effect of modality thus:

Modality expresses affinity – or lack of it – of speaker with hearer via an affirmation of their affinity about the status of the mimetic system. Affinity is therefore an indicator of relations of solidarity or of power [...] A high degree of affinity indicates the expression of solidarity between participants. A low degree of affinity indicates that power difference is at issue. (Hodge & Kress 1988: 123)

74 In the conversational writing annotation, initialisms like *idk* and *nm* (meaning “not much”) were tagged as if their constituents were spelled out, finding analytic negation in them. (This treatment, however, was not applied to sentiment initialisms such as *lol* and *lmao*, to be treated in section 4.6, as explained in section 3.2.)

The present chapter has revealed a high degree of modality in SSCMC, e.g. a great number of modal auxiliaries (figure 4.1), frequent switches into the interrogative mood (figure 4.6), prevalent use of the polarity indicator *not* (figure 4.7) and insertion of a hedge such as *idk*. The findings all highlight a significant situational circumstance of the ICQ communication: the ICQ chatters are interpersonally involved in not just the online medium, but also in the offline world, and experience close affinity in both modes. This high degree of affinity, expressed through highly modalized language, indicates, in Hodge & Kress' terms, "solidarity between participants" (1988: 123). Chatters in IRC modalize their utterances to about the same degree as speakers as regards modal auxiliaries, but slightly less than speakers as regards analytic negation. On the other hand, IRC chatters switch into the interrogative mood more than speakers, which reveals that they, too, are interpersonally focused, even though the relationships formed in the public IRC channels tend to be of a more superficial nature. In conclusion, relating Halliday's metafunction of tenor to relevant features annotated in the conversational writing texts has shed light on the relationships among interlocutors and yielded insights into the functions served by the respective media. In what follows, Halliday's metafunctions are left aside for a while, but we will find reason to return to Halliday in other respects shortly.

The next two features that deviate from the mean for all of Biber's spoken and written genres (Appendix II table 4) by more than two standard deviations, in either of the conversational writing genres, are demonstrative pronouns and indefinite pronouns. Biber (1988) subsumes these two features, together with "pronoun *it*," under the heading "impersonal pronouns," in contrast to "personal pronouns" (Biber 1988: 225–226). In the present chapter, personal pronouns have been discussed at length, as first and second person pronouns are clear markers of involved discourse. Pronoun *it*, furthermore, was mentioned in connection with the finding of slightly more grammatical items in the face-to-face conversations from the SBC subset than in the conversational writing genres, which rendered a lower lexical density for face-to-face conversations. Pronoun *it* was found to occur more often in face-to-face conversations partly because of the deictic function it can serve there (cf. the discussion of example (8) above), in addition to the function of substituting for nouns, phrases and whole clauses. Demonstrative pronouns (*that*, *this*, *these*, *those*)⁷⁵ and indefinite pronouns (e.g.

75 "Demonstrative pronouns" here constitute Biber's (1988) feature no. 10, that is (a) *that/this/these/those* followed by verbs, clause-punctuation, tone-unit boundaries, wh-pronouns or conjunction *and*, (b) *that's* and (c) *that* immediately after a tone unit boundary; see Biber (1988: 226) for algorithms. *That* as relative pronoun is not included. Note

everyone, somebody, anything, nothing), as we shall see, can serve similar functions in conversational writing.

Biber (1988: 226) notes that demonstrative pronouns can refer to “an entity outside the text, an exophoric referent, or to a previous referent in the text itself.” Biber et al. (1999: 349) find demonstrative pronouns “far more common in conversation than in the written registers” and demonstrative pronoun *that* in conversation “by far the single most common demonstrative pronoun” (ibid.). As regards indefinite pronouns, they find all groups (the *every-*, *some-*, *any-* and *no-* groups) to be most common in conversation and fiction, and least common in academic prose (1999: 353). The distributions of demonstrative and indefinite pronouns in the five media contrasted in the present chapter are given in figures 4.8 and 4.9.

Figures 4.8 and 4.9 are best understood by studying sample occurrences of the features in the texts. Among the demonstrative pronouns, *that* is the most frequent one in the face-to-face conversations from SBC and in conversational writing genres, and it is used in analogous ways in the three genres, that is, to denote a previous referent in the text itself; see italicized phrases in examples (10) from SBC, (11) from IRC and (12) from split-window ICQ.

Figure 4.8: Demonstrative pronouns.

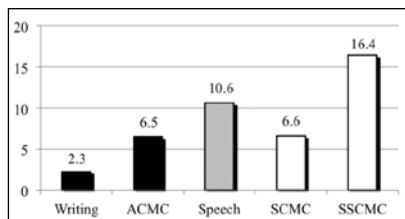
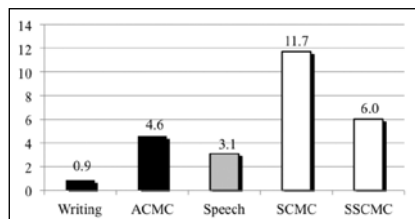


Figure 4.9: Indefinite pronouns.



- (10) Phil: they asked me *to meet with them about ... Teresa's thing.*
 Brad: Mhm
 Phil: ... *that* .. I find v- really,
 ... nothing,
 ... to be honest,
 nothing of any validity.

Face-to-face conversations SBC text 10

that feature no. 10 differs from feature no. 51 “demonstratives” *that/this/these/those*, in that the demonstratives in feature 51 are followed by nouns (e.g. *this thing*).

- (11) <furryman> so will it be a *long interview* blondii
 <blondii> *that* depends on you
 Internet relay chat text 2b (UCOW)

- (12) <I> So *what do you think about Joey?*
 <A> what kind of question is *that*
 Split-window ICQ chat text 1 (UCOW)

Pronouns *that* in examples (10) through (12) refer to events, states, or phrases, rather than to nominal referents. Chafe (1985) finds demonstrative pronouns referring to events and states to occur predominantly in speaking, prescriptively claiming that they are among the several “grammatical devices that are not accepted in written English” (1985: 114). The demonstrative pronouns *that*, *this*, *these* and *those* are naturally inherently deictic in all three corpora, referring to phrases or whole utterances, but also referring to specific nominal referents, such as the italicized noun phrase in (13).

- (13) <AdamSxy35> oups why dont you try a *business chat room on yahoo?*
 <_oups> hm...well do they have *that*..
 Internet relay chat text 5b (UCOW)

Demonstrative pronouns are typically found in passages of involved discussion in all three sampled corpora (see examples 10–13). However, since such affective, involved passages are more rare in IRC than in the face-to-face conversations from SBC or in split-window ICQ, the overall incidence of demonstrative pronouns drops for SCMC in figure 4.8. In speech, demonstrative pronouns can also refer to nominal referents *outside* of the text, e.g. *this is cream soda* (although one might argue that the referent is cataphoric here). Such use is frequent in, for instance, a minor part of LLC, a physics demonstration, but occurs only marginally more in the annotated face-to-face conversations from SBC than in conversational writing. The number of demonstrative pronouns in the SBC face-to-face conversations is 16.0 per thousand words – roughly the same as in ICQ. Chatters in split-window ICQ, in other words, well manage to bridge over the spatial distance between themselves and put the demonstrative pronouns to text-internal deictic use. Their conversations via the written online medium largely follow the same pattern as face-to-face conversations, as regards demonstrative pronouns.

Indefinite pronouns (e.g. *anybody*, *everyone*, *something*) are another feature that split-window ICQ chatters and face-to-face conversationalists employ to an approximately equal extent: 6.0 per thousand words in ICQ, see figure 4.9, and 6.6 in the face-to-face conversations from SBC. Examples are *do you like someone else* and *its great being someone who can be mentor*, in ICQ, and *you have something on your tooth*, in SBC. The SBC example reveals a usage not found in

the conversational writing texts – a reference to something specific, visible to the speaker, but not to the listener. *Something* in the chats, just like the other indefinite pronouns, always refers to a general idea, concept or phrase, or an indefinite person or thing, not immediately visible. Indefinite pronouns are “markers of generalized pronominal reference, in a similar way to *it* and the demonstrative pronouns” (Biber 1988: 226, original italics). The split-window ICQ and face-to-face conversations display functionally analogous usage of indefinite pronouns, numerically on a par, but as can be seen in figure 4.9, indefinite pronouns are almost twice as common in IRC as in ICQ.

What brings about the high frequency of indefinite pronouns in IRC? The answer to the question is very simple, and it is immediately discernible in the various occurrences sampled from IRC in (14 a–f).⁷⁶

- (14) a. *anyone* wanna chat
b. *anyone* from sydney
c. hello *everyone*
d. how old's *everyone*?
e. *Anybody* here???
f. wassup with *everyone* today

Internet relay chat (UCOW)

IRC chatters employ indefinite pronouns when angling for conversational partners in the channel, but also in greetings and questions intended for indefinite recipients. This kind of usage accounts for half, or more, of the indefinite pronouns in the IRC texts, which wholly explains the high frequency of indefinite pronouns found for SCMC in figure 4.9. Noting similar results for his chat corpus, Ko (1996) relates the high frequency of indefinite pronouns to the situational context; “[u]sers do not know for certain who their audience is at any given moment” (1996: no page number available).

As mentioned, Biber et al. (1999) find in LSWE approximately the same number of indefinite pronouns in fiction as in conversation (a rough estimate is 5 per thousand words, in each genre, for the same indefinite pronouns that Biber 1988 considers), a number significantly higher than in the other genres they studied: news and academic prose. Biber et al.'s (1999: 352) finding for fiction, however, does not tally with Biber's (1988) figures for fiction. Among Biber's (1988) written genres in figure 4.9, a fiction genre, adventure fiction, contains the highest number of indefinite pronouns (2.7 per thousand words) but most other genres, including

76 As explained in section 1.4, no text numbers are given for examples that contain turns sampled from several texts.

the other fiction genres, contain fewer than 2 per thousand words. Thus, no parallel similar to Biber et al.'s (1999) finding can be drawn between fiction and conversations, or conversational writing, in the present study. Among Biber's (1988) spoken genres in figure 4.9, the highest number (3.9) is recorded for face-to-face conversations, closely followed by telephone conversations (with 3.6). All three corpora annotated in the present study thus display a usage of indefinite pronouns beyond previously recorded findings. In conclusion, indefinite pronouns in conversational writing are worthy of mention not just for contributing to the obviously oral character of the online communication, but also for distinguishing functionally among the conversational writing genres. Their use in IRC reveals one of the main functions of the public medium: finding conversational partners.

Out of the ten features in conversational writing that deviate by more than two standard deviations from Biber's mean for speech and writing, only two, in themselves, may constitute lexical items: present tense verbs and predicative adjectives, both to be taken up next (direct WH-questions and prepositional phrases, of course, may also contain lexical items). However, while predicative adjectives are lexical by default, a vast number of present tense verbs are forms of the verbs *be*, *have* and *do*, which are grammatical items (cf. Quirk et al. 1985: 67 and section 4.3 here). Judging from the low lexical density found for conversational writing in the previous section, the prevalence of grammatical items among the most salient features tallies with possible expectations; conversational writing indeed contains many more grammatical than lexical items. The distributions of present tense verbs and predicative adjectives in the five media are shown in figures 4.10 and 4.11.⁷⁷

Figure 4.10: Present tense verbs.

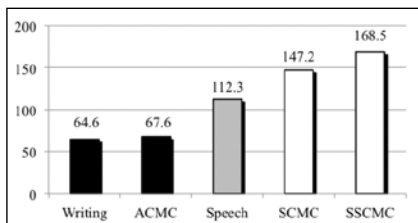
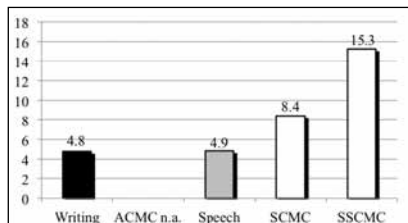


Figure 4.11: Predicative adjectives.



77 The figure for predicative adjectives in ACMC is not available (Collot 1991: 69).

Present tense verbs and predicative adjectives do not share pragmatic-functional properties like the pairs of features treated above (direct WH-questions and analytic negation as parts of the modality system, and demonstrative and indefinite pronouns as markers of impersonal pronominal reference). The ten features in this account are all, naturally, entirely unforeseen, as they have crystallized from their sheer frequency in the conversational writing genres and not by kinship or the author's choice. The pairing of features in the present section is thus mostly incidental and applied for practical, rather than necessarily linguistically motivated, reasons. All the same, when studying the textual occurrences of one of the ten features, very often another one pops up. Such is the case for nearly all of the occurrences of predicative adjectives in the annotated corpora; examples from SBC are *That's not bad, Tha- that's right, They're cool*; from IRC *this is slow, i'm lost, your welcome* and from ICQ *its ok, that's cute!!, That's pretty cool...*⁷⁸; in which the predicative adjectives tend to co-occur with present tense verbs, very often with demonstrative pronoun that, and sometimes with analytic negation (not).

Present tense verbs is one of the features that carry the largest weight on Biber's (1988) first dimension, distinguishing texts with highly involved, interactive discourse from texts with more informational content. Figure 4.10 illustrates the pervasiveness of present tense verbs in speech, as opposed to writing. Present tense verbs "deal with topics and actions of immediate relevance" (Biber 1988: 224), whereas past tense and perfect aspect verbs are typically markers of narrative or descriptive, mostly written, texts (Biber 1988, Biber et al. 1999). On Biber's (1988) first dimension ("Informational vs. Involved Production," to be discussed in chapter 5), present tense forms indicate a verbal (involved), as opposed to nominal (informational), style. Spoken language is typically verbal, interactional and affective, whereas written language is nominally elaborated (cf. Wells' 1960 verbal and nominal styles). Judging from figure 4.10, the verbal, involved style found in speech is augmented further in conversational writing – in split-window ICQ chat, practically every sixth word is a present tense verb, in writing, by contrast, only every sixteenth.

Among the spoken genres in figure 4.10, the face-to-face conversations from SBC contain 141.6 per thousand words, face-to-face conversations from LLC 128.4 and telephone conversations 142.6; the three highest figures for speech. Sample occurrences of present tense verbs in SBC and the conversational writing genres can be found in any one of the numbered text examples given in this chapter, requiring only a few examples to be given here. Furthermore, in section

78 Biber's (1988: 238) algorithm 41 (b) for finding predicative adjectives was interpreted "be+adv+adj+xxx (where xxx is not adj or n)."

5.2.1, the impact of present tense verbs for distinguishing among genres will be discussed, in conjunction with examples from the annotated corpora. Clearly, present tense verbs contribute to a sense of orality in conversational writing, a sense that is further born out in that they are frequently also private verbs (e.g. *feel, know, think, guess*). Present tense verbs are about as indicative of speech as nouns are of writing. In fact, if asked to distinguish among genres by one word class alone, opting for verbs might prove a felicitous choice, as by their character, tense and frequency, verbs reveal a great deal about a text's genre affiliation. Private verbs, for instance, on average occur twice as often in Biber's (1988) spoken, as compared to his written genres. Present tense verbs, moreover, occur as private verbs twice as often in split-window ICQ as in IRC; in ICQ to about the same extent as in face-to-face conversations and in IRC almost as infrequently as in writing. Examples of present tense private verbs in ICQ are *I know, i think i'll just..., that's cool I guess, I mean if you like him* – typically used to introduce evaluative and emphatic utterances. The relative rarity of private verbs in IRC is likely to be due to the superficial character of relationships in the public channels; interlocutors simply do not know each other well enough to discuss preferences, express evaluation or give supportive advice. The low frequency of private verbs in IRC, in turn, is a likely explanation for the slightly fewer present tense verbs found for IRC (SCMC), as compared to split-window ICQ (SSCMC), in figure 4.10. What this means for IRC, with regard to Biber's first dimension, will be further explored in the next chapter.

Turning now to predicative adjectives (figure 4.11), a few remarks are in order. Firstly, "predicative adjectives" is not a factor that distinguishes among Biber's (1988) genres of writing and speech. As seen in figure 4.11, writing and speech contain approximately the same number of predicative adjectives (4.8 and 4.9 per thousand words, respectively). Consequently, predicative adjectives, as a linguistic feature, did not load on any of Biber's (1988) dimensions of genre variation (unlike all other features discussed in this section).⁷⁹ Collot (1991), therefore, decided not to count the feature, which is why no result is available for ACMC in figure 4.11. For IRC and split-window ICQ, however, identifying and summing the predicative adjectives has proved highly valuable: figure 4.11 indicates that SCMC contains nearly twice as many, and SSCMC more than three times as many predicative adjectives as writing and speech, respectively. This means

79 Predicative adjectives loaded tentatively on Biber's (1988) fifth dimension, but their low weight (0.31) was below the cut-off point (0.35) for the feature to be considered in dimension score calculations.

that, if a new factor analysis was carried out with the inclusion of chatted texts, predicative adjectives may turn out to load on one of the resulting dimensions.⁸⁰

While predicative adjectives do not distinguish between written and spoken genres in Biber's (1988) study, nor in Chafe's (1982) account, attributive adjectives do. In Biber's (1988) methodology, attributive adjectives are identified as all adjectives preceding nouns, or otherwise "not identified as predicative" (1988: 238). Attributive adjectives are used to elaborate nominal information, and thus highly integrative in their function (Chafe 1982, Chafe & Danielewicz 1987, Biber 1988), while predicative adjectives "might be considered more fragmented" (Biber 1988: 237). Chafe (1982) notes that the use of attributive adjectives "allows states to be expressed as modifiers rather than assertions," e.g. "the *old* house," as opposed to "the house was *old*," and calls them integrative devices and a prevalent feature of written language (1982: 41–42, original italics). To delve beyond the equal figures for writing and speech in figure 4.11, therefore, we will consider the ratio of predicative to attributive adjectives in writing, speech, SCMC and SSCMC. The results of such a calculation reveal that, in Biber's (1988) genres of writing, only every fifteenth adjective is predicative; in speech, every tenth (although in the SBC subset as many as every fifth); in SCMC, every seventh; and in SSCMC, as many as every third adjective is a predicative adjective. If attributive adjectives are typical of nominal, written discourse, then the relative rarity of attributive, and the prevalence of predicative adjectives, is typical of conversational writing. Although previous studies have not found predicative adjectives to be typical of speech, the present study finds predicative adjectives to be highly typical of conversational writing.

Biber (1988) notes that predicative adjectives are frequently used for marking stance. Biber et al. (1999) find that, "(s)emantically, the most frequent predicative adjectives of conversation tend to be evaluative and emotive, e.g. *good*, *lovely*, and *bad*" (1999: 516, original italics). The examples of predicative adjectives from SBC, IRC and split-window ICQ given above (in connection with the discussion of present tense verbs) confirm Biber et al.'s finding for conversation, for the conversational writing genres: the predicative adjectives in conversational writing are also largely evaluative and supportive responses to statements made by partners in the online conversations. Two additional such examples conclude the account of predicative adjectives here: example (15) illustrates a typical occurrence in IRC and example (16) one in ICQ, both evaluative and/or supportive.

80 Interestingly, face-to-face and telephone conversations from LLC contain 4.2 and 6.0 predicative adjectives per thousand words, respectively, whereas face-to-face conversations SBC contain 8.2, possibly suggesting that they are becoming more frequent in conversations, or simply are more frequent in American than in British English conversations.

- (15) <yazzie^> I'm coming to Aussie next Xmas 404!!
 ...
 <Guest22> wow yazzie, thats *great*...for how long
 Internet relay chat text 4b (UCOW)
- (16) <6> but if practice makes perfect and no ones perfect then y practice?
 <F> gotta practice! It makes perfect ya know
 <6> i try
 <F> thats *deep*
 Split-window ICQ chat text 5 (UCOW)

The final two features, out of the ten that collectively epitomize the character of conversational writing (by deviating from Biber's mean for speech and writing by more than two standard deviations), are contractions and prepositional phrases. Contractions deviate by their high frequency in split-window ICQ, and prepositional phrases by their striking *infrequency* in both conversational writing corpora. The two features are entirely independent of each other in the texts, but they share the ability to distinguish among writing, speech and conversational writing. The distributions of the features are shown in figures 4.12 and 4.13.

Figure 4.12: Contractions.

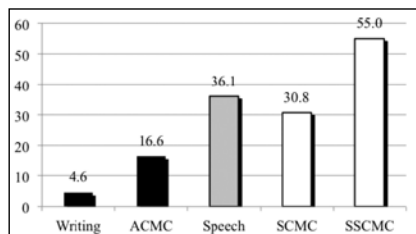
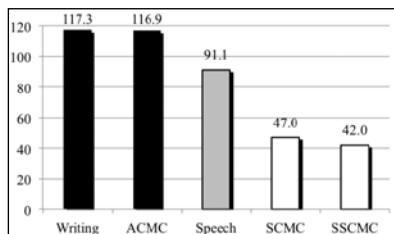


Figure 4.13: Prepositional phrases.



Chafe & Danielewicz, in their 1987 account of the properties of spoken and written language, find contractions (e.g. *it's*, *I'm*, *don't*), as well as prepositional phrases, to be distinguishing factors between speech and writing. "Spoken language commonly employs contractions" whereas "[s]uch items are rare in academic written language" (Chafe & Danielewicz 1987: 93). Finegan and Biber (1986) find contractions to be distributed as a cline: most frequently used in conversation, least frequent in academic journals, and with intermediate frequencies in broadcast, public speeches and press reportage. Biber (1988) presents very similar findings, except that he also finds official documents to be virtually void of contractions; see figure 4.12 for the average figures for speech and writing found here (based on Biber 1988, the former supplemented with the SBC subset). In the face-to-face conversations from LLC, there are 46.2 per thousand words;

in the SBC subset, there are 48.5; and in telephone conversations, as many as 54.4 per thousand words, whereas there is only one (1) contraction in the 28,000-word official documents component of LOB studied. “Contractions are the most frequently cited example of reduced surface form,” says Biber (1988: 243). Biber et al. (1999) separate the reduced surface form into verb contractions (e.g. *she’s going*) and *not*-contraction (e.g. *couldn’t go*), but find the distribution of both types to follow the same decreasing cline among their four genres: conversation > fiction > news > academic writing, in order of frequency. In the present study, contractions were found to be slightly more frequent in SSCMC than in spoken conversations, but less frequent in SCMC than in speech overall; see figure 4.12.⁸¹

Detecting contractions, as well as most other features, in conversational writing requires meticulous manual annotation. For written and spoken texts (that is: texts transcribed by linguists), automatic detection of contractions is usually possible; one simply queries the text for apostrophes and then sorts out irrelevant hits (e.g. to exclude genitive inflections). Most contractions in chat, by contrast, do not contain apostrophes, and like other chatted words, they are frequently misspelled. Chatters, governed by economy of typing, by leaving out apostrophes take the reduced form one step further. Below are various occurrences of contractions that illustrate the intricacies of annotational detection. The examples in (17), from IRC, and (18), from ICQ, also illustrate the results of chatters’ economy of typing, i.e. some ultra-reduced surface forms.

- (17) *your* (you’re), *yvw* (you’re very welcome), *their*, *there* (they’re), *where* (we’re), *dunno*, *tis*, *whatcha* (what’re you), *lets*, *lits* (let’s), *wassup*, *wasssup*, *sup* (what’s up), *whats*, *whast*, *whts*, *whxx* (what’s)

Internet relay chat (UCOW)

- (18) *their*, *there* (they’re), *were*, *where* (we’re), *ur*, *your* (you’re), *no ones* (noone’s), *souldnt* (shouldn’t), *dunno*, *dunnp*, *donno* (don’t know), *idk*, *idjk* (I don’t know), *whos*, *itz*, *lets*, *thas*, *ain’*, *shoulda* (should’ve), *can’s*, *caznt* (can’t), *whats*

Split-window ICQ chat (UCOW)

81 Contractions in Biber’s (1988: 243) terms are those on pronouns, on auxiliary forms (negation) and suffixed on nouns (except possessive forms). The following are examples of contractions accordingly left unannotated in the corpora annotated in the present study: in IRC: *wheres*, *where;s* (“where has/is”), *where*, *WR* (“where are”), *whered* (“where did”), *theres* (“there is”), *heres* (“here is”), *old’s* (“old is”), *hows* (“how is”), *kinda*, *fliirt’n*, *dutch’n*; in ICQ: *pound’n*, *laugh’n*, *where* (“where are”), *theres* (there is), *kinda*, *tell’n*, *turn’n*, *pay’n*, *wait’n*, *frik’n*, *look’n*, *let’n*, *kid’n*; and in the SBC subset: *how’s*, *kinda*, *where’s*, *there’s*.

As mentioned in section 2.5, Freiermuth (2003) uses Chafe & Danielewicz's (1987) methodology to contrast linguistic features in synchronous political chat data, with spoken discussions from a political television talk show, and written (political) newspaper editorials. In his SCMC data, Freiermuth finds chatters to use contractions less frequently than speakers, but more often than writers – a finding that is corroborated in the present study, but that does not lend itself to easy explanation. What partly limits the number of contractions in IRC, compared to split-window ICQ, is the relative rarity of analytic negation in IRC (as this includes the contracted form *n't*; see discussion of analytic negation above), but compared to speech this explanation is insufficient (as IRC contains only marginally fewer instances of analytic negation than speech). A different, more likely, explanation for the rarity of contractions in IRC is presented in section 5.2.1; in the present section, we acquiesce in Freiermuth's and the concurrent finding, for the genre of SCMC, and simply observe that the conversational writing genres diverge from each other as regards frequency of contractions. As seen in the examples in (17) and (18), however, contractions in both genres show analogous composition, and both groups deviate from writing and transcribed speech in that they are occasionally realized as ultra-reduced forms. The SSCMC users employ contractions to about the same degree as speakers in conversations (55.0 per thousand words in ICQ vs. 48.5 in the SBC subset), whereas the SCMC users employ them less.

The final linguistic feature that deviates from Biber's mean for speech and writing by more than two standard deviations ($|s.d.| > 2.0$) is prepositional phrases; see figure 4.13. This feature deviates negatively for both conversational writing genres; SCMC and SSCMC both display a remarkable paucity of prepositional phrases. The frequency of prepositional phrases is nearly three times as high in writing as in conversational writing. Chafe & Danielewicz (1987, as well as Chafe 1982, 1985) find prepositional phrases, and sequences of them, to be factors that distinguish written discourse from spoken discourse, as represented by academic papers and conversations respectively. In connection with the lexical density discussion in the previous section (4.3), Chafe & Danielewicz's concept of the "intonation unit," roughly equivalent to a clause, was touched upon. Chafe & Danielewicz (1987), in their discussion of the intonation unit, expound on linguistic devices that writers, more than speakers, employ to increase the size of the unit. One of the devices is prepositional phrases (other devices are attributive adjectives, also discussed above, and e.g. nominalizations). Prepositional phrases thus typically elaborate the nominal information and expand the length of clauses. Biber (1988) postulates that prepositional phrases are "important device[s]"

for packing high amounts of information into academic nominal discourse” (1988: 237), but in his study they are also found to be frequent in other kinds of written discourse and, actually, most frequent in official documents. In the LSWE corpus, Biber et al. (1999) find prepositional phrases most common in academic prose and least common in conversation. The results of Biber’s (1988) study with regard to prepositional phrases in LOB writing and LLC speech, are reflected in figure 4.13: writing contains on average nearly 30 percent more prepositional phrases per thousand words than speech. Among the genres amalgamated into the speech bar in figure 4.13 are face-to-face conversations (LLC with 85.0, and the SBC subset with 61.1) and telephone conversations (with 71.8). Spoken American English (the SBC subset) somewhat restricts the elevation of the speech bar; yet, conversational writing contains significantly fewer prepositional phrases than the SBC subset. Apparently, very little clausal elaboration by way of prepositional phrases (or e.g. attributive adjectives and nominalizations; cf. Appendix II) takes place in conversational writing. Ko (1996) and Freiermuth (2003) both find a similar sparsity of prepositional phrases in their chat corpora, Ko making the observation that the chatted clauses “tend to be stripped down to their obligatory core, minus optional adjuncts such as prepositional phrases” (Ko 1996: no page number available).

In the lexical density discussion in the previous section, conversational writing was found to display more grammatical than lexical items. A prepositional phrase is initiated by a preposition (a grammatical item), and in written texts the phrase typically contains at least one nominal (lexical) item. Prepositional phrases, as a feature, therefore, are practically neutral in the lexical density calculation for written texts (as 1 grammatical + 1 lexical item “cancel” each other out). In spoken language and in conversational writing, however, the composition of the prepositional phrase is usually different. In these media, a typical prepositional phrase contains just a stranded preposition (grammatical) or a preposition followed by other grammatical items (such as pronouns). Prepositional phrases thus typically contribute to *lowering* the lexical density for spoken and conversational writing texts. On the other hand, prepositional phrases are extremely rare in the latter genres, as shown in figure 4.13. The effect of prepositional phrases and other elaborating devices on the mean length of written clauses, however, is palpable. It was seen in the discussion of lexical density per clause (section 4.3) that the average “intonation unit” (roughly: clause) in academic writing is 9.3 words long (Chafe & Danielewicz 1987). Table 4.5, furthermore, revealed that the average clause length in face-to-face conversations is around six words (also found by Chafe & Danielewicz 1987 for conversation), and that the average

conversational writing clause only is about four words long. Considering average clause length in conjunction with figure 4.13, consequently, we find what is partly missing in conversational writing clauses: clause-extending devices, such as prepositional phrases.

A few textual examples will shed further light on average clauses, and the effect of prepositional phrases in them. The excerpts from academic prose (19), face-to-face conversation SBC (20), IRC (21) and split-window ICQ (22), below, serve to illustrate the typical distribution of prepositional phrases in respective genres. The prepositional phrases are marked by their preposition in bold script.⁸²

- (19) It is not clear that the growth **of** the spread **between** earnings and wage rates **in** the UK over the period **of** our sample can be plausibly explained **in** cost terms. If it is argued that such a gap is automatically opened **by** the rise **in** piece-workers' earnings as productivity increases, or **by** changes **in** the amount **of** overtime worked, such changes may themselves be traced back **to** the existence **of** a high level **of** demand.

Academic prose LOB J: text 44

- (20) Jamie: Aren't you guys gonna stick up **for** me?
and beat up **on** him or something?
Miles: He's bigger **than** I am.
Pete: (laughter)
Miles: He's not bigger **than** you.
Pete: No.
Harold: But he's my –
Miles: (laughter)
Harold: he's my friend
Pete: Tha- that's right.
Miles: (laughter)
Jamie: (laughter)
Pete: You know who I'll stick up **for**
Miles: (laughter)
Pete: ... I stuck up **for** you today **at** that store.
Harold: That's true.
Jamie: ... You did.
You made me get the,

82 Biber's list of prepositions (used to detect prepositional phrases) is taken from Quirk et al. (1985: 665–667), but excludes prepositions “that have some other primary function, such as place or time adverbial, conjunct, or subordinator (e.g., *down*, *after*, *as*)” (Biber 1988: 236–237) as well as, for instance, *over*. Examples (20) and (21) contain one and two stranded prepositions, respectively, which also count as instances of feature no. 61 (stranded prepositions; see Appendix II).

Pete: Mhm,
 Jamie: um,
 Pete: that's right.
 Jamie: the green scarf.
 ... That's right.
 ... He was my fashion consultant today.

Face-to-face conversations SBC text 2

- (21) <Guest_258> wassup **with** everyone today
 <SoulSearchR> Be Back Later
 <italan> hi lily lily
 <bored> hey P where you **from**?
 <furryman> still least your still be young when they grow up.lol
 <darth> well do it again barbiegirl
 <blondii> yeah, thats how i look at it
 <Guest_258> oh isee
 <Lilly_Lilly> hi iatalan
 <furryman> so whens the next one.
 <blondii> the youngest is 3, so i dont know
 <carrots35ca-bbl> hb SoulSearchR
 <blondii> no more!!
 <italan> where you **from**
 <barbiegirl> cool rock
 <nane> Mart??
 <brokenwing-ange> if one day you dont see me anymore...it means i given up **of** my life

Internet relay chat text 2a (UCOW)

- (22) <9> who said i hooked up **with** her
 <I> if u dont wanna be **with** laurie anymore, why did u just hook up **with** her **on** saturday???
- <9> we were both lying there and i kissed her but i wouldnt say we hooked up
 <I> i asked her yesterday when th elast time u hooked up and she told me satrunday. but dont tell her that im telling u this.
 <9> cause she thought katie was still awake
 <9> i dunnp

Split-window ICQ chat text 8 (UCOW)

Excerpts (19) through (22) are approximately equally long (c. 75 words), but whereas the academic prose example (19) contains 13 prepositional phrases, the sampled SBC face-to-face conversation (20) contains seven, and the conversational writing excerpts, (21) and (22), only five and four, respectively. The sloping cline for prepositional phrases across written and spoken genres, from academic prose to conversations, found by other scholars (Chafe 1982, 1985, Chafe &

Danielewicz, Biber 1988, Biber et al. 1999) thus continues its descent across conversational writing, as seen in figure 4.13.

That prepositions “serve to integrate high amounts of information into a text” (Biber 1988: 104) is distinctly shown in example (19) from academic prose. In (19), the prepositional phrases each contain at least one lexical item (*spread*, *earnings*, etc.) and the phrases extend and elaborate clauses to make the text extremely integrated. Moreover, prepositional phrases are stacked upon each other (*by changes in the amount of overtime*) in sequences, which Chafe & Danielewicz (1987) find typical of academic writing. In the other three examples, (20) through (22), however, the prepositional phrases display an entirely different distribution. Not only are the prepositions here often left stranded, which Chafe (1985: 115) cites as examples of “errors” typical in speech due to production constraints, but also the prepositional phrases contain mostly grammatical items, and therefore less clearly serve the function of elaborating clauses. Halliday (1987) calls the complexity of written language “crystalline,” “whereas the complexity of spoken language is choreographic” (1987: 66). He explains the latter thus:

The complexity of spoken language is in its flow, the dynamic mobility whereby each figure provides a context for the next one, not only defining its point of departure but also setting the conventions by reference to which it is to be interpreted. (Halliday 1987: 66–67)

Consequently, the difference between writing and speech lies not just in the presence vs. absence of prepositional phrases, or in the relations between lexical and grammatical items, but also in the usage of these items. Halliday (1987) criticizes Chafe (1982) for describing both writing and speech “using a grammar of writing” (Halliday 1987: 67). Halliday instead proposes a kind of choreographic grammar, one that recognizes the intricacy of spoken language; that “its mode of being is as process, not as product” (1987: 67). For Halliday spoken language has:

[...] a considerable degree of intricacy; when speakers exploit this potential, they seem very rarely to flounder or get lost in it. In the great majority of instances, expectations are met, dependencies resolved, and there are no loose ends. (Halliday 1987: 67)

Halliday explains that the intricacy of spoken language is of a grammatical kind; it has multiply linked clause structures. This intricacy requires the use of grammatical items, as they provide the glue that connects the parts of a spoken utterance together (Halliday 1987, Yates 1993).

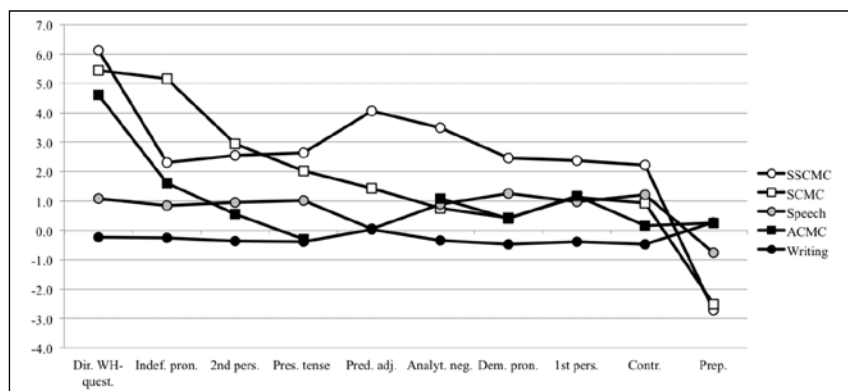
Whether we side with Chafe or Halliday is of secondary importance in the account of conversational writing here. Chafe and Halliday, of course, have both

developed their stances over the years; Halliday into his (choreographic) functional grammar (e.g. Halliday 2004) and Chafe along a more cognitive linguistic track (e.g. Chafe 1994); though both constantly in tune with natural language data. Their interpretational quibble apart, the excerpts of face-to-face conversation and conversational writing, (20) through (22), have managed to elucidate the important primary finding here: the striking similarity of the three conversational genres. Prepositional phrases are distributed in analogous ways in conversations and in both genres of conversational writing, ways that sharply distinguish these genres from the most “written” mode of writing, the genre of academic prose (19). With regard to clausal elaboration by means of prepositional phrases, other genres of writing, and speech, are intermediate between these two poles. That academic prose constitutes the “written” end of the pole, as regards prepositional phrases, was a well-established fact. The present study has extended the “spoken” end beyond conversations, finding prepositional phrases in conversational writing not just to be rare, but also possibly to serve other functions than just clausal elaboration.

In conclusion, to sum up the ten salient features of conversational writing, i.e. first and second person pronouns (described in section 4.2) and the eight features explored in the present section, we will take advantage of the standard scores calculated for each feature. Recall from chapter 3, section 3.5, that a standard score was computed for each feature, which equals the feature’s number of standard deviations from Biber’s mean for speech and writing (Appendix II table 4, from Biber 1988: 77–78). These standardized scores are ideal for enabling the comparison of features across texts and genres, and crucial for the calculation of comparable dimension scores. The present chapter has exploited the fact that the features with the highest standard deviation in conversational writing are the features that collectively epitomize the nature of conversational writing. The cut-off point for a feature’s inclusion as a salient feature in the present section was two standard deviations, which meant that a convenient number of ten features crystallized. (Modal auxiliaries, word length, TTR and the lexical density measures are included in this chapter for other justified reasons.) The ten most salient features are not necessarily the most frequent features, but the features that together distinguish English chatted texts, on average, from English written and spoken texts, on average. Naturally, the make-up of conversational writing is more complex and many-faceted than what the ten most salient features depict, and in the next chapter, therefore, all of Biber’s 67 features will be taken into account to more accurately describe the chatted material. It was decided, nevertheless, that the features that deviate from Biber’s mean by more than two

standard deviations would be of statistical interest, and that the account of them in the present chapter serves well as an introduction to the more all-round investigation of the conversational writing genres in chapter 5. Figure 4.14, finally, sums up the ten most salient features of conversational writing, or rather: those features, which, in either SCMC or SSCMC (or both), deviate from Biber's mean by more than two standard deviations, for each showing its distributions in the other three media as well. The zero point in figure 4.14, by inference, constitutes Biber's (1988) mean for speech and writing. The standard scores are based on numbers from Biber 1988: 247–263 for writing, Collot 1991: 69–70 for ACMC, Biber 1988: 264–269 and Appendix II table 3 for speech, Appendix II table 1 for SCMC and Appendix II table 2 for SSCMC, contrasted with the mean numbers and standard deviations for Biber's speech and writing overall (Appendix II table 4, from Biber 1988: 77–78); see section 3.5 for a description of the procedure of standard score calculation.

Figure 4.14: Standard score distribution of the linguistic features that, in SCMC or SSCMC, deviate by more than 2 s.d. from Biber's (1988) mean.⁸³



In the next two sections, other important features of conversational writing will be taken up – features that are characteristic of conversational writing, but not identified through Biber's (1988) methodology; firstly, the paralinguistic cues and extra-linguistic features of chat, the latter most common in IRC, and lastly, two salient linguistic features: inserts and emotives.

83 The standardized score for predicative adjectives in ACMC is unavailable (Collot 1991: 69).

4.5 Paralinguistic features and extra-linguistic content

Before the advent of computer-mediated conversational writing among the general public in the late 1980s, linguists justifiably concluded that writing is unable to incorporate all the features of speech. Halliday (1985a), for instance, points out that:

There are various aspects of spoken language that have no counterpart in writing: rhythm, intonation, degrees of loudness, variation in voice quality ('tamber'), pausing, and phrasing – as well as indexical features by which we recognise that Mary is talking and not Jane, the individual characteristics of a particular person's speech. (Halliday 1985a: 30)

The features that writing typically leaves out are what in spoken language are known as prosodic and paralinguistic features. Prosodic features are part of the linguistic system; they extend across long stretches of speech (e.g. rhythm, intonation, pausing and phrasing) as systematic phonological realizations, as in an intonation contour (Halliday 1985a: 30–31). Paralinguistic features can also extend across varying stretches of speech, but they are “not systematic – they are not part of the grammar, but rather additional variations by which the speaker signals the import of what he is saying” (1985a: 30), as by the degree of loudness, variation in voice quality (“tamber”), tempo and facial/bodily gestures. Halliday (1985a: 31) considers “prosodies” and paralinguistic to be of linguistic status, but calls a third group of features non-linguistic, “indexical.” Indexical features are not part of the language at all, but rather “properties of the individual speaker” (1988: 30), such as individual preferences for certain prosodic and paralinguistic patterns. The prosodic, paralinguistic and indexical features are difficult to represent in writing, says Halliday (1985a: 30) “because they do not belong at any particular point.” Yet, Halliday proceeds to challenge and partly dismiss the notion that these features are entirely missing from writing. Spacing and punctuation (comma, semicolon, full stop, question mark, parenthesis, etc.), he claims, are used in writing to overcome the omission of prosodic features. Spacing marks off words, and punctuation marks off grammatical units, or prosodic units, giving written text systematic variation similar to the intonation contour in speech. Nevertheless, Halliday inevitably resorts to the conclusion that “[w]ritten language never was, and never has been, conversation written down” (1985a: 41). Except for the linguistic transcription of natural spoken recordings (for linguistic research) the task of writing down speech is not what writing is about. “Why?” Halliday asks rhetorically, and answers:

—because in its core functions, writing is not anchored in the here-and-now. The particular conditions that obtain at the time of writing are not going to be present to the

reader anyway, who is usually at some distance from the writer both in time and place; so much of the message that is contained in the rhythm and tamber of speech would simply be irrelevant. (Halliday 1985a: 32)

Having made this case for writing, it is easy to see how conversational writing differs from writing: conversational writing, in its core function, *is* anchored in the here-and-now (cf. Ooi 2002). The particular conditions that obtain on the computer screen, the ideational “field” (Halliday 1985a) in the “ideational metafunction” (Halliday 2004), are present to both interlocutors at once. The text is presented to them dynamically – it happens, much like airwaves traveling through the air in speech. Linguists inquiring into SCMC therefore generally concur in describing computer chat as speech-like communication. Dresner (2005) goes so far as to propound that the visual perception of the transmitted text is analogous to auditory reception:

In a simple (i.e., single-window [...]) chatroom situation all participants sit in front of their computer screens. All of them are seeing the same thing—the text lines accumulating in front of them. As opposed to visual perception in spoken conversation, where each participant sees a completely different picture, in textual conversation vision functions somewhat like hearing in auditory discourse—it enables mutual focus on the buffer on which communication takes place. We see that the affinity between ordinary and textual chat goes beyond (or, rather, deeper) than synchronicity. The structure of mutual visual perceptual intake in computer mediated textual chat is topologically similar to its auditory counterpart. (Dresner 2005: 15–16)

This means that computer chatters, like what Halliday claims for listeners in conversations, are “predisposed to take a dynamic view of what [the text] means” (1985a: 81). Conversational writing thus turns text from “product” into “process” and writers from authors into interlocutors, that is, almost into speakers and listeners.

Interlocutors in conversational writing use a number of prosodic, paralinguistic and indexical devices, here generically called “paralinguistic features,” to enrich their writing with cues that assimilate speech, or at least to assimilate a situation similar to face-to-face interaction. Conversational writing by no means incorporates all the paralinguistic features of speech, but several of the devices employed, as we shall see, are passable attempts to bridge the gap to face-to-face spoken discourse. In conversational writing, the paralinguistic cues are applied to written text, and not spoken, and therefore differ somewhat from Halliday’s definition. Paralanguage, in this section, is used as a broad term covering several salient aspects of conversational writing that Biber’s (1988) features fail to include, aspects ranging from nicknames, personalization tropes and self-imposed

spoken language transcription, to abbreviations, graphic devices, “leet,”⁸⁴ interlanguage and code-switching. The paralinguistic of conversational writing is realized in the messages of the communication. The paralinguistic devices therefore also provide clues to the role language plays in online communication, the semiotic mode of conversational writing, i.e. what the language is being used to achieve, as regards, for instance, conscious self-representation (Halliday 1985a, Halliday & Hasan 1989), reflecting the textual metafunction of language (Halliday 2004). Besides paralinguistic, this section will also cover extra-linguistic factors in the communication that are not always realized in the user-generated messages, such as pictures and music shared among the chatters. Extra-linguistic factors form important parts of interlocutors’ shared time and space in conversational writing (their field), influencing their communication. The survey of paralinguistic devices and extra-linguistic factors is essentially brought in here to complement the comprehensive linguistic analysis to be undertaken in chapter 5.

Paralinguistic features, innovative orthography and neologisms in textual computer and cellphone communication have been pet areas for linguistic researchers over the past few decades as witnessed by a host of publications dealing with these (inter alia Wilkins 1991, Yates & Orlikowski 1993, Werry 1996, Jonsson 1998, Schulze 1999, Crystal 2001, 2008a, Gajadhar & Green 2003, Baron 2008, Waldner 2009, Rowe 2011, to name but a few). As mentioned, the primary concern of the present study is to apply Biber’s (1988) methodology, with its 67 linguistic features, to the conversational writing data, no feature of which covers paralinguistic and extra-linguistic factors. The chatted texts, as described in chapter 3, were annotated for Biber’s list of linguistic features after the texts had been purged from bracketed nickname turn indicators, server-generated messages, action commands and certain other strings of text (e.g. graphic noise and mass-advertising dumped into the IRC channels) that were impossible to tag and/or apt to skew the results (for examples of excluded material, see Appendix IV). This purging was kept to an absolute minimum, as it was of utmost importance that texts remain as intact as possible, and that all user-generated, i.e. keyed-in, linguistic messages, in English, were retained. The present section, however, is devoted to bringing some of the excluded material temporarily back into the account.

84 Leet, “leet speak” or “1337 5p34k” denotes the language of “elite” chatters, such as online gamers and hackers, who e.g. incorporate symbols and numbers as substitutes for letters in words. It is partly used as a means for experienced users to demarcate themselves from “newbs” or “n00bs” (those new to the medium) (see e.g. Van de Velde & Meuleman 2004, Blashki & Nichol 2005, Nichol & Blashki 2006). LeBlanc defines leet or “l33t” as “elite geek speech” (LeBlanc 2005: 72).

The first paralinguistic device employed by chatters, in both IRC and ICQ, is the choice of a nickname, decided upon before logging in. IRC nicknames (nicks) are usually easily changed, whereas ICQ nicknames (more like user-IDs) are connected to an account. (The ICQ nicknames in the present study, however, were not chosen by participants, but pre-set on lab computers by the present researcher, for practical reasons.) As seen in the IRC text samples in the present chapter, chatters make a conscious choice of nicknames; examples are big-dog, River, Chaser, } melons{, Sweet_Victoria, Cheeky1, BillClinton and blondii. “The nick is their electronic identity,” says Crystal (2001: 160); “it says something about who they are, and acts as an invitation to others to talk to them” (ibid.). Anglemark (2009: 89) notes that “[t]he nick is often the only identity a chat room participant displays in a chat session.” Indeed, quite frequently, IRC chatters “lurk” in the channel, “eavesdropping” without contributing to the ongoing communication. Occasionally, chatters signal their presence with an empty turn, displaying only their nickname, as <remut> in (23).

- (23) <Heart35> some using Mark
 <remut>
 <dony> c\free
 <biro> hi nuttygrl :o)
 <remut>
 <bergs> ..it is Brad!

Internet relay chat (UCOW)⁸⁵

Other chatters put their nicks to creative use in combination with their turn, as in the example of attempted flooding (dumping repeated jabberwock) in (24), and with the graphic feature in (25).

- (24) <Can[You]Handle[This]> %%f8738200%%
 <Can[You]Handle[This]> %%f8738200%%
 ...
 <Can[You]Handle[This]> %%f3738003bf9c129fec%%
 *** Can[You]Handle[This] was kicked by Sheila (flood)

Internet relay chat (UCOW)

- (25) <dj_19_m_uk> <===== any girls with pic message me!!!

Internet relay chat (UCOW)

85 Unnumbered Internet relay chat texts in examples (23) through (45) are from the part of the corpus that exceeds the texts sampled for annotation; see description of corpus creation in chapter 3.

The chatter's nickname is indicated within angle brackets, by the software, in the chatter's every turn. These bracketed nickname turn indicators are not part of the annotated IRC corpus, yet it must be recognized that in the ongoing communication they have a certain discourse value. Crystal (2001) points out that "they provide a crucial means of maintaining semantic threads in what is otherwise a potentially incoherent situation" (2001: 161). Moreover, the nicknames that are used as address terms in messages provide invaluable links in the conversational threads. Crystal considers the function of these links "analogous to the role of gaze and body movement in face-to-face conversation involving several people" (2001: 162).

The second paralinguistic device available to chatters in ICQ and web-based chats (not in IRC) is the choice of font style, font color and font size. The chatters in the split-window ICQ corpus employ this device diligently, with constant changes, to personalize their messages in a way comparable to the vocal variation of intonation found in speech. The changes in font style, color and size are retained in the corpus, though not reproduced in textual examples given here. Other personalization schemes are exemplified in (26) and (27), whereby individuals attract attention in the flow of IRC turns.

(26) <^mekrisi^> hi guys does any one wanna
chat ???
Internet relay chat (UCOW)

(27) <}}melons{{> \V/elcome Back angeldelight
Internet relay chat text 1b (UCOW)

Chatters typically mark their entrance into the chat room/channel/program/site by a greeting, e.g. *Hello All, hii all, hey room imback*, in which the first element is an interjection. (Interjections are particularly pervasive in the IRC texts, and all interjections were tagged in the corpora in the present study, but as they are not among Biber's (1988) list of features, they will be treated separately, among "inserts" in section 4.6.) Greetings, like other turns, can be personalized; see the two alternative enthusiastic responses to electrolite's modest general greeting in (28), in which BK is trying to attract electrolite's attention.

(28) <electrolite> hi all
<BK> _,-*^*~,-,_,-*' electrolite _,-*^*~,-,_,-*'
<BK> □□□□□□□□ Hello electrolite □□□□□□□□
Internet relay chat text 4b (UCOW)

The keystrokes in BK's turns in (28) are combined into iconographic effects, making up sets of decorative strings. While the IRC interface used by the chatters in

UCOW has no readily available supply of graphic icons, the ICQ program (and web-based chats) provide users with a choice of graphic emoticons, e.g. ☺. Moreover, as mentioned in section 3.3, ICQ has a supply of graphic action tropes for users to employ ad hoc by a simple click. A graphic action trope is realized in the text as e.g. “B picks a flower and hands it to you”; see Appendix IV. The chatters in the split-window ICQ corpus used these readily available graphic devices to a moderate extent. However, as the inclusion of a graphic icon or an action trope implies no conscious linguistic typing on the part of the chatter, neither device was retained in the purged material for annotation. For consistency then, in IRC, action commands were also purged away before the annotation of Biber’s (1988) features (see section 3.2 for a description of the purging process and Appendix IV for examples). Both the IRC and the split-window ICQ chatters, however, used textual emoticons (e.g. :, ;) which were preserved in the texts and tagged as emotives – and therefore to be treated separately, along with inserts, in section 4.6.

In their messages, chatters employ a vast number of paralinguistic devices to assimilate spoken interaction, i.e. to transcribe their own texts as if into speech. Enthusiasm, surprise, anger, or mere emphasis, is signaled through repeated exclamation marks (*Sweetie!!!!!!!, whoops!!!!, NOOOOOO!!!!, i know!!!!!!!!!!!!!!!*) and puzzlement through repeated question marks (*uhhhh....?????, when???????*). Punctuation is also used to signal pauses (*for sure chanel...can’t match up to ours huh...lol*). Capital letters mark off text expressed in a loud voice, sometimes as if it was screamed (*i’m very ANGRY!, Well i DO like those skateboarders... especially MATT!, THAT IS SOOOO MEAN, CAPS ARENT COOL THEY WILL GET U KICKED*). Repeated letters denote added emphasis, e.g. *this suxxxxxxxxxxxxxxxx* or, for instance, long vowel sounds (*oooooooooooooooooooo u didnt say that before well then thats a whole different ball game, YOU’VE GOT WORMS EWEWW-WWWWWWWW*). As seen in the latter, the two devices can be combined for increased effect (capital, repeated letters). Capital letters are fairly common in split-window ICQ, but very rare in IRC – their use in IRC is regarded as screaming, for which the channel operator may “kick” the user from the channel. Two passages from the conversational writing texts serve particularly well to highlight interlocutors’ sense of being on the verge of an auditory medium (as suggested by Dresner 2005): example (29) from IRC and (30) from ICQ.

- | | | |
|------|--|---|
| (29) | < mad_max >
< mad_max >
<Cheeky1>
< mad_max > | missing my voice
to scream a little ...
hahaha
screaaaaaaaaammmmmmmm!!!!!!!!!!!!!!!!!!!!!! |
|------|--|---|
- Internet relay chat text 3a (UCOW)

- (30) <Pilot1> yo, did you read that capian underpants bok
 <Pilot1> dude i'm not reading when i'm typing. i,m outof practice, i haven't typed any school paper's or e-mails in a while. yeah ne way....
 <esoteric> hi dude you can't spell. dude why are your eyes brown? you are boring to talk to so i have to get someone else to type. you are slow. yes. yes YES!!! dang you are slow. just use 2 fingers neither am i i am looking at the keyboard. duh! oh yeah whatever shut up yeah yeah i'm not listening..... la la la la la la la la oooh saaaaay can you seeeeeeee!!!!!!!!!

Split-window ICQ chat text 12 (UCOW)

Example (30) ends in the transcribed equivalent of the user esoteric singing the first stanza of the US national anthem. It is part of a turn in which the user is trying to “make his voice heard” over the conversational partner in the split ICQ window (as if they were speakers in the auditory medium). The users were new to SSCMC and were at once intrigued and annoyed by the supersynchronicity, which entailed that most of esoteric’s turn in (30) was overlapped by the other interlocutor, Pilot1’s, turns. That there is overlapping “speech” in (30) illustrates the similarity between split-window ICQ and face-to-face conversations. On the other hand, the supersynchronous mediation of text in split-window ICQ goes beyond speech, in that it does not require interlocutors to “stop and listen” at the same points as in the auditory medium. Experienced interlocutors in split-window ICQ can carry on with their interaction and simultaneously listen (read) and speak (write), only pausing in strategic moments to maintain a certain consecutiveness in the communication. If this were done in the auditory medium (in long completely overlapping passages), the communication would be rendered incomprehensible. Supersynchronous CMC thus not only resembles auditory conversation; it surpasses it.

IRC surpasses auditory conversation by another type of simultaneous “speech,” in that a vast number of online chatters can engage in a conversation at once. Chat channels “provide virtually unlimited access to people who want to chat on a particular channel in a moment in time” (Freiermuth 2003: 31). However, chat channels rarely contain only one conversation; rather, several conversational threads are interlaced, requiring untangling skills from users in order for threads to be followed. Elsner & Charniak (2008, 2010) find an average of 2.75 conversations active at a time in their IRC corpus. If IRC were an auditory situation, it would be a cocktail party (Crystal 2001). Dresner (2005) notes for the auditory situation that a person can “admittedly catch his name in a conversation going on in another part of the room, but the rule is that we do not, and cannot, follow more than one conversation line for a substantial period of time” (2005: 20). IRC

chatters, on the other hand, are “continually perceptually aware of more than one conversation line” (2005: 21). Dresner goes on to explain that it is the “visual spatiality” of the synchronous texts that enables chatters to untangle conversations; “(p)ictorial processing abilities seem to help us sort out the entanglement of conversation lines” (2005: 21). Following Dresner’s reasoning, then, computer chat approximates auditory face-to-face interaction; yet, it is only through the visual medium that simultaneous speech, as in split-window ICQ, and simultaneous threads, as in IRC, can be perceived. In either format, to be sure, chatters must be apt typers to keep up with the simultaneous reception and production of text.

IRC chatters, possibly more than split-window ICQ chatters, are concerned about keeping pace with the conversations at hand. Certainly, IRC chatters have slightly more processing time than speakers, but in order to stay abreast of the unfolding conversation, they must construct text quickly. When typing in on-line chat, “it becomes imperative to use precious construction time efficiently” (Freiermuth 2003: 171). Werry (1996) points out that:

The language produced by users of IRC demands to be read with the simultaneous involvement of the ear and eye. One can discern an intensified engagement with the sounds of language, with the auditory and iconographic potential of words. (Werry 1996: 59)

This “intensified engagement with the sounds of the language,” with the auditory potential of words, brings chatters to impose spoken language transcription schemes upon their discourse, such as those discussed among the paralinguistic features above. The “iconographic potential of words” (Werry 1996: 59) is further explored below.

Earlier in this chapter, we observed the short clause length of conversational writing, which indicates brief turns. The brief turns, moreover, consist of very short words. Occasionally, the short words constitute abbreviations (initialisms such as *idk* (I don’t know), *brb* (be right back), *lol* (laughing out loud), *lmao* (laughing my ass off), *a/s/l?* (age/sex/location?)), which chatters employ to speed up typing, but which really represent several longer words in themselves. The answer to the latter initialism (*a/s/l?*), for instance, might be almost as brief as the question, and yet, impart a great deal of information (e.g. *31/blk m/usa tx, 20/m/syd*). (As mentioned in section 3.2, abbreviations were retained in the present study and annotated for their constituent linguistic items; *idk*, for instance, was tagged with Biber’s (1988) features nos. 3, 6, 56, 59 and 67.) Werry (1996) observes a general tendency for IRC words to be stripped down to “the fewest possible letters that will enable them to be meaningfully recognized” (1996: 55). The same tendency is observed in both the IRC and the split-window ICQ

corpus in the present study, though more markedly so in IRC. Abbreviations of the IRC kind, once deciphered, are linguistic; yet, the initiated users of chat abbreviations exploit their paralinguistic, iconographic potential to control the orthographic “prosody” of their message, to accelerate its tempo. The initialisms are more common in the IRC than in the ICQ chats, and naturally wanting from the spoken and written language corpora (though adolescents were overheard to employ them playfully in spoken discourse around the turn of the millennium, and a few chat initialisms, like *irl*, “in real life,” seem to linger in speech; more on this shortly). Another reduced form of language in online chat is apostrophe-less contractions, discussed in the previous section (4.4). Freiermuth (2003) notes that it is likely that production time plays a role when chatters leave out apostrophes; “one less character to type means that the time it takes to post a message is reduced by a few precious milliseconds” (2003: 101).

While it is true that chatters are concerned with economy of typing, it is equally true that they occasionally post pre-composed strings of text, or graphic textual compositions, into the chat (more so in IRC than in split-window ICQ). The actual posting takes only a copy-paste-enter move, even though the pre-composing, possibly in a word processor, may have been a more cumbersome task, as in the instances in (31), (32), a US flag, and (33), a rose.

- (31) <EasterBookCase> ,°°'KiSs'°° ,LoVe,°°'KiSs'°° ,LoVe,°°'KiSs'°°
 ...
 <EasterBookCase> Hi MOM_OF_3_BRATZ! I'm just so happy to
 see you today! :)
 <EasterBookCase> ,°°'KiSs'°° ,LoVe,°°'KiSs'°° ,LoVe,°°'KiSs'°°
 Internet relay chat text 4a (UCOW)
- (32) <GaGaSue_NYC_NYU> * * * * * _____
 <GaGaSue_NYC_NYU> * * * * * _____
 <GaGaSue_NYC_NYU> * * * * * _____
 Internet relay chat (UCOW)
- (33) <Guest_698> @---}----- 4 all you ladies
 Internet relay chat text 2a (UCOW)

As a rule of thumb, any string of text containing any linguistic item found among Biber's (1988) features was annotated for this constituent feature. This means, for instance, that the first and third turns in (31) contain five nouns each (Biber's feature no. 16), but also that a few limited graphic features, like the rose in (33), were annotated as nouns as well. The decorative elements in (28) were retained in the annotated corpus, but without annotations as they do not constitute as clear equivalents of nouns as, for instance, the rose in (33). Graphic features extending

beyond the turn, as in (32), however, were removed before the annotation (this particular instance was found to occur five times, but it was the only graphic feature to extend beyond one turn in the corpus annotated).

The examples of conversational writing in this section illustrate that chatters are masters of their keyboard. They exploit its every key to enliven their textual interaction, rendering turns spoken-like to the point of their being sung, and graphic to the point of their being art. Just like in spoken discourse, there are slips of the tongue in computer chat, or rather, slips of the key. A famous slip of the key from 21st-century CMC, more specifically from computer gaming lingo, is found in the verb *pwn*, meaning “own.” In “elite” computer-mediated chat lingo (so-called “leet,” “l33t” or “l337”), used by e.g. gamers and hackers, *pwn-ing* stands for “owning” (Pichlmair 2010). A computer gamer taking over an enemy base, or a hacker taking over a server, would say that they pwn it. A slip of the key thus perpetuated in this sub-language and eventually became a symbol of how leet-speakers, advanced chatters, “pwn” the English language (Pichlmair 2010). Moreover, at the beginning of the 21st century, *pwn* (pronounced /pəʊn/ in British, /poʊn/ in U.S. English) and other leet terms (e.g. *noob*, meaning “newbie,” i.e. inexperienced users; *lol*; *irl*, as in “meeting in RL” and *leet* as a term in itself) passed over into spoken language, mostly among adolescents (Bennett 2007). The persistence of these terms in the spoken medium, of course, remains to be proved. If *pwn* is given the persistence of the term *qwerty*, which denotes a standard for keyboards introduced in the 19th century, also derived from the adjacency of keys, the term *pwn* is likely to stay in the language outside of CMC for some time. Unfortunately, no matter how intriguing the subject matter, a more thorough analysis of the lexis of leet is beyond the scope of this study; instead, interested readers are referred to e.g. Van de Velde & Meuleman (2004), Blashki & Nichol (2005) and LeBlanc (2005).

The lexis of the conversational writing corpora in the present study, in both IRC and ICQ, is English (in which leet is reflected, of course). English is the only language allowed in the recorded chat channels, and it was the only language allowed in the recording of the split-window ICQ chats. In IRC, language rules are often displayed automatically upon the user’s entrance and channel operators are particularly quick to enforce them. Nevertheless, users in IRC are globally dispersed, and English is not the native language of all of them, which means that a few instances of interlanguage, code-switching and non-English fonts inevitably surface in the IRC corpus, as in examples (34) through (36).

- (34) <DJ-XNS|Vs|DJ_RMX> halooow are ther some one ho will talk
with a swedich boy?
Internet relay chat text 5a (UCOW)

(35) <CLAUDIAA> si somos muchos lo que hablamos espanolporque/
Internet relay chat (UCOW)

(36) < mouad__> □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □
...
*** mouad__ was kicked by AussieDino (You have been kicked
for using non-english fonts. Please speak in English next time.)
Internet relay chat (UCOW)

Naturally, the interlanguage in (34) causes no action from the channel operator. The user in (35) is politely reminded of the language rule, whereas the channel operator's enforcement upon the use of non-English font in (36) is severe (and included here as an illustration of channel operator interference). IRC communication is ASCII-based, and therefore unable to render non-English fonts correctly. (The non-English orthography in (36), consequently, was not correctly represented in the log, either.) As partly touched upon in section 3.2, instances such as (34) through (36) were treated in the following way in the present study: all interlanguage, such as in (34), was retained and annotated, whereas all foreign language turns, such as those in (35) and (36), were removed, as well as all foreign language items within English turns (extremely few). The same procedure applied to the ICQ and SBC texts, whereby a few foreign turns and words were removed.

The omniscient presence of a vigilant channel operator is perceptible to all users in IRC. The operator's and interlocutors' nicknames are displayed as a list in the software, which is constantly updated upon users joining and leaving. The chat channel is thus the virtual equivalent of a room, in which people mingle, chat and act upon each other's actions. In sharp contrast to rooms in real life, however, the chat channel is textual – the mingling, chatting and acting is carried out via written characters. Or rather, they are carried out in written characters for the most part. A few extra-linguistic factors make their way into the communication. These will be tended to shortly.

The linguistic environment – the co-text, or context – of any linguistic item is crucial for the interpretation of the item. Halliday & Hasan (1976, 1989) call text-internal reference “endophoric” (co-textual) and situational reference “exophoric” (contextual). Reference is one of the cohesive devices that enact the textual metafunction in language, reflecting the semiotic mode of the interaction. Endophoric reference makes a text cohere within itself and exophoric makes it cohere with the context of situation (Halliday & Hasan 1976). Endophoric reference is realized, for instance, in the use of personal and demonstrative pronouns (referring to antecedents) and other text-internal deictic devices. Another cohesive

device is ellipsis – the omission of words that are recoverable in an earlier passage of text. Ellipsis is frequent in the conversational writing corpora, for instance in answers to questions; see B's answers to 2 in (37).

- (37) <2> do u know were ur going to college yet?
 umm i dont know
<2> were do u want to go
 umm 7 places

Split-window ICQ chat text 2 (UCOW)

Endophoric reference and ellipsis are thus co-textual (Halliday & Hasan 1989), i.e. inferable from the surrounding (here: preceding) text. Example (37) illustrates what typically happens in spoken interaction, as well as in conversational writing; Hughes (1996) postulates that there is far more ellipsis in speech than in writing as “speakers can assume that listeners will ‘fill in’ the gaps from their shared knowledge” (1996: 155). Exophoric reference, by contrast, typically depends on the speaker “pointing to” something in the text-external, situational context, explicitly or implicitly (as in a nod or a gaze), “for example, ‘she’s nice’ said with a nod towards a person in the vicinity” (Hughes 1996: 155). Exophoric reference, per definition then, is more common in speech than in writing (cf. Coleman 1996: 43). In writing, exophoric reference, expressed in for instance dialogs, needs to be explicitly explained by the author in order for the reader to understand what is referred to. In face-to-face conversations, the extra-linguistic content is evident in the surrounding environment in which speakers are situated. The extra-linguistic situation is thus often brought to bear on typical spoken discourse, or, rather, speech typically depends on the extra-linguistic, contextual situation.

Conversational writing is carried out in a contextual situation distinct from both speech and writing. The interlocutors’ physical surroundings may be vastly, that is globally, separated, but the communication takes place in a shared, virtual space on the interlocutors’ computer screens. This virtual space (the chat window itself, or an adjacent window) can carry shared extra-linguistic information, or content that affects both interlocutors at once. The IRC protocol preceded the hypertext protocol (the World Wide Web), and was in its earliest form a mere textual affair (although occasionally complemented with file transfer via protocols like ftp and gopher). Over the years, however, IRC users increasingly complemented their communication with information shared via other protocols (the direct client-to-client protocol, for instance). The IRC chat corpus in the present study gives proof of, or suggests, a few instances in which extra-linguistic content is being shared. Posting web addresses into the public

IRC channel is rarely tolerated, and the interlocutors in the IRC corpus are not found to discuss web content. The direct client-to-client contacts and the private chats, however, may involve the sharing of web sites from which, for instance, scripts can be obtained. Scripts obtained for free may surreptitiously program a user's leave-message to display an advertisement of the free script site, as in (38). The high frequency of leave-messages like (38) in the corpus therefore suggests that users engage in the sharing of scripts, which in the direct client-to-client protocol or private chat (outside of the public channel) most likely yields instances of exophoric reference. The chatter in (39) is using a script that automatically detects what music is being played on the user's computer and displays this as an action in the public channel, an action which may lead to the user being asked to share the file. In example (40) a brief sound is played into the channel (audible only in a few chat clients, i.e. IRC "programs"), and in example (41) a chatter initiates a trivia game to be played with fellow chatters in the public channel.

- (38) *** Tina^^B has quit IRC (»;« Scoop Script 2001 »!« The best script ever seen! Get yours copy at www.scoopsite.com)
Internet relay chat (UCOW)
- (39) * I_C_Triple is now playing: Artist: Tukan | Title: Light A Rainbow [CJ Stone Rmx] | Genre: Trance | Year: 2001 | Comment: <http://mphase.6x.to> | Quality: 160kbps 44kHz | Position: 1:31 | Length: 8:02
Internet relay chat (UCOW)
- (40) [SittingBull SOUND]
Internet relay chat (UCOW)
- (41) <sOLDierZ____> 04Starting the trivia. Round of 035
04questions. 03!strivia 04to stop. Total: 037841
Internet relay chat (UCOW)

Text such as that in (38) through (41) was not retained in the annotated corpus; (38) is a server-generated join- and quit-message; (39) and (40) are action commands, and the turn in (41) is not consciously keyed in as a linguistic message by the user, but rather produced through strict programming. Nevertheless, the examples provide clues to what extra-linguistic content might be shared in pending private chat windows or via the client-to-client protocol. Shared music files are more common than parlor games in the IRC corpus (the game in (41) is the only instance). On the whole, the sharing of extra-linguistic content leaves remarkably few imprints on the discourse in the public channel. Example (42) is a rather amusing exception, in which |mad_max| hums a song being played

(recites its lyrics) and eventually asks Brutal_Beauty for a dance, and in (43) a chatter expresses his/her enthusiasm over another song played. Overall, the most commonly shared extra-linguistic content seems to be photographs. In example (44) two chatters share photos via file transfer and discuss these, and (45) exemplifies another turn with exophoric reference to a shared photo.

- (42) <|mad_max|> looking back over my shoulder
 <Tha-Kappo-tan> hey what up people
 <|mad_max|> i can se e that look in ur eyes
 <|mad_max|> eyes*
 <Brutal_Beauty> Tha-Kappo-tan, Nothings up here. :)
 <Tha-Kappo-tan> any people from the land of Oz msg me
 <|mad_max|> hey, bartender gimme some more of that!!!
 <Brutal_Beauty> |mad_max| :S
 <|mad_max|> wow
 <|mad_max|> hi, beauty
 <|mad_max|> u wanna dance?
- Internet relay chat text 3a (UCOW)

- (43) <yazzie^> !BK I-Will-Survive.wav
 WooHoooo!!!...like taking candy from a baby!!!
 ...
 <yazzie^> can you send that song to me plz BK
- Internet relay chat text 4b (UCOW)

- (44) <Genie500> oh river just a sec I gotta turn something off for
 you to send okay
 <River> this one is from 95 without the glasses .
 <River> ok
 ...
 <Genie500> okay try again
 ...
 <River> but the hair is almost the same now as then
 <River> plus a wee bit more grey in it
 <Genie500> Laughing Out Loud ok
- Internet relay chat text 4a (UCOW)

- (45) <SittingBull> [Bahamut] i need to send a newer picthat one
 was in england and from 2 years ago
- Internet relay chat (UCOW)

Whereas there is a relative paucity of exophoric reference to shared audible and physical (i.e. virtual) extra-linguistic devices in the public channels, more subtle kinds of exophoric reference permeate throughout. It is evident in the corpus, for instance, that the IRC chatters experience their software window, and the textual

flow, as a confined, shared space, much like a room in real life. Spatial pro-forms and other exophoric references to the room abound (*here, where, back, on*, cf. Quirk et al. 1985: 514ff); see the various turns in (46). Chatters *look for* people in different rooms, *see* each other in rooms, or refer to other, private, rooms, as in the various turns in (47), and they refer deictically to both the room, and the ongoing interaction, (as *this*) in (48).

- (46) a. Anybody *here*???
- b. I'll be *here* for a while...
- c. *where* shes not answering me
- d. *where* have yu been
- e. i will be *back*
- f. \\elcome *Back* ^xelle^
- g. matt is *on*...lmao

Internet relay chat (UCOW)

- (47) a. h0rnyale you just *missed* ann...she was *lookin for* ya
- b. *looking for* saba
- c. Hi MOM_OF_3_BRATZ! I'm just so happy to *see* you today! :)
- d. *see* ya barbie
- e. she is *in my room* hcmk

Internet relay chat (UCOW)

- (48) a. well now *this* is fun isnt it
- b. just getting use to *this*
- c. is *this* slow tonite or what?

Internet relay chat (UCOW)

Moreover, chatters refer exophorically to the shared time in the room (*while ive been away, 2night, tonight, later*), as in (49), incidentally ignoring that, in their global dispersion, a time adjunct like *tonight* may be perceived differently in a different time zone.

- (49) a. ah been talking *while ive been away* have you ?
- b. no ops here *2night*
- c. hm not much talking in here *tonight*
- d. u r really a big help *2night*
- e. tks see you *later* ulsterman
- f. hey i'll talk to ya all *later* i need to jet for a *lil while*

Internet relay chat (UCOW)

Besides spatial and temporal adverbials, Halliday & Hasan (1976) and Halliday (2004) also consider, inter alia, the definite article and personal pronouns to be carriers of exophoric reference; “the definite article is the item that, in English, carries the meaning of specific identity or ‘definiteness’ in its pure form” (1976:

32) and this definiteness can sometimes be achieved only through an examination of the situational context. The first and second person pronouns “do not normally refer to the text at all” but rather are “normally interpreted exophorically” (1976: 48), whereas the third person essentially refers to the text, but also “may refer exophorically to some person or thing that is present in the context of situation” (1976: 49). In (50), two chatters are exchanging files and experiencing trouble opening the files because of an unknown file format. From their use of the definite article (in *the first one, the extention*), and the subsequent pronoun *it* (*di it open*), it is evident that both chatters from their situational context can infer which file and which extension are referred to. All the while, their exophoric reference is obscure to other chatters, who do not have access to the same extra-linguistic material. In (51), the definite article (in *the server*) signals shared common knowledge among all chatters on the same server, but to an outsider, reading this log, it is not evident which server is referred to. Thus, extra-linguistic information plays an important role in both cases.

- (50) <River> oops Genie500, *the first one* you may not be able to
 open, forgot to look at *the extention*.
 <Genie500> back
 <big-dog> 'WeLCoMe BaCK.Genie500'WeLCoMe BaCK.
 <River> wb Genie500
 <River> *di it* open for you ?
 <Genie500> Thank You River big-dog
 <Genie500> not yet I froze up when I tried

Internet relay chat text 4a (UCOW)

- (51) <River> looks like big troubles on *the server* today

Internet relay chat text 4a (UCOW)

The various turns in (52), finally, exemplify exophoric reference whereby chatters refer to other chatters in the room, almost as if they were nodding or gazing at the intended referent. A plural second person pronoun (*u girls, you 2, u*) is used to address two participants, or the *ladies* identified in the room. Third person pronouns (*he, she*) refer to foregoing speakers, and it is evident to all chatters that pronoun *they* (in *theyd boot you*) refers to the rigid channel operator. In neither case does the pronoun refer to an explicitly stated, anaphoric, referent, but rather to persons simply identified as present in the room, inferred from the extra-linguistic context (for instance, from the list of logged-in participants). The first person plural *us* (in *let's sing him*) is also clearly exophoric, including all chatters as referents and a foregoing speaker (*him*) as the recipient of the intended action.

- (52) a. *u girls* are from the uk right
b. hey *you 2* gonna quit fighting and talk to me or what?
c. hello ladies *any of u* care too chat with me
d. chanel *he's* here...lmao
e. *he's* not tooking with you
f. *she* is here hcmk23
g. didnt know *theyd* boot you for saying s\$cks
h. let's sing *him*

Internet relay chat (UCOW)

To sum up, the *explicit* extra-linguistic content shared in connection with the IRC communication (e.g. music, pictures, a game) is found to leave remarkably few traces in the discourse, whereas the *implicit* extra-linguistic content (the shared space, the shared time, the turns themselves, and the people apparent in the room) gives rise to prevalent exophoric reference. Naturally, defining the latter extra-linguistic content as exophoric is an intricate matter, as the content is indeed reflected in users' messages (as if endophoric) – nevertheless, the reference to it is contextual, not co-textual, as examples (46) through (52) have shown. As mentioned, Halliday & Hasan (1989) defined endophoric reference as co-textual, referring to the surrounding text, and exophoric reference as contextual, referring to the shared situation. In conversational writing, the shared situation is largely made up of text, and yet, this mass of text and the shared window, together, make up an extra-linguistic environment, a room, in which people interact.

The present section has explored the paralinguistic features of conversational writing, finding the account of them to elucidate the semiotic mode of conversational writing, the “particular part that the language is playing in the interactive process” (Halliday & Hasan 1989: 24). Chatters' nicknames are conscious choices for self-representation, and chatters' personalization tropes and self-imposed spoken language transcriptions all tinge their turns, just like their abbreviations, graphic devices and instances of “leet,” interlanguage and code-switching. Most turns in the chat carry a clue to the identity of their producer, regardless of whether chatters consciously exploit their major means at hand, the keyboard, to construe the identity or not. A majority of the section has tended to the circumstances of IRC, but naturally, several of the features equally apply to the medium of split-window ICQ. In split-window ICQ, however, the virtual room is usually shared by only two participants, who know each other outside of the medium, which means that more intense chatting goes on, and less action, joining, leaving, and conscious self-representation. Moreover, the ICQ chatters in the present study were instructed not to leave their chat window, and were therefore unable to share extra-linguistic content, such as music, graphics, or web sites.

The split-window ICQ corpus has fewer references to the shared space and time in the chat, but more to the shared real-life environment (*how bout we bounce outta here, you shoudl come down, yesterday, satruday, last night, last weekend, next year*). In both IRC and split-window ICQ, exophoric reference is made to the shared contextual situation, but whereas the IRC chatters share only the virtual room, the split-window ICQ chatters share both the virtual and the real life “room” (cf. section 3.3), and this is reflected in their chats.

In the next section, two salient *linguistic* features of conversational writing are discussed: inserts and emotives. They are not found among Biber’s (1988) list of linguistic features, but emerged in the annotation process as decidedly characteristic of chatted texts.

4.6 Inserts and emotives

Neither of Biber’s (1988, 2006) two major multidimensional analyses of the English language considers the use of interjections, or “inserts” overall, in the spoken and written genres studied. Yet, linguistic intuition suggests that inserts are one of the most immediate discriminating markers of spoken discourse, apt to be an influential factor in any analysis distinguishing among spoken and written registers. At an early stage, therefore, it was decided that the corpora annotated in the present study should be tagged for their inserts. In the annotation of the IRC corpus (SCMC), moreover, it soon became evident that without this feature, nearly every tenth word would have been left untagged (typically greetings). Biber et al. (1999) describe “inserts” as a class of words typically found in conversations, recognizing that “[i]f we are to describe spoken language adequately, we need to pay more attention to them than has traditionally been done” (1999: 56). Accordingly, Biber et al. (1999) devote a subsection of the chapter entitled “The grammar of conversation” to inserts, grouping them into nine major functional types: interjections (e.g. *oh, ah, wow*), greetings and farewells (e.g. *hello, bye*), discourse markers (e.g. *well, right, now*), attention signals (e.g. *hey, yo*), response elicitors (e.g. *right?, huh?*), response forms (e.g. *uh huh, mhm*), hesitators (e.g. *uh, erm*), polite-speech formulae (e.g. *thanks, sorry*) and expletives (e.g. *shit, geez*) (1999: 1082–1099).

The annotation of inserts in the corpora in the present study, UCOW and the SBC subset, proceeded in three steps. First, all occurrences of interjections were manually annotated (i.e. those classified as interjections in OED). This annotation ran parallel with the annotation of Biber’s (1988) 67 linguistic features and

was essentially done in an effort to assign a tag to every token.⁸⁶ Without a tag for interjections, approximately every twentieth to every tenth word would have been left unannotated in the texts (e.g. *oh, wow, hi, hello, hey, yah, no, uh, um*), even if certain interjections also received a tag, or two, from Biber's features (e.g. *well*, tagged as both adverb and discourse particle). After the annotation of Biber's (1988) features, and interjections, was complete, the second step was taken. In the second step, Biber et al.'s (1999) definition of inserts was used, which meant that approximately ten percent additional occurrences, in each corpus, were found to belong to the category, all words that rightfully had been assigned Biber tags (and that, naturally, also keep those, e.g. *well*). In the third and final step, all interjections were renamed "inserts" and the total occurrences were summed up. The number of inserts per thousand words in the three corpora is shown in figure 4.15 (based on table 4.7). No equivalent annotation of inserts was carried out for writing, or speech overall. Unlike previous diagrams, the speech bar in figure 4.15 thus represents face-to-face conversations from the SBC subset only.⁸⁷

Table 4.7: Frequencies of inserts

	Writing	ACMC	f-t-fSBC	SCMC	SSCMC
inserts	n.a.	n.a.	60.0	97.3	66.6

Table 4.8: Frequencies of emotives

	Writing	ACMC	Speech	SCMC	SSCMC
emotives	0.0	n.a.	0.0	23.4	3.8

Figure 4.15: Inserts (normalized freq.).

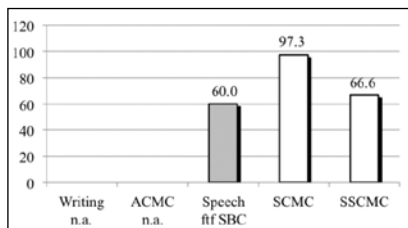
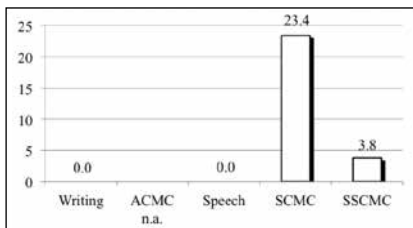


Figure 4.16: Emotives (normalized freq.).



The annotation of "emotives" (a new linguistic feature, introduced in the present study; see section 1.5) was also begun alongside the annotation of Biber's (1988) features, but completed after the annotation of all inserts. The new linguistic

86 Several tokens (such as abbreviations and contracted words), of course, were assigned two or more tags.

87 The results of statistical tests of the frequencies in the relevant media are found in Appendix VI. In tables 4.7 and 4.8 and figures 4.15 and 4.16, "n.a." means that the figure is "not available," as the texts were not annotated for the feature.

feature assigned tags to a few more tokens otherwise ignored, ultimately rendering practically all tokens bestowed with tags. Emotives are items typically found in conversational writing whereby chatters add an emotional zest to their utterances, e.g. :), ;), :(, :P, lol, rofl, lmao (partly taken up as emoticons or smileys in the previous literature; see e.g. Werry 1996, Jonsson 1998, Schulze 1999, Mar 2000, Crystal 2001, 2011a, Ooi 2002 and Baron 2008). Emotives thus comprise both emoticons *and* the initialisms that typically denote the sentiment in which an utterance is produced or intended to be received. Both emoticons and such sentiment initialisms illustrate chatters' intention to ensure that their message, produced on the fly, is correctly interpreted by the recipient. The number of emotives in the corpora is shown in figure 4.16 (based on table 4.8). No figure for the corpus of APMC is available. It was mentioned in the discussion of initialisms, in section 4.5, that all abbreviations in the chatted corpora were annotated for their constituent linguistic items (*idk*, for instance, was tagged with Biber's (1988) features nos. 3, 6, 56, 59 and 67). The initialisms that constitute emotives, however, did not receive this treatment, but rather were assigned the emotive tag only. Emotives will be discussed further, shortly.

Inserts and emotives can both be regarded as operators within the interpersonal metafunction, the tenor of communication, enacting social relationships. Previous sections of the present chapter explored how interpersonal meaning is carried lexico-grammatically by modal auxiliaries and personal pronouns, but also by e.g. markers of mood (WH-interrogatives) and negation (Halliday 1978, 1985a, Halliday & Hasan 1989, Halliday 2004), all part of the modality system of language. In the present section we will explore the ways in which inserts and emotives also, among other things, serve as lubricants in the social machinery.

Hodge & Kress (1988) introduce their discussion of the modality system of language thus:

In every day communication it manifestly matters a great deal what weight we are to attach to an utterance. A statement may be said emphatically, without qualifications, and we know that we are being asked to believe that it is true. Or it may be hedged with 'I think', 'it may be that'. Perhaps it is spoken with a rising intonation like a question, and we know that a speaker is offering the statement more tentatively. Or it may be said with a laugh or an ironic sarcastic tone, and we know that the speaker does not believe in the statement at all. (Hodge & Kress 1988: 121)

Inserts comprise discourse markers and hesitators, which, like the hedges Hodge & Kress mention, construct relations between the communicating parties, signalling their tentative, pending attitudes to messages. Emotives modalize utterances

by indicating the tone in which a “prosodic” unit might be read. Modality is at play in the semiotic act of chatting, as well as in face-to-face interaction, and inserts and emotives may be regarded as important carriers of modality in synchronous CMC, reflecting the tenor of the communication. The primary focus of the present section, however, is not to bear out the modality status of these features, but rather to point to their salience in conversational writing and to contrast their distributions in the annotated corpora (that these features act within the modality system of language is merely background information, implicitly understood).

Biber et al. (1999) note that inserts “comprise a class of words that is peripheral, both in the grammar and in the lexicon of the language” (1999: 1082). They are “stand-alone words” that are generally unable to “enter into syntactic relations with other structures” (ibid.). Nevertheless, they tend to “attach themselves prosodically to a larger structure, and as such may be counted as part of that structure” (ibid.). The inserts found in the annotated corpora are exemplified in table 4.9, below, along with the “larger structures,” i.e. the turns, in which they appear. Inserts either stand alone in the corpora (and comprise a turn in themselves), or else typically introduce larger “prosodic” units. Whereas Biber et al. (1999) classify as interjections only inserts that have an “exclamatory function, expressive of the speaker’s emotion” (1999: 1083), inserts classified as interjections in OED are represented among all the insert types in table 4.9. The few additional inserts found in annotation step two mainly belong to the types “discourse markers” and “polite speech-act formulae.”

The quantitative distribution of each type of inserts in the three corpora investigated is largely depicted by the proportions of exemplified turns in table 4.9. As seen in figure 4.15, SCMC contains the greatest number of inserts; in fact, inserts rank as the third most prevalent linguistic feature in IRC, if seen from the perspective of Biber’s (1988) list (next to present tense verbs and nouns, cf. Appendix II table 1). Table 4.9 reveals the insert type to which the most abundant SCMC inserts belong: greetings and farewells. The abundance of greetings and farewells in SCMC fully accounts for the higher number of inserts in SCMC (i.e. in IRC) overall, as compared to the other annotated corpora. Approximately half of the inserts in IRC are greetings, farewells and attention signals. IRC communication is a textual cocktail party involving the circulation of dozens of participants who, at any given moment, enter and leave rooms, continually greeting each other, calling for attention or bidding each other farewell. Greetings are the most

common initiators of social contact in face-to-face situations and conversational writing alike. In the chat room environment, the initiators often incorporate the nickname of a new participant and serve to confirm that the participant entering the room has been noticed (Anglemark 2009). Biber et al. mention that greetings are usually “reciprocated in a ‘symmetrical’ exchange” (Biber et al. 1999: 1085). In IRC, the reciprocation is not symmetrical (if it was, the quantity of greetings would be intolerable). The split-window ICQ communication contains symmetrical exchanges of greetings and farewells, although considerably fewer than IRC as each ICQ conversation for its full duration here involves only two participants (one conversation involves three). The SBC subset face-to-face conversations contain no greetings or farewells exchanged between informants; the instances found are reported speech. Apart from the disproportion of greetings and farewells, inserts are distributed fairly equally in the three corpora (see table 4.9), except for response forms and hesitators, which appear to be more common in split-window ICQ than in IRC.

The largely similar distribution of inserts in the three corpora makes a strong case for conversational writing as regards orality. Chatters, like face-to-face conversationalists, express emotional involvement by way of interjections. They readily accept the effort it takes to not just produce the conventionalized *oh* and *ah*, but also to create phonological spelling equivalents of other exclamations; see examples in table 4.9 (a finding echoing Ooi’s in 2002). Interjections convey chatters’ and conversationalists’ intensity of feeling alike: their surprise, their sympathy, their laughter – as well as their disgust, and their pain, among other sentiments. Chatters use slightly fewer discourse markers than oral conversationalists, for the reasons adduced in section 4.3 (with regard to Biber’s 1988 discourse particles). The discourse markers used, however, just like in speech, signal transitions in the evolving conversations, as well as “an interactive relationship between speaker, hearer, and message” (Biber et al. 1999: 1086).⁸⁸

88 Biber et al. (1999) suggest the finite verb formulae *I mean*, *you know* and *you see* as discourse markers, admittedly “open to debate” (1999: 1086), but these were not tagged as inserts in the present study.

The chatters (mostly those in IRC) use slightly more attention signals than the SBC speakers, but fewer response elicitors. The paucity of response elicitors briefly reminds us of the textuality of the medium; whereas spoken turns may need repetition to be correctly overheard, turns in conversational writing linger long enough to be re-read. Response forms in conversational writing array themselves in approximately the same orthography as in transcribed face-to-face conversations, with canonical *yeah* overriding the less frequent *yes*, for instance, but differ with regard to backchannels. Speech includes a variety of vocalized sounds as backchannels (transcribed *mhm*, *uh huh*, *unhunh*, etc.), which are not found in the chats. Chatted response forms tend to array themselves in variants of *yeah*, even when used as backchannels and, as mentioned, they are more common in the split-window ICQ chats than in the IRC chats. In both media, response forms, including backchannels, nonetheless serve the same functions as in spoken conversations; they provide answers to yes/no questions, responses to statements, or simply signal feedback to the conversational partner that the messages are understood and accepted – all in order to further lubricate the social machinery and ensure that the communication is functioning well. Backchannels were also found in IM texts in Nuckolls' (2005) study, although fewer than in face-to-face conversations recorded in the same study.

Hesitators are “pause fillers, whose main function is to enable the speaker to hesitate, i.e. to pause in the middle of a message, while signaling the wish to continue speaking” (Biber et al. 1999: 1092). Hesitators are very common in the SBC subset, and interestingly, these “pause fillers” to some extent also occur in conversational writing, despite users' inability in the textual CMC media to audibly hold the floor over their conversational partners. Whereas, in IRC, the hesitator merely signals that the message required some contemplation from its producer, in ICQ, it potentially signals the interlocutor's intention to keep or take over the conversational floor. The higher frequency of response forms and hesitators in split-window ICQ, compared to IRC, thus indicates a certain supersynchronicity effect in ICQ. Just as in oral conversations, while the conversational partner is producing their turn, the ICQ chatters may interpose these inserts to signal simultaneously their understanding, puzzlement or intention to “speak,” whereas this is not possible in IRC. To investigate whether the higher incidence of response forms and hesitators is an effect of the supersynchronicity, however, would require a close examination of the overlapping sequences in ICQ, which unfortunately is unfeasible due to the varying quality of the video recordings of the split-window ICQ material at hand.

The penultimate type of insert, polite speech-act formulae, provides another interesting contrastive finding in the corpora. These inserts are used in conventional speech acts, such as thanking, apologizing and requesting, and are interestingly found to be much more common in IRC than in face-to-face conversations or split-window ICQ. Possibly, the IRC users' lack of acquaintance with each other, and their tentative, forming relationships, trigger a higher degree of politeness among users, a desire to appear polite. Finally, expletives are rare in all three corpora, with taboo expletives non-existent in the IRC chats (in the channels recorded, users were immediately "kicked" upon their use).

All in all, the use of inserts in conversational writing distinctly resembles the use in spoken conversations, both as regards quantity (except for the abundance of greetings and farewells in IRC) and as regards functional quality. Chatters are not just chatters, but also (presumably) experienced speakers and, to further their human relationships, they bring their conversational routines to bear on both social media alike (face-to-face as well as computer-mediated conversations). Inserts provide valuable links between utterances in both forms of social exchange. The distribution of inserts in the written genres remains to be expounded, but is expected to contrast sharply with the corpora annotated here, for which reason future studies of the variation among written, spoken and computer-mediated genres are encouraged to take inserts into account. Halliday (1985a) makes the point that "[t]he spoken language is every bit as highly organised as the written, and is capable of just as great a degree of complexity. Only, it is complex in a different way" (1985a: 87). Whereas written language is "static and dense," spoken language is "dynamic and intricate" (*ibid.*). The present study regards Halliday's (1985a) claim regarding speech equally applicable to conversational writing, and finds inserts to be some of the most central markers of this "spoken language" complexity.

Turning now to emotives, the first point to make is with regard to their "linguistic" status adopted here. Emotives in their current form have been around in the English language since the 1980s.⁹⁰ Common emoticons (e.g. :) ;(;) :-))

90 Scott Fahlman in 1982 suggested the use of :-)) and :-(in a messageboard (ACMC) exchange as "joke markers" and "to mark things that are not jokes," respectively, widely recognized as the original use of what later came to be called emoticons (see <<http://www.cs.cmu.edu/~sef/Orig-Smileys.htm>> for a clip of the original messageboard thread). Lol, meaning "laughing out loud" is claimed to originate from messages in a bulletin board system (ACMC) in the "early-to-mid-80s" (<<http://pages.cpsc.ucalgary.ca/~crwth/LOL.html>>) (cf. Morgan 2011).

and sentiment initialisms (e.g. *lol*), are used and understood by a wide Anglophone, and international, audience. In fact, in 2011, *lol* (meaning “laughing out loud”) entered into OED, as both interjection and noun, with the pronunciations /,ɛləʊ'ɛl/, /lɒl/ in British, and /,ɛl,ou'ɛl/, /lɒl/ in US English. Walther & D'Addario (2001: 329) state that “[a]lthough emoticons may be employed to replicate non-verbal facial expressions, they are not, literally speaking, nonverbal behavior.” They go on to explain that in face-to-face interaction a person may smile unconsciously, whereas in CMC “it is hard to imagine someone typing a :-) with less awareness than of the words he or she is selecting” (ibid.). Marvin (1995) similarly recognizes that smiles in face-to-face conversations can be strategic, spontaneous, or unintentional, whereas in SCMC (more specifically in the mode of MOO that she studied, a text-based online virtual reality system) every smile is consciously indicated: “a conscious choice must be made to type it out” (Marvin 1995: no page number available). Moreover, an SCMC participant “might frown at the keyboard” and yet “decide to type a strategic smile” (ibid.). An emoticon can thus be both strategic and spontaneous, but rarely unintentional (except as a slip of the key). Smileys are not just appended to statements that are ironic or ambiguous; they are also incorporated as “friendly gestures, indications of approval or appreciation” (Marvin 1995: no page number available), much like smiles in face-to-face interaction.

The conscious typing of emotives in the conversational writing corpora in the present study yields a nearly finite set of types, almost as if they belonged to a closed grammatical class. On the other hand, individual emoticons display something like morphological inflections, as :((((((((is a variant of :(. Emotives are at once paralinguistic (indicating the tone of the utterance) and linguistic, constituting tokens in their own right (usually set apart from other words orthographically). They resemble other paralinguistic features of chat (like repeated exclamation marks appended to words for emphasis), but are not appended to other words – rather, more like inserts, stand-alone words or appended to prosodic units, like the laughter particles identified as inserts above (e.g. *hehe*). On the other hand, emotives do not lend themselves easily to phonology; only pronounceable ones (*lol* and *rofl*) have crossed over into speech and thereby become lexicalized. CMC studies to date have typically regarded emoticons and sentiment initialisms as paralinguistic features of the communication, substituting for the lack of non-verbal cues (e.g. Dery 1993, Thompson & Foulger 1996, Werry 1996, Schulze 1999, Derks et al. 2007, Waldner 2009). Crystal (2001),

however, hesitates to call emoticons paralinguistic, emphasizing that “they have to be consciously added to a text” (2001: 34). Dresner & Herring (2010) also extend the function of emoticons beyond substituting for non-verbal cues, construing them as “textual indicators of illocutionary force” (2010: 260). The present study recognizes the paralinguistic denotation of emotives; they are chatters’ own ways of transcribing their “speech.” The present study, nevertheless, is an investigation into the variation between genres of writing and speech, and such a study needs to recognize every token of the texts. Once- or twice-occurring graphological tokens, like `_,-^*^*-,__,-*’` (see example 28 in section 4.5), are easily dismissed as hapax legomena or as void of meaning, whereas emotives carry modal meaning and can be expected to recur in texts. After all, the most common ones have recurred in texts for thirty years, to date. It is, consequently, high time that emotives be given linguistic status as markers of CMC discourse. In variation studies, they effectively set computer- and cellphone-mediated texts apart from other texts, and thus, clearly, constitute a linguistic feature to take into account in future multidimensional studies of the variation of the English language. The remainder of this section presents the distribution of emotives in the annotated conversational writing corpora.

Recall from figure 4.16 that SCMC (that is, IRC) contains far more emotives than SSCMC (that is, split-window ICQ). In IRC, emotives are the ninth most common feature, more common than for instance past tense verbs, third person pronouns and pronoun *it*. *Lol* is the predominant marker of emotional involvement in both modes of CMC; in IRC *lol* accounts for 56 percent of all emotives, in ICQ it accounts for as much as 73 percent. In spite of this, the use of *lol* is much more rare in split-window ICQ than in IRC. The distribution of the individual emotives in the conversational writing corpora is illustrated in figure 4.17, detailing their overall distribution from figure 4.16, per thousand words.


```

<Cheeky1>      lol
<Cheeky1>      i will be back
<|mad_max|>    ok .....
<|mad_max|>    take care
<Cheeky1>      gotta go for 5 minutes
<Cheeky1>      u 2 max sweety
<Cheeky1>      c ya in a sec u hunk of spunk
<Cheeky1>      hehehe
<|mad_max|>    c u
<Cheeky1>      cya
<Cheeky1>
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
<Cheeky1>      :0)
<USA__MALE>    hi

```

Internet relay chat text 3a (UCOW)

One and the same user (Cheeky1), first signaling his/her appreciation of a fore-going joke about women, and later signaling his/her continued sympathetic presence, produces all *lol*-turns in the example. At the end of the example, the user announces his/her exit, and before leaving flashes a brief :0) “grin.” Stand-alone emotives (constituting a turn in themselves) are found in both IRC and split-window ICQ, but in IRC, stand-alone *lol* appears to function more often as a jovial presence marker than as a transcription of actual laughing. Chatters in IRC are in initial stages of contact and are concerned about appearing congenial. *Lols* and smiles are therefore sprinkled into the IRC conversations much as friendly smiles would be in face-to-face first encounters. Such use of emotives seems to account for much of the discrepancy in the emotives distribution between the two modes of CMC. In the split-window ICQ chats (SCMC, right bars in figure 4.17), the *lols* seem more co-textually motivated, as for instance in the amusement 10 expresses over the comment J makes about his sister in example (54).

(54) <J> to tell the truth.. i dont think i've ever seen my sister go 10
feet away from the shore.. let alone anywhere else in a big body of water
<10> lol

Split-window ICQ chat text 9 (UCOW)

In both modes of CMC, initialisms appended to turns appear in both initial and end positions, with a few rare instances in medial position. Emoticons in IRC turns appear in medial and end positions, whereas in split-window ICQ they are exclusively appended at the end. It seems as though IRC chatters are more concerned than the ICQ chatters to set the tone of their message as early as possible. Smiles in both media represent friendly smiles more than laughter, and in IRC

they are typically appended to turns close to greetings and farewells; see the various IRC-turns in (55). The winking smiley is, surprisingly, more common in IRC than in split-window ICQ, possibly because in IRC it also appears at the end of greetings and farewells as in the last two turns in (55). The winking smiley otherwise prototypically signals tongue-in-cheek comments, as in (56), and given the ICQ chatters' previous acquaintance, they could be expected to use them more.

- (55) a. hi again rainman19 :)
 b. puck....hello to you too...:))
 c. AdamSxy35 :)
 d. REVOLI, Im fine, how are you? :)
 e. Raha,take care, bye :)
 f. hiya CityWoman and y'all;)
 g. Ta ta Adam...;)

Internet relay chat (UCOW)

- (56) <AdamSxy35> oups why dont you try a business chat room on yahoo?
 <_oups> hm...well do they have that..
 <AdamSxy35> it works for me when i cant fall asleep ;)

Internet relay chat text 5b (UCOW)

In general, the IRC corpus displays a wider emoticon repertoire than the split-window ICQ corpus. IRC chatters are presumed to be experienced emotives-users, and often thought to proliferate emoticons. In a large-scale emoticons study, however, Schulze (1999) plays down the need for smiley dictionaries. His 28,345 "line" long IRC chat corpus contains no more than eight major types of emoticons (with several minor variations) (1999: 76). Ten years later, Waldner (2009) finds no more than 15 emoticons used regularly in IRC (2009: 81). The IRC corpus in the present study is about ten percent the size of Schulze's, but can be said to proportionally agree with his findings. Out of Schulze's (1999) eight major types of emoticons, the present study finds representatives of four: the "smiley" :), the "frowney" :(, "sticking out tongue" :P and "slight frown" :/, but also two additional major types: the "winking smiley" ;) and the "indifferent" one :I, i.e. altogether six types. In the split-window ICQ corpus, the emoticon repertoire is even more limited, with representatives of only four major types.

A few writers have investigated linguistic gender differences in computer-mediated communication, putting e.g. Coates (1993) and Tannen's (1990, 1994) findings of gender-differentiated conversational styles to the test on empirical CMC data. Herring (1996b) finds women and men to present different styles of interaction and information exchange on two Internet mailing lists (ACMC), styles that she terms the "aligned variant" (supportive, mostly used by women)

and the “opposed variant” (more insulting or aggressive, mostly used by men). Echoing this finding, Herring (1998, 2003) notes that, in SCMC, women type three times as many representations of smileys or laughter as do men. Wolf (2000) finds women to use more emoticons in same-gender newsgroups (ACMC), but finds no significant difference between women’s and men’s use in mixed-gender newsgroups. Baron (2004) describes a study of instant messaging (IM) data, collected among college students, in which she found differences e.g. in women’s and men’s use of emoticons (women used more) and contractions (men used more). Replicating these studies on the UCOW IRC data is not feasible, as no record of the Internet relay chatters’ gender exists, but for the ICQ data a comparable investigation yields interesting results with regard to emoticons (no comparable investigation was carried out on contractions).

Baron’s (2004) IM corpus is approximately the same size as the UCOW split-window ICQ corpus and thus comes in handy for a comparison. A total of 49 emoticons were used in Baron’s data. Females were found to be the prime users of emoticons; out of the 16 female participants three-quarters used one or more emoticons. Of the 6 male participants only one used emoticons (2004: 415). The results for the comparable analysis of the UCOW split-window ICQ data are presented in the first row of table 4.10.

Table 4.10: *Individuals’ emotives usage in the split-window ICQ corpus, by gender; f=female (7), m=male (18). N.B. raw figures*

	f f f f f f f	m m m m m m m m m m m m m m m m m m m
Emoticon	3 0 0 0 0 0 0	2 2 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Sentiment initialism	0 0 0 0 1 2 8	0 2 2 0 0 0 0 0 0 0 0 0 0 0 1 1 2 2 3 4

A total of nine emoticons are used in the split-window ICQ corpus. *Males* are here found to be the prime users of emoticons; out of the 18 male participants, four used emoticons, whereas out of seven females, only one did the same. All the while, however, 28 sentiment initialisms were used (all of them *lol*, except one *lmao*; see the second row in table 4.10); 43 percent of the females used sentiment initialisms, and 44 percent of the males. About half of the males used emotives overall, whereas fewer than half of the females did. In other words, the findings for the UCOW split-window ICQ corpus do not corroborate the findings in Baron (2004) with regard to emoticon use. On the other hand, the average number of emotives used by males in the ICQ corpus is only 1.3, whereas for females the same number is 2. One of the females produced the highest number

4.7 Chapter summary

The present chapter has expounded on the salient features in conversational writing. The bulk of the chapter zoomed in on the ten linguistic features which, in either mode of CMC, synchronous or supersynchronous, or in both, deviate from Biber's (1988) mean for spoken and written language overall by more than two standard deviations. The chapter set out from two of the important carriers of interpersonal meaning in language: modal auxiliaries and personal pronouns, the latter of which reveals salient traits in the chatted texts, the pervasive first and second person pronouns. Next, the lexical properties of the conversational writing genres, writing, and speech were investigated, through the employment of contrasted measurements of word length, type-token ratio and lexical density, essentially revealing the latter to be most appropriate for capturing the grammatical intricacy (or lack of lexical density) of the chatted texts. The fourth section presented the salient features annotated in the corpora and what each of them reveals about the communication, as regards, quantity, quality, orality and Halliday's metafunctions, most notably about the tenor of the discourse. Paralinguistic cues and extra-linguistic features were then surveyed, which further incorporated consideration of the textual and ideational metafunctions (the semiotic mode and field of the discourse). Finally, the last section proposed two linguistic features to be incorporated into future accounts of the variation of the English language, inserts and emotives, which both serve important functions in computer-mediated communication. In the next chapter, the more granular, yet all-round, results of the application of Biber's (1988) methodology will be presented and the positions of the conversational writing genres on Biber's dimensions of linguistic variation revealed. It will be seen there, that most of the salient linguistic features presented above load on one and the same dimension of variation (Dimension 1), distinguishing involved production from informational. As mentioned in chapter 3, inserts and emotives are not linguistic features in Biber's (1988) methodology and consequently have no bearing on the dimension scores to be presented in chapter 5 (except for a few items contained within inserts, which are also tagged in Biber's methodology). The present chapter subsumed numerous written genres into mean figures for writing, and numerous spoken genres into mean figures for speech. In chapter 5, these two multitudes of genres will be split up to further diversify the contrastive analysis, but in the end to make for a unified picture as regards the nature of synchronous and supersynchronous computer-mediated communication.

Chapter 5. Conversational writing positioned on Biber's (1988) dimensions

5.1 Introductory remarks

In the previous chapter, the most salient features of conversational writing were narrowed down and their distributions in traditional writing, speech and a genre of asynchronous computer-mediated communication were contrasted. The chapter explored, for instance, the distribution of modal auxiliary verbs, inserts and emotives, and surveyed a number lexical and paralinguistic features of conversational writing. In the present chapter, we turn to investigating the overall lexico-grammatical patterns found in the corpus of conversational writing (UCOW). The purpose of the chapter is to position the UCOW genres "Internet relay chat" (IRC) and "split-window ICQ chat" on Biber's (1988) dimensions of language variation (cf. Conrad & Biber 2001a), taking into account all of Biber's 67 linguistic features (see table 2.1 for a complete list of these). For updated reference as regards face-to-face conversation, the subset sampled in the present study (the SBC subset) from the Santa Barbara Corpus of Spoken American English part 1 (SBC, recorded in the 1990's) will also be studied and positioned on the dimensions, but reference will no longer be made to the corpus of ACMC (since, as mentioned in section 4.1, only feature count data is available for the ACMC corpus, not the comprehensive raw texts).

In section 3.5 of chapter 3 (Material and method), the procedure for calculating dimension scores was presented. To recapitulate, dimension scores were computed for each text in the UCOW (10+12 texts) and the SBC subset (14 texts). First, frequencies of all the linguistic features were recorded for each text and normalized to text lengths of 1,000 words. The normalized frequencies for IRC's 10 texts, split-window ICQ chat's 12 texts and the SBC subset's 14 texts are documented in Appendix II (tables 5–7). Next, a table of descriptive statistics for each of the genres was compiled; see Appendix II (tables 1–3). The frequencies of the linguistic features in each text, and the mean frequencies in each genre, were then contrasted with the mean frequencies of Biber's corpus as a whole (all 23 genres), summarized in Appendix II table 4. All frequencies were standardized to conform to a single scale, i.e. a mean of 0.0 and a standard deviation of 1.0, before the dimension scores were computed. Table 5.1 here summarizes the resulting dimension score statistics for the genres of principal concern in the present chapter: Internet relay chat, split-window ICQ and face-to face conversations

from the SBC subset (also in this chapter called “face-to-face conversations SBC”). An analogous summary of dimension score statistics for each of Biber’s genres is found in Appendix VIII.

Table 5.1: Descriptive dimension statistics for the UCOW genres and the SBC subset

Conversational writing (UCOW)	Dimension	Mean	Minimum value	Maximum value	Range	Standard deviation
Internet relay chat	Dimension 1	25.6	14.4	35.9	21.5	7.1
	Dimension 2	-4.2	-5.9	-1.9	4.1	1.4
	Dimension 3	-4.7	-8.3	-0.9	7.4	2.5
	Dimension 4	-2.6	-7.5	0.1	7.6	2.4
	Dimension 5	-3.9	-4.8	-0.9	3.9	1.2
	Dimension 6	-3.5	-4.5	-1.0	3.5	1.0
Split-window ICQ chat	Dimension 1	47.2	19.3	66.4	47.0	13.3
	Dimension 2	-2.2	-3.6	2.1	5.7	1.5
	Dimension 3	-4.1	-6.3	-1.5	4.8	1.5
	Dimension 4	0.2	-1.8	3.5	5.4	1.7
	Dimension 5	-3.3	-4.8	1.3	6.1	1.7
	Dimension 6	-1.9	-3.9	0.2	4.1	1.3
<hr/>						
Speech (SBC subset)	Dimension	Mean	Minimum value	Maximum value	Range	Standard deviation
Face-to-face conversations	Dimension 1	43.7	9.1	63.0	53.9	14.9
	Dimension 2	-0.6	-4.9	5.7	10.6	3.0
	Dimension 3	-2.4	-5.2	2.1	7.3	2.0
	Dimension 4	-1.3	-6.5	4.0	10.5	3.3
	Dimension 5	-3.3	-4.8	2.3	7.1	1.9
	Dimension 6	-0.2	-3.1	6.9	10.0	2.6

Dimension 1: Informational versus Involved Production

Dimension 2: Narrative versus Non-Narrative Concerns

Dimension 3: Explicit/Elaborated versus Situation-Dependent Reference

Dimension 4: Overt Expression of Persuasion/Argumentation

Dimension 5: Abstract/Impersonal versus Non-Abstract/Non-Impersonal Information

Dimension 6: On-Line Informational Elaboration

The sections below will present the dimension score statistics for all genres on Biber’s six dimensions as graphic adaptations of the dimension figures found in Biber (1988: 128ff). Each dimension graph plots the genres of conversational

writing and face-to-face conversations SBC in relation to the 23 genres Biber studied (17 genres of writing and 6 of speech; see Appendix I). The plotting of the new genres (Internet relay chat, split-window ICQ chat and face-to-face conversations SBC) on Biber's dimensions follows the tradition developed in numerous post-Biber variation studies in that new genres are positioned in relation to Biber's established genres' dimension scores (Appendix VIII) without being part of the calculation of their mean⁹² or the conception of the dimensions in the first place,⁹³ i.e. without the application of a new factor analysis (cf. Conrad & Biber 2001a: 41, 43–183, Biber 2008: 844). The reader is presumed to be familiar with this tradition when interpreting the dimension plots.

Table 5.2 summarizes the results of an analysis of variance (ANOVA) carried out among the new genres “Internet relay chat,” “split-window ICQ chat” and “face-to-face conversations SBC,” and the results of Biber's tests among his genres (from Biber 1988: 127). As dimension scores for Biber's individual texts are unavailable, tests of all genres in combination were not carried out. Pairwise t-tests among the new genres were performed with respect to the dimensions for which the ANOVA returned significant differences (Dimensions 1, 2 and 6), in order to establish which genres differ. Table 5.3 reports the p-values from the t-tests. For dimensions on which the ANOVA yielded no significant differences, no t-test was carried out and, consequently, no p-values are given in table 5.3; instead, the genres on those dimensions are indicated as not significantly different, “n.s.”

92 This standpoint was taken after a trial inclusion of UCOW and the SBC subset in the calculation of new means for an amalgamated corpus consisting of UCOW, the SBC subset and Biber's 23 genres, which yielded essentially comparable scores. The UCOW and the SBC subset together consist of approximately 30,000 words (see table 3.1), whereas Biber's established genres total approximately 960,000 words (see Appendix I). The resulting scales altered the zero point marginally, but in all significant respects genres kept their ordinal positions and relative distances on the dimensions.

93 As mentioned in footnote 92, the size of the corpora representing the new genres (UCOW and SBC subset) is small, meaning that the inclusion of the corpora in a new factor analysis of the spoken and written texts studied here might only very marginally alter the layout of dimensions. This notwithstanding, such an effort is naturally both feasible and commendable if ventured into, in future studies, with regard to new corpora of writing and speech.

Table 5.2: Results from ANOVA among the new genres and from Biber's (1988: 127) tests among his genres. (The p-values for the new genres have been multiplicity adjusted)

IRC, split-window ICQ chat and face-to-face conversations SBC				Biber's 23 genres (Biber 1988: 127)		
	F-value	Probability (p)	R-squared	F-value	Probability (p)	R-squared
Dimension 1	9.0	p=0.0208	31.3%	111.9	p<0.0001	84.3%
Dimension 2	7.7	p=0.0432	27.7%	32.3	p<0.0001	60.8%
Dimension 3	4.7	p=0.2788	17.3%	31.9	p<0.0001	60.5%
Dimension 4	3.2	p=0.7826	11.0%	4.2	p<0.0001	16.9%
Dimension 5	0.5	p=1.0000	-3.0%	28.8	p<0.0001	58.0%
Dimension 6	9.3	p=0.0006	32.2%	8.3	p<0.0001	28.5%

Table 5.3: Results from t-tests among the new genres. Values for probability (p)(no multiplicity adjustment needed)

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6
Internet relay chat vs. split-window ICQ chat	0.0001	0.0039	n.s.	n.s.	n.s.	0.0037
Internet relay chat vs. face-to-face conversations SBC	0.0008	0.0009	n.s.	n.s.	n.s.	0.0004
split-window ICQ chat vs. face-to-face conversations SBC	0.5333	0.1081	n.s.	n.s.	n.s.	0.0420

The dimension score of a new genre essentially marks the genre's position relative to the mean of all of Biber's spoken and written genres (which constitutes the zero point of each dimension scale). The dimension scores may assign a genre to the positive or negative end of a dimension scale, but more important than the absolute dimension score is the genre's position relative to neighboring genres and opposing genres, as will become apparent in the presentation of the dimension plots (figures 5.1a through 5.6b) and associated textual examples. Biber (1988: 129) notes that a proper interpretation of a dimension entails consideration of 1) similarities and differences among the genres, 2) the linguistic features constituting the dimension and 3) the underlying functional and situational parameters associated with the dimension. The graphic presentation of each dimension scale will thus be followed by a discussion of sample texts in the new genres and contrastive genres, with reference to features constituting the dimension, both positive and negative, and to the communicative functions they serve.

A summary of Biber's factorial structure is found in table 5.4 (adapted from Biber 1988: 102–103, repeated here from section 2.3 for convenience). Please recall from the presentation of Biber's (1988) dimensions, section 2.3, that underlying each dimension are the combined sets of features, i.e. the co-occurrence patterns that reflect underlying communicative functions. For Dimensions 1 and 3, the sum of the standard scores of features with negative loadings (features below the dashed line) has been subtracted from the sum of the standard scores of features with positive loadings to obtain a dimension score; for all other dimensions, the standard scores of relevant features have simply been added up. The respective loads of features in table 5.4 were not included in the calculations, other than as indicators of which features to add up and which to subtract, in order to produce the dimension score for each genre.

The genres of focal concern in the present chapter are the conversational writing genres Internet relay chat and split-window ICQ chat, contrasted with Biber's (1988) numerous genres of writing and speech, as well as with face-to-face conversations SBC. APMC will not be plotted on the dimension graphs nor discussed beyond this point in the present chapter, since, as mentioned, the unavailability of comprehensive raw APMC texts renders further textual analysis unfeasible. Nevertheless, one important disclosure with regard to Collot's APMC corpus deserves to be made here, as it pertains to dimension scores.

In chapter 4, Collot's feature counts for the "ELC other" corpus of BBS conferencing were brought in, for contrastive purposes, to represent APMC (see section 4.1 for an introduction to the corpus). In her 1991 study, Collot applies Biber's (1988) methodology to compute dimension scores for the APMC corpus, reportedly based on the standard scores computed, which Collot calls FDS, "feature deviation scores" (1991: 73, results also presented in Collot & Belmore 1996). Upon studying Collot's (1991) dimension scores, however, the present author found a considerable mismatch between the dimension scores and the constituent standard scores. As a result of the mismatch, Collot's (1991) dimension scores for the "ELC other" corpus fail to adequately represent the "ELC other" genre of BBS conferencing on several of Biber's (1988) dimensions, most notably on Dimension 1. To remedy this situation, a new computation of dimension scores for BBS conferencing was carried out in the present study, based on Collot's (1991: 69–70) "feature deviation scores" for the "ELC other" corpus. Table 5.5 presents the results of the new calculation.

Table 5.4: Summary of co-occurring features on each dimension (Biber 1988: 102–103)

Dimension 1		Dimension 3	
private verbs	0.96	WH relatives: object position	0.63
THAT deletion	0.91	WH relatives: pied pipes	0.61
contractions	0.90	WH relatives: subject position	0.45
present tense verbs	0.86	phrasal coordination	0.36
second person pronouns	0.86	nominalizations	0.36
DO as pro-verb	0.82	time adverbials	-0.60
analytic negation	0.78	place adverbials	-0.49
demonstrative pronouns	0.76	adverbs	-0.46
emphatics	0.74		
first person pronouns	0.74		
pronoun IT	0.71	Dimension 4	
BE as main verb	0.71	infinitives	0.76
adverbial subordinator – cause	0.66	prediction modals	0.54
discourse particles	0.66	suasive verbs	0.49
indefinite pronouns	0.62	adv. subordinator –condition	0.47
hedges	0.58	necessity modals	0.46
amplifiers	0.56	split auxiliaries	0.44
sentence relatives	0.55		
direct WH-questions	0.52		
possibility modals	0.50	Dimension 5	
non-phrasal coordination	0.48	conjuncts	0.48
WH clauses	0.47	agentless passives	0.43
stranded prepositions	0.43	past participial clauses	0.42
nouns	-0.80	BY passives	0.41
word length	-0.58	past participial WHIZ deletions	0.40
prepositional phrases	-0.54	adverbial subordinator –other	0.39
type/token ratio	-0.54		
attributive adjectives	-0.47		
		Dimension 6	
Dimension 2		THAT verb complements	0.56
past tense verbs	0.90	demonstratives	0.55
third person pronouns	0.73	THAT relatives object position	0.46
perfect aspect verbs	0.48	THAT adjective complements	0.36
public verbs	0.43		
synthetic negation	0.40	Dimension 7	
present participial clauses	0.39	SEEM/APPEAR	0.35

On Dimension 1, the correct dimension score (25.3) positions the “ELC other” genre of ACMC considerably closer to face-to-face conversations than what Collot (1991: 77) and Collot & Belmore (1996: 22) imply (which is c. 8.5). On other dimensions, the scores in table 5.5 represent less grave adjustments of Collot’s results, although significant enough to warrant their documentation here. On Dimensions 4 and 6, Collot’s (1991) dimension scores appear properly computed. Interested readers are referred to table 5.5, Collot (1991) and Collot & Belmore (1996) for further contrastive analysis. The present chapter now turns to the genres of synchronous and supersynchronous CMC and their positions on Biber’s (1988) dimensions of linguistic variation.

Table 5.5: Corrected dimension scores for the “ELC other” corpus of BBS conferencing presented in Collot (1991) (“n.c.” means that a corrected value was not calculated)

Asynchronous CMC (ELC other)	Dimension	Mean	Minimum value	Maximum value	Range	Standard deviation
BBS conferencing	Dimension 1	25.3	n.c.	n.c.	n.c.	n.c.
	Dimension 2	-2.3	n.c.	n.c.	n.c.	n.c.
	Dimension 3	0.4	n.c.	n.c.	n.c.	n.c.
	Dimension 4	2.1	n.c.	n.c.	n.c.	n.c.
	Dimension 5	4.7	n.c.	n.c.	n.c.	n.c.
	Dimension 6	1.8	n.c.	n.c.	n.c.	n.c.

5.2 Dimension plots

In the subsections below, the positions of the conversational writing genres “Internet relay chat” and “split-window ICQ chat” and of “face-to-face conversations SBC” (denoting the SBC subset annotated in the present study) are plotted on Biber’s (1988) six dimensions of linguistic variation (alongside Biber’s 17 genres of writing and 6 genres of speech) and discussed – one dimension per subsection (the first three genres based on table 5.1, and Biber’s 1988 genres on Appendix VIII). To follow up on the bar charts in chapter 4 (e.g. figure 4.4), the written genres are plotted in black in the dimension graphs of this chapter, the spoken genres in gray, and the conversational writing genres in white; see e.g. figure 5.1a. The black and gray dots may thus be regarded as the granular follow-up of the bars for writing and speech in chapter 4, respectively, whereas the white dots were each represented by their own bar in the chapter 4 figures.

The dimension plots in this chapter are all drawn from Biber (1988), but have been graphically adapted to make room for the new genres alongside Biber's (1988) great number of established genres, in the format of the present publication. As in Biber (1988), the interpretation of a genre's position always focuses on the scale of the y-axis, that is, the vertical position of the genre (Biber's 1988 dimension plots accordingly only have vertical axes). To afford room for the multitude of genres here, however, the plots are slightly slanted; see e.g. figure 5.1a. This means that the genres are plotted in ordinal sequence one step apart on the x-axis, even though the x-axis, of course, is insignificant for the interpretation of a dimension and, consequently, has no scale.

5.2.1 Dimension 1: Informational versus Involved Production

The first of Biber's dimensions is labeled "Informational versus Involved Production" (Biber 1988: 107). In the adapted dimension scale, figure 5.1a, the genres are plotted in ordinal sequence one step apart on the x-axis ranging from most informational at the bottom left end to most involved on the upper right (adapted from Biber 1988: 128).

Although most variationists after Biber heed the admonition that there is no overall absolute difference between writing and speech, they nevertheless agree that the first of Biber's dimensions reflects a near-perfect literate vs. oral dichotomy (the dimension scores of written genres are plotted in black and spoken genres in gray in figure 5.1a). New in the Dimension 1 plot here, however, are the genres of conversational writing (Internet relay chat and split-window ICQ chat, plotted in white and with labels in capital letters) and the face-to-face conversations SBC genre (in gray).

The conversational writing genres score well into the involved end of the Dimension 1 scale, with split-window ICQ chat exhibiting a score beyond all other genres. The following text example, (1), taken from the split-window ICQ chat corpus, illustrates the patterns found in texts with very high dimension scores on Dimension 1. Several of the salient features in the text were explored in section 4.4 (in fact, as many as nine of the ten most salient features found in conversational writing in section 4.4 are features that loaded on Dimension 1 in Biber's (1988) study).⁹⁴

94 Of the most salient features in conversational writing explored in section 4.4 (see table 4.6), eight are "positive" features on Dimension 1 (first and second person pronouns, direct WH-questions, analytic negation, demonstrative and indefinite pronouns, present tense verbs, predicative adjectives and contractions) and one is "negative" (prepositional

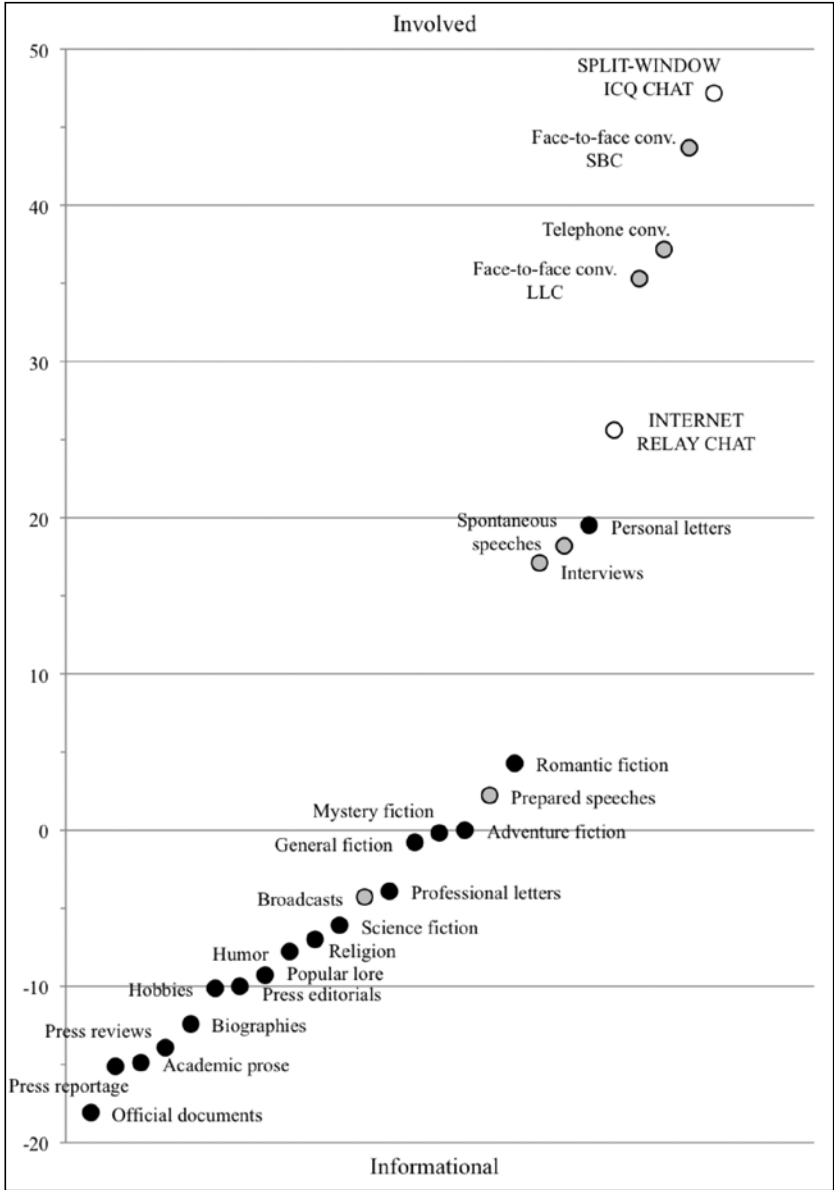
- (1) <A> there is no reason to hurt me just cause i'm attractive

 nah, it's more that you have that insecure charm
<A> hehe i know i do it so well
<A>
 girls think your pathetic, and thus flock to love you
<A> hey i'll let you know that almost all of tamis friends think
i'm hot
 you know i never saw that movie
<C> Fuck B stool my font
 heard it was really twisted
<A> stole
 no, you stole it from me off AIM
<A> C learn how to spell
<C> i wad this font on my odl comp
 i've had it since i started using AIM
<C> damn
 and then figured out hoe to do the font change thingy
<A> see thats why i have a black font

Split-window ICQ chat text 11 (UCOW)

phrases). The tenth most salient feature, predicative adjectives, did not load significantly on any dimension in Biber's (1988) study.

Figure 5.1a: Mean scores on Dimension 1 for all genres (capitalization denotes conversational writing). Dimension 1: "Informational versus Involved Production" (adapted from Biber 1988: 128).



Example (1) displays intense personal involvement among the three supersynchronous conversation participants. The split-window ICQ communication in UCOW is highly interactive and affective, and many of its high-frequency features belong to categories with strong positive weights on Dimension 1. Private verbs abound (*know, think, saw, heard, learn, see*), and several of them are followed by subordinator-THAT deletion (e.g. *i know Ø i do it so well*,⁹⁵ *you know Ø I never saw, girls think Ø your pathetic, friends think Ø I'm hot, heard Ø it was really twisted*). Contractions are the norm (*i'm, it's, your, i'll, i've, thats*), with *our* without apostrophe or standard spelling. Most verbs are in the present tense (*is, -'m, -'s, have, r, flock, let, -'ve, s, have*) like the private verbs. Possibility modals (such as *can*) occur more frequently in split-window ICQ than in any other genre. Proverb DO frequently substitutes more elaborated constructions (*do it so well, do the font change thingy*) and general emphatics add force and mark certainty (*just, more, really*). The supersynchronous texts bristle with first and second person pronouns (*me, i, my, you, your*), but the most conclusive contribution to the high dimension score is brought by direct WH-questions (3.9 per thousand words, compared to 0.2 per thousand words in Biber's corpus as a whole). All of these linguistic features together, contrasted with the sparsity of features with negative weights (infrequent use of nouns, prepositional phrases and attributive adjectives), yield a mean dimension score for split-window ICQ beyond all spoken genres (although the score is not significantly different from that of face-to-face conversations SBC, as indicated in table 5.3).

The spoken genres that come closest to split-window ICQ chat on Dimension 1 are face-to-face and telephone conversations. Worthy of notice among them is the face-to-face conversations SBC genre, which scores higher than both the LLC face-to-face and telephone conversation genres (a position also noted by Helt 2001 for American telephone conversations). As can be seen in example (2), the SBC texts have a strong interpersonal focus, and some are fairly intimate in character. They display frequent private verbs (*think, know, find, guess*), some of which are followed by subordinator-THAT deletion (e.g. *I don't think Ø I am*).

- (2) Nathan: ... Am I doing that right so far?
 Kathy: ... Mhm.
 Nathan: ... All the way down to that?

95 Identified by Biber's algorithm as containing a subordinator-THAT deletion (Biber 1988: 244), despite the possible alternative reading as two asyndetically coordinated main clauses. To attain results comparable with Biber's, his algorithms were closely observed at all times (Biber 1988: 222–245). Setting Biber's algorithm aside would have rendered incomparable results.

Kathy: ... Mhm.
 ... I think.

Nathan: ... I don't think I am.
 Do you?

Kathy: ... And you'd have to have that plus or minus.
 What.

Nathan: I don't know what I did to get that.
 Where did I get that square root of-
 um,
 ... ex squared.

Kathy: Because you brought this over here.
 ... You brought ... three over here.
 ... divided by three,
 and then you have ex squared,
 so if you want to find ex,
 you have the square root of ex squared.

Nathan: ... I guess all I can't figure out is,
 what the square root of negative two thir- .. thi- .. two
 thirds is.

Face-to-face conversations SBC text 9

Contractions (*don't, you'd, can't*) are more common in the SBC subset spoken American English conversations than in the LLC British counterpart, and more verbs are in the present tense (*am, think, do, have, want, guess*). Analytic negation is frequent and usually contracted (*n't*). As in most spoken genres, first and second person pronouns together by far outweigh third person pronouns, although in IRC and split-window ICQ chat, first and second person pronouns individually show superior distribution over third person, as seen in section 4.2. Features with negative weights on Dimension 1 boost the mean dimension score with their low frequencies. Prepositional phrases, for instance, are markedly infrequent in both the SBC and split-window ICQ texts; in the SBC texts they are on average 61.1 per thousand words, and in ICQ only 42.0 per thousand words. Nouns are equally rare in both genres of conversational writing, and slightly more frequent in SBC. The type/token ratio is lower in face-to-face conversations SBC, although more notably the mean word length is higher. All in all, the distribution of features renders a very high mean Dimension 1 score for face-to-face conversations SBC, although not as high as for the split-window ICQ chats.

Dimension 1 is one of the dimensions most clearly associated with a literate/oral dichotomy, and we find that both conversational writing corpora reside in the oral end. Interestingly, however, Internet relay chat scores lower than face-to-face and telephone conversations. Let us find out why by way of an example.

- (3) <chatty`1> hi cheeky1 hows things:
 <{{{odew[[[[[> hey allllll
 <Cheeky1> ok thanx and u chatty1?
 <{{{odew[[[[[> how u guy doing
 <chatty`1> hi mad max how u doing:))
 <Sexy-Xhick> hey
 <Sexy-Xhick> grrrrrr
 <Sexy-Xhick> grrggr
 <Sexy-Xhick> ?
 <Sexy-Xhick> ;
 <Cheeky1> grrrr u
 <chatty`1> gret thanks cheeky
 <Cheeky1> well eat u
 <Cheeky1> lol
 <|mad_max|> fine, chatty and u+
 <|mad_max|> ?
 <chatty`1> great thanx max
 <|mad_max|> glad to hear
 <Cheeky1> me 2
 <chatty`1> hey cheeky we finally got your weather:(
 <REVOLI> hi everyone how are you?
 <Cheeky1> lol
 <Cheeky1> lucky u
 <Cheeky1> its horrible here at the moment
 <Cheeky1> so wet and muggy
 <chatty`1> same here think the house is about to float away
 <Cheeky1> lol
 <Cheeky1> get those paddles out
 <Guest75862> hello
 <chatty`1> hehehe got the floaties just incase

Internet relay chat text 3b (UCOW)

As the flow of the IRC conversation is rapid, messages are kept very short and occasionally consist of only one keystroke. Example (3) illustrates how the competition for attention calls for minimal turns, and how these turns manifest themselves with abundant abbreviations and misspellings. Paradoxically, the irregularity of spelling renders a seemingly varied vocabulary, i.e. a great number of types, at least according to the discretion of available type/token ratio calculators (the mean TTR of IRC is 54.9), as seen in section 4.3. The type/token ratio of Internet relay chat is thus comparable to that of written texts such as press reportage, editorials and reviews (whose means range from 54.4 to 56.5) – a factor that inevitably slightly reduces the dimension score of both conversational writing genres.

Interlocutors in public IRC channels are concerned with finding conversation partners; greetings abound, and the conversations rarely evolve beyond superficiality. The IRC chatters appear less inclined than the split-window ICQ chatters to share personal information, and private verbs common in the ICQ chat (*think, know, feel*, etc.), where interlocutors are previous acquaintances, are more rare in the IRC corpus. Along with the private verbs go THAT deletions, which are markedly few in the IRC chats compared to the split-window ICQ chats (although found in the IRC example (3) as *think Ø the house is about to...*).

A surprising, counter-intuitive, finding in the IRC texts is the relatively low frequency of contractions (30.8 per thousand words), as noted in section 4.4. The manual tagging ensured that no apostrophe was needed for them to be found, yet they turn out to be fewer than in the face-to-face conversations, telephone conversations or split-window ICQ chats (which range from means of 46.2 to 55.0 contractions per thousand words). The relative rarity of analytic negation (including its contracted form *n't*) in IRC partly helps to explain the low frequency of contractions (as noted in section 4.4). Another likely explanation for the rarity of contractions, however, lies not with the contractions themselves, but rather with the irregular orthography, and lexico-grammatical homogeneity, of the IRC tokens, as well as with the prevalence of inserts and emotives, and nicknames used as address terms. A great number of tokens (besides pronouns *I* and *u*) are one mere keystroke long (such as *?, , ; 2*), while other tokens are the result of fingers resting on a keyboard key for the entire turn (*grrrrrr*). This irregularity yields an abundance of tokens, some of them nonsensical; that is, lexically empty. As standard scores are based on relative differences in frequencies per thousand words, several of the linguistic features with positive weights on Dimension 1 score low in the IRC chats (this also partly accounts for the relatively low scores of e.g. DO as pro-verb and demonstrative pronouns). Their common denominator is simply a vast number of irregular tokens, frequently consisting of one mere keystroke (but also repeated greetings, compliments, phatic expressions and attention-attracting tropes such as *hi green, hey, gret thanks cheeky, grrrr u*). The IRC tokens, moreover, represent a collection of rather few of Biber's (1988) linguistic features (nine Biber features, for instance, have no textual representation in IRC). In addition, roughly every fifth word in IRC is an insert, an emotive or a nickname address term, i.e. belongs to a category not included in the dimension score calculations. Markedly frequent categories included, however, are direct WH-questions (*how are you?*) and indefinite pronouns (*anyone*), general forms of address, which act together with prevalent first and second person pronouns to raise the Dimension 1 score of Internet relay chat.

At the literate end of Dimension 1 are the texts of traditional writing. Contrasting them with the chatted texts immediately elucidates two of the major distinctions between traditional writing and conversational writing – the salience versus sparsity of nouns, and the sparsity versus salience of first and second person pronouns. In example (4), from the official documents genre, more than every fourth word is a noun (roughly every fifth in the whole official documents genre, compared to roughly every ninth in the split-window ICQ texts).

- (4) As before, the record made during the enumeration lists all buildings, residential premises and temporary places of abode, and all households occupying them, as the basis of the enumeration is the household schedule. The number of structurally separate dwellings (that is, houses or flats or other quarters built or adapted for separate occupation and forming a private and structurally separate unit) was obtained as previously, together with the number of households with sole occupation or sharing such dwellings, and the number of living rooms occupied by each household.

Official documents LOB H: text 1

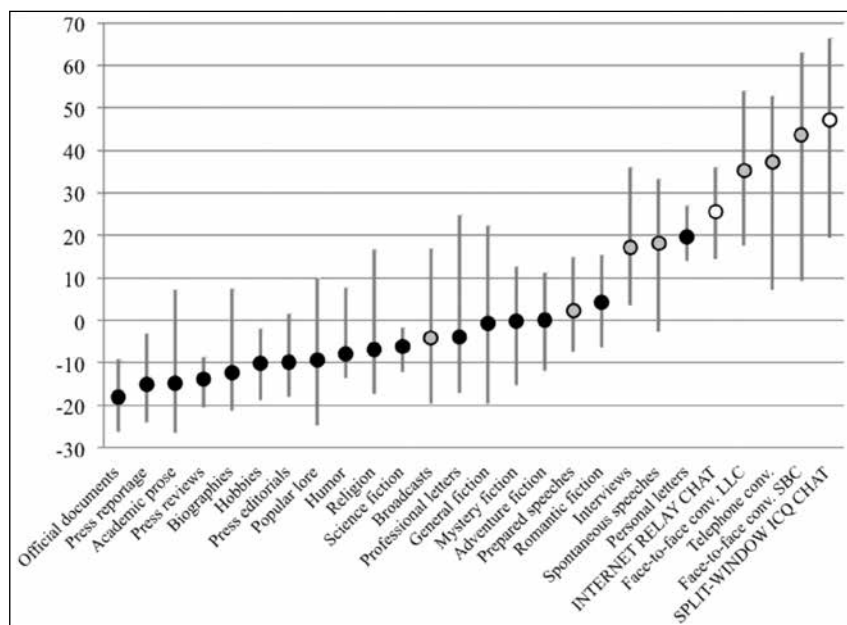
First and second person pronouns are extremely rare in official documents; only every hundredth and every thousandth word belongs to these categories respectively (Biber 1988: 254), compared to nearly every tenth and every twentieth word respectively in split-window ICQ chat. The common characteristic of texts on the literate end is their informational density, resulting from production circumstances that permit careful planning, redrafting and selective word choice. The texts have no affective content and consequently very few private verbs, THAT deletions and emphatics (e.g. *just, really, for sure*). Contractions and direct WH-questions are completely absent from the official documents genre of the LOB corpus,⁹⁶ whereas they are pervasive in the chatted UCOW texts. On the other hand, attributive adjectives are twice as common (*residential, temporary, separate, private, sole*), and prepositional phrases are more than three times as common in official documents as in chatted texts. Words in official documents are on average one character longer than in chats, but their TTR fails to compete with chats, for reasons addressed in section 4.3. In sum, the near-absence of features with positive weights on Dimension 1 combined with the impact of features with negative weights (nouns, attributive adjectives, prepositional phrases and word length) results in very low dimension scores for traditional writing, as illustrated in figure 5.1a. In short, judging from the mean dimension scores, traditional writing is informational whereas

⁹⁶ Actually, there is one (1) contraction in the 28,000-word official documents component studied.

conversational writing is involved. Now, let us briefly consider the genre-internal *spread* of scores along Dimension 1 and find out if this proposition holds.

Besides the mean dimension scores, table 5.1 (in section 5.1) indicates the genre-internal variation, i.e. minimum and maximum scores, of texts in the genres of IRC, split-window ICQ and face-to-face conversations SBC, in analogy with the descriptive dimension statistics for Biber's genres, found in Appendix VIII (taken from Biber 1988: 122–125). As mentioned in section 3.5, dimension scores were, in fact, computed not just for each new genre, but for each text in the genres. To illustrate the spread of these scores, figure 5.1b plots the mean and range of scores along Dimension 1 in each genre studied (along with those of Biber's genres, Biber 1988: 122–125). The texts of split-window ICQ chat, for example, range on Dimension 1 from a minimum dimension score of 19.3 to a maximum of 66.4, a range that is illustrated in figure 5.1b as whiskers around the mean score of 47.2. Our focal concern now is to contrast the whiskers of writing (black dots) and those of conversational writing (white dots).

Figure 5.1b: Spread of scores along Dimension 1 for all genres (capitalization denotes conversational writing). Dimension 1: "Informational versus Involved Production" (adaptation of Biber 1988: 172–177 and 122–125, supplemented with the new genres).



A quick look at figure 5.1b strengthens the proposition that traditional writing is informational whereas conversational writing is involved – even the least involved texts of IRC and split-window ICQ chat have dimension scores that exceed those of most of the written genres. The least involved IRC conversations, however, show considerable overlap with personal letters and some overlap with professional letters (but as these letters are not available more specific analysis is unfeasible). Three more written genres are close on the heels of IRC: romantic fiction, general fiction and religion. A closer look at the standard deviation of these three, however, suggests that their texts are fairly tightly distributed around the mean (s.d.<10; see Appendix VIII), which is also the case for IRC (s.d. 7.1; see table 5.1), meaning that only few of their texts overlap on the on the “involved” end of the scale. Compared to the spoken genres, IRC is truly intermediate.

Split-window ICQ chat surpasses all other genres not just in mean dimension score, but also regarding the extent of its scores into the involved end. One third of the ICQ conversations are more “oral” and involved than the LLC face-to-face conversations, although only one conversation in isolation surpasses the SBC subset texts. More striking at the other end is that even the least involved split-window ICQ chat conversation (a statistical outlier with a dimension score of 19.3) has very little in common with traditional writing, except for personal letters. The split-window ICQ chats are highly interactive, personal and affective, which apparently is a characteristic displayed in some of the personal letters and a few of the professional letters, but among the written genres from LOB, only a few general fiction texts display any resemblance – naturally deriving from dialogue such as example (5).

- (5) ‘But the Old Man doesn’t care for using double-barrelled names, as he calls them. And I think I agree with him. That’s why I use just the plain “Lee” on my cards. But if you think’ – and his expression changed quickly to deliberation – ‘that I should use the Stratford-Lee, just out here I mean, then of course-’ ‘Oh Lord, no,’ I said, perhaps just a little too abruptly. ‘There are far too many double-barrelled names out here as it is.’ He sat back again, obviously satisfied. ‘I’m inclined to agree with you, sir,’ he said.

General fiction LOB K: text 2

The split-window ICQ chat texts above the bottom outlier text range from dimension scores of 32.7 to 66.4, a range which clearly separates supersynchronous chat conversation from all written genres and renders only analogies with face-to-face and telephone conversation fruitful.

In conclusion then, on Dimension 1, conversational writing most closely resembles face-to-face and telephone conversation, and although the Internet relay chat features are occasionally difficult to interpret and show some parity with e.g.

personal letters, the results indicate firmly that the conversational writing genres are distinct from traditional writing and that supersynchronous conversational writing occasionally even appears to exceed the involvement of speech. However, no single dimension in itself will account for the full range of variation in language, a fact that will become clear as we move on to consider the second of Biber's dimensions.

5.2.2 Dimension 2: Narrative versus Non-Narrative Concerns

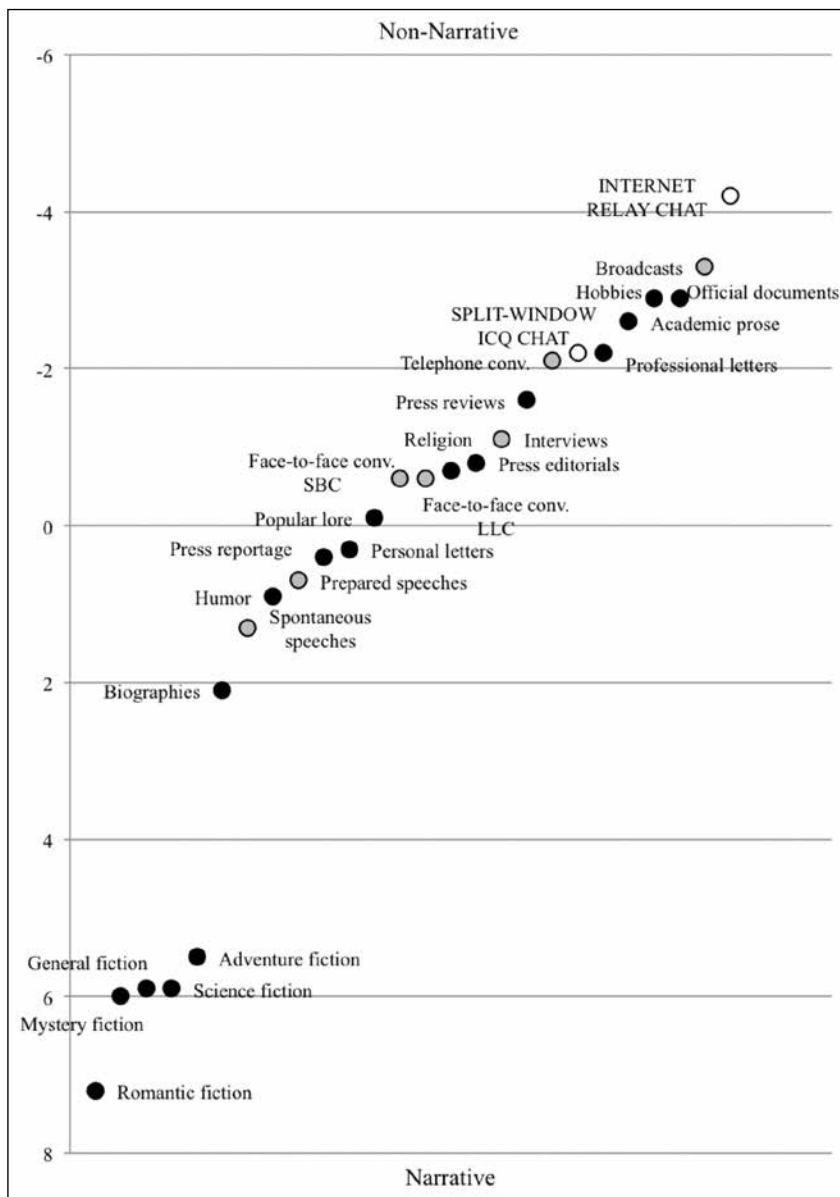
Several scholars who pioneered the analysis of patterns in writing and speech discussed the varying patterns in terms of speech styles (Ervin-Tripp 1972, Hymes 1974, Brown and Fraser 1979). Other scholars studied linguistic features across social groups and situations and came up with labels for basic discourse dichotomies such as high and low varieties (Ferguson 1959), nominal versus verbal styles (Wells 1960), elaborated versus restricted codes (Bernstein 1970), formal versus informal registers (Ervin-Tripp 1972, Irvine 1979) and planned versus unplanned discourse (Ochs 1979). Chafe (1982) was one of the first to empirically identify sets of co-occurring features that characterize written and spoken texts into underlying dichotomous dimensions, which he labeled "integration vs. fragmentation" and "detachment vs. involvement" (as seen in chapters 2 and 4), and Tannen (1982a, 1985) discussed linguistic variation in terms of oral versus literate discourse – the type of discourse distribution partly elucidated by Biber's Dimension 1 (Biber 1988). However, not many of the scholars were able to account for the range of variation among written and spoken texts as regards *narrative* concerns.⁹⁷ Biber's discovery of the second dimension of variation therefore threw light on a continuum that distinctly separates fiction genres from other written genres and distinguishes among genres of writing and speech, but only in intricate ways without association to divisions of literacy and orality.

Dimension 2 is labeled "Narrative versus Non-Narrative Concerns" (Biber 1988). Genres with high positive scores on Dimension 2 are all associated with past-time narration, whereas genres with high negative scores are similar to each other only in that they lack narrative concerns. As seen in figure 5.2a, the fiction genres cluster by themselves in the bottom left corner (most narrative) and an array of genres share the upper right end of the scale (non-narrative), with Internet relay chat standing out at the top.⁹⁸ Intermediate in the continuum is a variety of written and spoken genres.

97 Although see Tannen (1982b) for a vigorous exception.

98 The y-axes in Dimensions 2, 3 and 5 are reversed to facilitate comparison across dimensions and for the dimensions' interpretive names to read correctly from left to right across genres.

Figure 5.2a: Mean scores on Dimension 2 for all genres (capitalization denotes conversational writing). Dimension 2: "Narrative versus Non-Narrative Concerns" (adapted from Biber 1988: 136).



The linguistic features with a bearing on Dimension 2 were shown in table 5.4. The five fiction genres are all characterized by a high concentration of past tense and perfect aspect verbs, public verbs (e.g. *insist, mention, say*), third person pronouns, synthetic negation (*no, neither, nor*) and present participial clauses – typical markers of narrative action. In romantic fiction these features co-occur with very high frequencies, as illustrated in example (6).

- (6) He reached over into the back and lifted out his bag.
“But not yours, Mrs. Landry. I attend only to the lower members of your household.”
He said it quite without rancour, and I was positive none was intended.
“But you could be mine,” I insisted.
He inclined his head. “I could, yes. But I would advise you to see your own man, one who knows and understands you.” He shut the door and leaned down through the window to ask, “Are you coming in, Mrs. Landry?”

Romantic fiction LOB P: text 15

The majority of verbs in example (6) are in the past tense (*reached, lifted, etc.*). Public verbs are prevalent in connection with dialogue (*said, insisted*), and the reference to characters in third person naturally carries the story forward. Synthetic negation is more common in fiction than in other writing, although not a decisive factor in (6). Present participial clauses add description and narrative action to stories, e.g. *Seizing a piece of carpeting* Mr. Herman attempted to... (LOB P: text 1), as well as conclusive import to the score – these clauses are nearly absent in non-fictional writing, even more infrequent in speech and completely absent from the chats.

Narration is by definition concerned with the rendering of (human) events, which the narrator communicates directly to the reader/listener. In fictional texts, authors (or their characters) are the narrators, but, to the extent that narration occurs, any communication can take on a narrative flavor. In face-to-face conversation, for instance, speakers typically switch back and forth between the rendering of past events and the discussion of current matters. Consequently, as the positive end of Dimension 2 indicates high density of narrative markers, and the negative end the absence of the same, we can see why the means of face-to-face-conversations (from both LLC and SBC) assume an intermediate position in the continuum, coinciding as they do on the score of -0.6.

On Dimension 2, no linguistic features carry a negative load in the calculation of dimension scores; yet, the paucity of relevant positive features adds up to negative numbers for many of the genres (as explained in section 3.5). Split-window ICQ chat is about as non-narrative as telephone conversations and professional

letters, whereas Internet relay chat is the least narrative of all. Interpreted differently, split-window ICQ chat is more narrative than Internet relay chat, even though neither of them is particularly concerned with narration. The slightly more narrative concern of ICQ chat is partly explained by looking at the relationship between the interlocutors. The ICQ chatters in UCOW are previous acquaintances, most of them high school classmates with friends in common and occasional stories to tell, as in part of example (7).

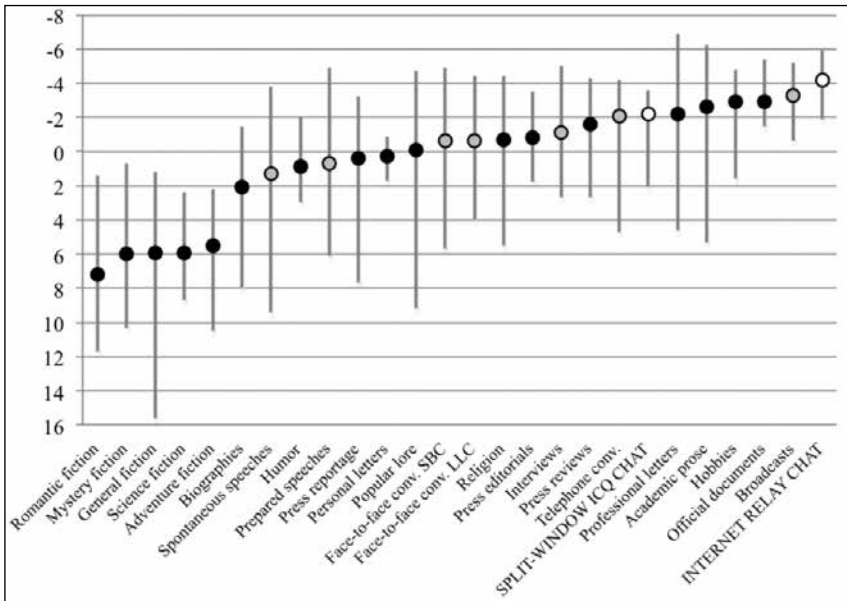
- (7) <10> did u go to the dance last weekend
 <J> uhh.. nope.. not reallly.. not much of a fan of dances
 <J> but i heard it sucked anyways
 <J> the only thing i hated about the dance when i didnt even go
 there is that i learned that michael took out my sis.. :-(
 <10> yeah
 <J>
 <J> nah.. i dont think so.. me and spencer is pretty cool.. so i
 really didnt care much that he went out wit my sister
 <10> does he like her
 <J> though they both say it wasnt a “date” because they went as “friends”
 <10> oh ok
 <10> oh well he might get with her
 <10> what would u do if he did
 <J> i would do absolutely nothing
 <J> my sister has her personal life.. i stay away from her personal life because
 she needs to live her life without me interffering.. ya know?

Split-window ICQ chat text 9 (UCOW)

IRC interlocutors, on the other hand, are not previously acquainted with each other, at least very rarely in real life, and therefore have few common referents. In public channels, their main concern is with finding conversation partners through superficialities (repeated greetings, compliments, phatic expressions and attention-attracting tropes), as seen in example (3) in this chapter. IRC communication in public channels therefore rarely evolves into the narrative state, where interlocutors share stories or relate to events in the past.

Judging from the spread of scores along Dimension 2, illustrated in figure 5.2b, IRC has the highest concentration of non-narrative texts, but two genres have a few texts that surpass IRC’s range into the non-narrative end: professional letters and academic prose. As the professional letters in Biber’s 1988 study are unavailable for scrutiny, we will briefly look at academic prose. Exemplifying the absence of features is tricky, but the non-narrativeness of academic prose might be found in texts such as example (8).

Figure 5.2b: Spread of scores along Dimension 2 for all genres (capitalization denotes conversational writing). Dimension 2: “Narrative versus Non-Narrative Concerns” (adaptation of Biber 1988: 172–177 and 122–125, supplemented with new genres).



- (8) Changes in voltage accompanying fluctuations of coolant temperature according to equation 6 vary only slightly with concentration and are proportional to the temperature change. Values at various oxygen concentrations of [FORMULA] together with apparent changes in oxygen level for temperature fluctuations of 14 100 C at 500 C are presented in Table 1.

Academic prose LOB J: text 1

Example (8) from academic prose and example (9) from IRC, below, despite their apparent lexical and functional differences, remarkably share the non-narrative space diametrically opposed to fiction. Neither (8) nor (9) has any past tense, perfect aspect or public verbs, third person pronouns,⁹⁹ synthetic negation or present participial clauses;¹⁰⁰ that is to say, they completely lack the typical

99 Pronoun *it* is not a feature on Dimension 2.

100 Biber’s algorithm detects only present participial clauses preceded by punctuation or a tone unit boundary (Biber 1988: 233).

markers of narration. Example (9) shows how IRC communication typically is concerned with the immediate present, in this case with greetings as participants are joining and leaving the channel.

- (9) <}}melons{{> \\\elcome Back ^^katty^
 <scorpio_> hello }}melons{{
 <^^katty^> ty mels...man this is slow
 <Guest_162> hello ladies
 <}}melons{{> yes ot is
 <}}melons{{> it
 <Rich23> later alls
 <^^katty^> bye rich
 <Rich23> [bye ^^katty^]
 <ROCK> hello ladies
 <Farkles> hey, rock
 <crisatie> hi rick
 <Chaser> hi christie
 <scorpio_> hm not much talking in here tonight
 <crisatie> hi chaser
 <Syl> Hi Y'all
 <scorpio_> just lots of coming and going

Internet relay chat text 1a (UCOW)

The spoken genre on the non-narrative end, broadcasts, also largely lacks markers of narration but is somewhat noteworthy for its unexpected location. The genre stems from the London-Lund corpus radio broadcasts recorded in the 1960–70s, which reported on events actually in progress (sports and events commentary, wildlife commentary and a physics demonstration). The description of sports, events and demonstrations occurred almost exclusively in the present tense and the present progressive, reflecting the ongoing action. Since then, however, radio and television broadcasts have developed into an array of modes: more affective commentary, interviews with players on past events, interaction among commentators with reflections on past events etc., which more likely would place modern broadcasts among the genres with intermediate status on Dimension 2.

To sum up Dimension 2, we can conclude that no functional analysis of variation across writing, speech and conversational writing is adequate unless it also takes into account the narrative dimension, the dimension that places fiction on one end and non-fiction, whether in the form of prose or conversational writing, on the other. We found that both genres of conversational writing largely lack markers of narration and, even though brief passages of narration are

found in the split-window ICQ chats, the genres generally display interaction with the immediate time as the main concern. In the sections below, Biber's further dimensions will each shed further light upon the genres of conversational writing, as we are ultimately aiming for an adequate, multifaceted description of conversational writing in relation to the textual variation of the English language.

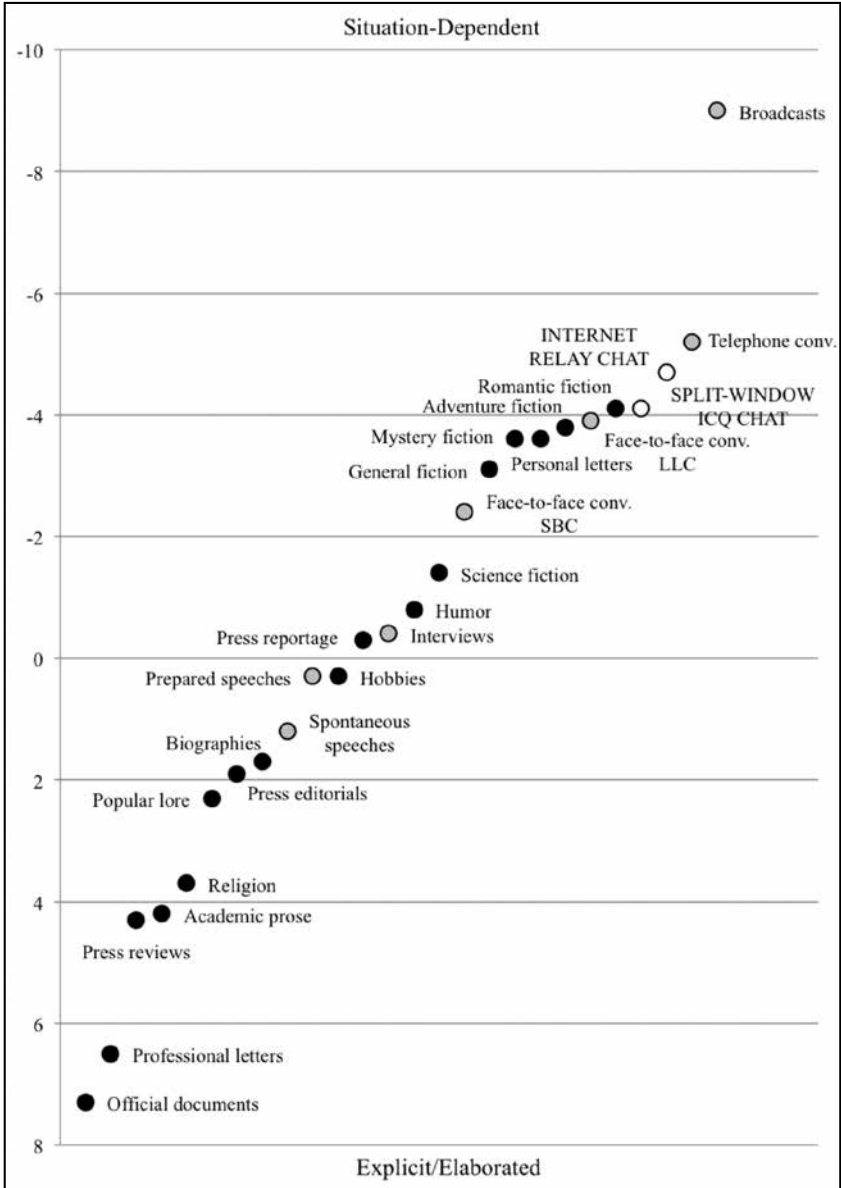
5.2.3 Dimension 3: Explicit/Elaborated versus Situation-Dependent Reference

When communication participants share time and space, as in face-to-face conversation, reference to external objects, events and people is frequently made by temporal deixis (*yesterday, tomorrow*), spatial deixis (*inside, outside*) and pronominal reference to objects or humans sharing the same space. In most such circumstances, no confusion arises as to which time, space, object or person is referred to, as the referent is either immediately discernible or can be inferred from shared knowledge or the shared physical context, i.e. by situation-dependent, exophoric, reference (Halliday & Hasan 1976, Biber 1988), as explained in section 4.5. When, by contrast, communication participants (sender and receiver) are temporally and/or spatially separated, reference to such objects usually must be made explicitly, in an elaborated, endophoric, way (Halliday & Hasan 1976). Biber's (1988) Dimension 3 distinguishes among texts with precisely these separate functions.

Explicit reference is by nature independent of physical context; the wording of a legal act (an official document), for instance, must be equally valid whether read in court, where proceedings are taking place, or in the street of a different city. Furthermore, it must be independent of temporal context, as its wording will stay valid for a long period of time, across generations of readers. To attain this general validity, highly specific reference is needed. Elaborate constructions involving WH relative clauses, pied-piping relative clauses, phrasal coordination and nominalizations are stacked up so that no doubt arises among readers as to intended referents (Biber 1988: 110). The resulting texts are highly integrated and informational and, except as markers of text-internal deixis, time and place adverbials are rare. The linguistic features mentioned are the very features that in co-occurring distribution carry positive weight on Biber's Dimension 3; see table 5.4.

On Dimension 3 (figure 5.3a), the features with positive weight are contrasted with the absence of features with negative weight and vice versa.

Figure 5.3a: Mean scores on Dimension 3 for all genres (capitalization denotes conversational writing). Dimension 3: “Explicit/Elaborated versus Situation-Dependent Reference” (adapted from Biber 1988: 143).



The genres with positive dimension scores are texts with explicit and elaborated reference, found in the bottom left-hand corner of figure 5.3a. Genres with negative dimension scores are texts with situation-dependent reference (frequent time and place adverbials and adverbs), located towards the upper right. As before, a chain of genres share intermediate positions on the continuum, and written and spoken genres intertwine. A faint pattern emerges, however, according to which traditional writing mainly resides towards the bottom left and speech and conversational writing towards the upper right. A few contrastive examples will serve to shed light on the distribution of texts and the status of conversational writing on Dimension 3.

Example (10), part of a legal amendment from official documents, illustrates vigorous use of constructions used for explicit reference: a WH-relative clause (*which provides for*), a pied-piping construction (*for which*), and several nominalizations¹⁰¹ (*subsection, section, detention, indictment, imprisonment*) – all in less than 100 words. Together with the near-absence of features with negative weight (no time and space adverbials, only one adverb), the text carries a dry, formal tone with large amounts of information packed into each phrase and clause.

- (10) In subsection 2 of section fifty-three of the Children and Young Persons Act, 1933 (which provides for the passing of a sentence of detention for a specified period in the case of children or young persons convicted on indictment of certain grave crimes therein mentioned) for the words from “an attempt to murder” to “grievous bodily harm” there shall be substituted the words “any offence punishable in the case of an adult with imprisonment for fourteen years or more, not being an offence the sentence for which is fixed by law”

Official documents LOB H: text 14

Conversely, in a sports radio broadcast, like example (11), reference is typically situation-dependent. That is to say, referents are not identified explicitly or elaborately, but must be inferred from the context of the message, and listeners are forced to invent a mental image of the setting and situations. The speaker assumes the listener’s keen familiarity with the physical context, and the communication is thus perfectly functional despite the spatial divide between broadcaster and listener.

- (11) Gowling is beaten #
Dearden comes in on it #

101 Biber’s algorithm for the automatic detection of nominalizations includes all words ending in *-tion, -ment, -ness* or *-ity*, whether with or without verbal origin or English stems.

temporally simultaneous, ranging from synchronous to supersynchronous communication (the latter with completely overlapping turns). The co-temporality factor seems to be at play here, but not conclusively as certain fiction genres and personal letters make almost as frequent use of situation-dependent reference. To sort this out, we will take a look at example (13), which sheds light on the experienced shared *space* of the split-window ICQ interlocutors.

- (13) <1> I dont like you anymore
 <A>
 <1>
 <A> what u could never not like me
 <1> Oh yea?
 <1> would you stop with the font... you're freaking me out!
 <A> juice JUICE
 <1> blah blah blah
 <1> you piece of crap!!
 <A> SORRY
 <1> uhh... okay whatever... freak
 <A> DONT EVEN START WITH ME!!!!!!
 <A>
 <A> U LOVE ME
 <A> SNAP
 <A> UR NASTY
 <1> wow... shot down... that hurt
 <1> i hate you :(

Split-window ICQ chat text 1 (UCOW)

ICQ chatters may be spatially distant in real life but their conversation takes place in a visually proximate interface on their computer screens. As the interlocutors inhabit this virtual space, the text of their interaction carries the social situation as well as their relationship to the situation and the objects under discussion (cf. Yates 1996). In chapter 2, this shared space was defined as the semiotic field (Halliday 1978), which, besides carrying the text, allows interlocutors to express themselves graphically beyond actual words (Yates 1993). Example (13) displays ample devices whereby the semiotic field is disarranged (in the original color script); font choice, color and size, upper case, repeated exclamation marks and an emotive. Among the grammatical features with a bearing on Dimension 3, we find a complete absence of positive features in example (13) but several adverbs (*anymore, never, out, even, down*).

The adverbial distribution in example (13) largely resembles that of face-to-face conversations, exemplified in (14) by a face-to-face conversation from SBC with a similar low score on Dimension 3. In example (14), situation-dependence

is evident in the speakers' exophoric reference to shared time (stay up *late*, do this *anymore*, what can I do *now*) and their abundant use of adverbs (*up*, *usually*, *like*). In combination with the absence of features for explicit reference (no WH-relative clauses, no phrasal coordination and no nominalizations), such features result in low dimension scores on Dimension 3 for conversational writing and face-to face-conversations alike.

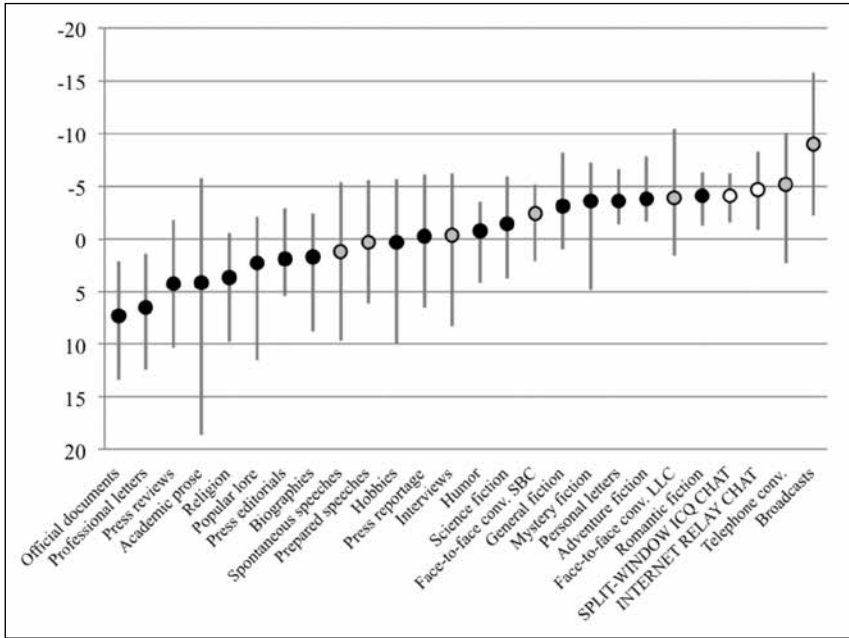
- (14) Mary: ... God,
 I said I wasn't gonna do this anymore.
 ... Stay up late.
 ... Kinda defeats the purpose of getting up in the morning.
- Alice: ... I know.
 .. And it's a hard habit to break.
 Usually I don't
- Mary: It is.
 s- Usually I don't stay up late.
 ... But it's like,
 if I'm up after midnight
 .. It's just like,
- Alice: ... Hm.
 ... Yeah yeah.
- Mary: What can I do now.

Face-to-face conversations SBC text 7

We can conclude then, that situation-dependent reference arises in contexts where both co-temporality and co-spatiality are at play, whether separately or in synergy. A few fiction genres manage to evoke a sense of shared time and space between characters and readers, through authors' vivid descriptions involving frequent use of adverbs. More explicably, however, situation-dependent reference occurs in the spoken genres and in conversational writing, where time, and a real or virtual space, is actually shared.

Noteworthy with regard to the spread of scores along Dimension 3, in figure 5.3b, are the whiskers of academic prose (Biber 1988: 174). Most texts in the genre are informationally dense, highly explicit and elaborated, with relative clauses galore, but to the extent that text-internal deixis is needed, time and space adverbials do occur, which for these texts yield a remarkable spread of scores. As the reader has perhaps noted in the present volume, academic prose indeed ranges from being extremely explicit to being fairly situation-dependent in character. On the next dimension to be considered, academic prose displays another record spread of scores, but more importantly, the two genres of conversational writing diverge slightly and assume positions on opposite sides of the zero point.

Figure 5.3b: Spread of scores along Dimension 3 for all genres (capitalization denotes conversational writing). Dimension 3: “Explicit/Elaborated versus Situation-Dependent Reference” (adaptation of Biber 1988: 172–177 and 122–125, supplemented with new genres).



5.2.4 Dimension 4: Overt Expression of Persuasion/Argumentation

On Dimension 4, figure 5.4a, which Biber’s early work calls “Overt Expression of Persuasion” (Biber 1988: 111), it seems only fair that academic prose scores below the zero point – scholarly publications, per traditional definition, should stay neutral and non-opinionated, as “author-evacuated” prose (Geertz 1988: 9) by tradition is “the standard of credibility in academia” (Surman Paley 2001: 31, also discussed by Elbow 1991 and Johns 1997). Dimension 4 has only features with positive weights, and when these add up to high frequencies in a text, the text is considered marked with persuasive and argumentative force; that is to say, it contains a speaker’s or writer’s expression of “likelihood or advisability” (Biber 1988: 148). Conversely, when the features are markedly infrequent, the text has no overt expression of persuasion or argumentation. All the same, the first sentence in this paragraph elucidates how academic prose occasionally may contain a *should* (as in *should stay neutral and non-opinionated*), a necessity modal which

adds argumentative but not necessarily persuasive force to an utterance. In later work, Biber seems to address remarks of this kind, and similar confusion arisen with regard to the dimension, by renaming the dimension “Overt Expression of Argumentation” (e.g. Biber 1995: 159). More recently, however, the two stances are combined with a solidus: “Overt Expression of Persuasion/Argumentation” (Conrad & Biber 2001b: 35), which is the label adopted here.

In figure 5.4a, the genres with overt expression of persuasion/argumentation are found in the top right end, those that lack this character drop down into the left bottom corner, and a chain of moderate genres range in between. The features carrying weight on the dimension were listed in table 5.4: to-infinitives, prediction and necessity modals (e.g. *will, would, shall, ought, should, must*), suasive verbs (e.g. *agree, ask, beg, recommend*), conditional adverbial subordinators (*if, unless*) and split auxiliaries (*they are objectively shown to...*).¹⁰²

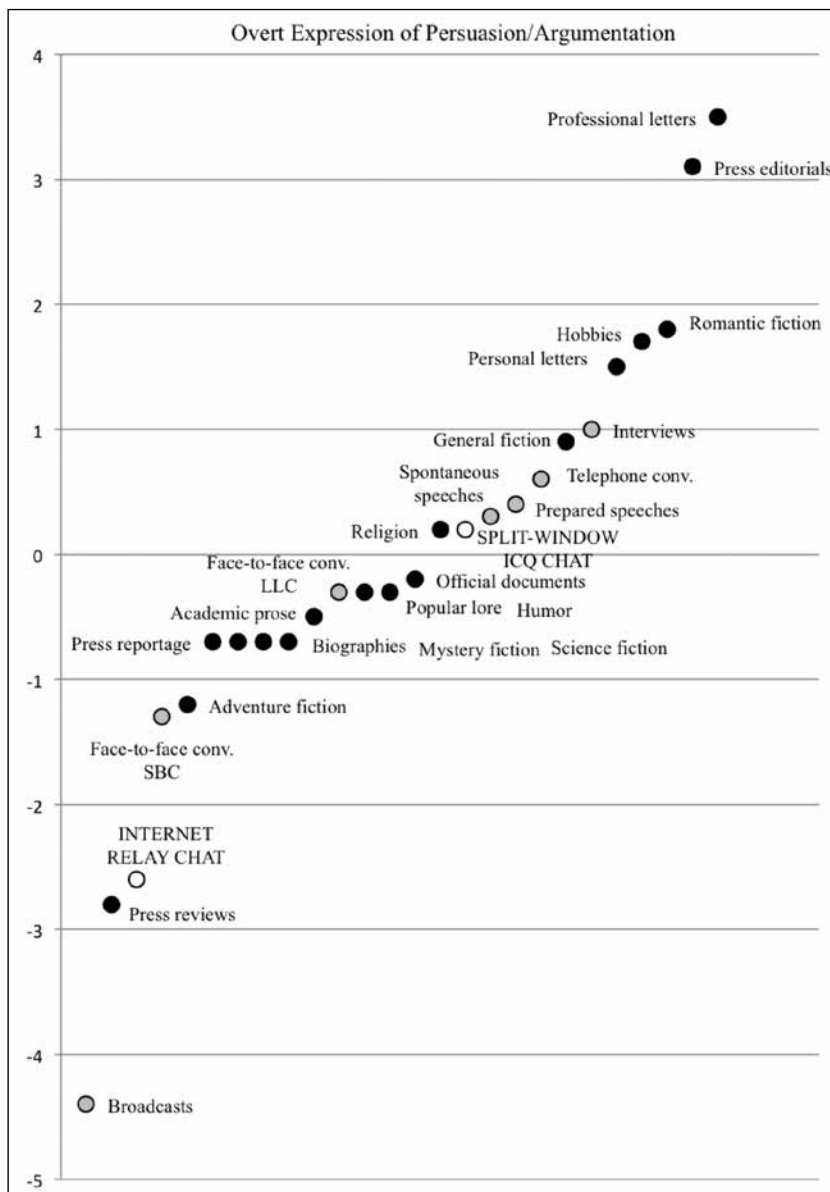
The main concern here will be to shed light on the slightly divergent positions held by the genres of conversational writing in the continuum. Judging from figure 5.4a, neither IRC nor split-window ICQ chat is overtly persuasive/argumentative, but the mean of split-window ICQ chat is higher than that of IRC (although not significantly so; see table 5.3). As it turns out, several split-window ICQ chat texts contain animated discussion, which, as we shall see, is instrumental in bringing the mean for split-window ICQ chat above the zero point (recall that the zero point, in all dimension plots, constitutes the mean of all of Biber’s written and spoken genres, listed in Appendix I).

The low R^2 values of Biber’s genres on Dimension 4 (16.9%) and of the added genres (11.0%; see table 5.2) indicate that the importance of the dimension is relatively small in distinguishing among the genres, which tells us we should not read too much into the scores.¹⁰³ Neither of the conversational writing genres is statistically different from face-to-face conversations SBC (as indicated by the p-values in table 5.3), but some IRC and split-window ICQ chats differ from each other more than others.

102 “Split auxiliary” means an auxiliary split from the main verb by adverb(s). Biber’s (1988: 244) algorithm for the detection of split auxiliaries was interpreted *aux+adv+(adv)+v*, the last element being *v* (any verb), instead of his suggested *vb* (base form of verb).

103 The R^2 value “indicates the percentage of variation in the dimension scores of texts that can be accounted for by knowing the genre category of the texts” (Biber 1988: 126). In our case less than 16.9% of the variation in dimension scores along Dimension 4 can be accounted for by knowing the genre categories.

Figure 5.4a: Mean scores on Dimension 4 for all genres (capitalization denotes conversational writing). Dimension 4: “Overt Expression of Persuasion/ Argumentation” (adapted from Biber 1988: 149).



immediate present; they are aware of each other's attitudes and occasionally dare to challenge them. Their discourse is often opinionated, even to the point of being adversarial, albeit usually exchanged with an ironical glint, as observed in split-window ICQ examples (1) and (13) earlier in this chapter. The adversarial discussions observed in examples (1) and (13), however, come from texts with fairly low dimension scores on Dimension 4 (-0.3 and -0.9, respectively). This means that very few of Biber's features of overt expression of persuasion/argumentation are found in them; rather, in those texts, argumentation manifests itself in other, more refined, ways. To find out what kinds of expressions bring about positive scores on Dimension 4, we will instead turn to example (16). The example comes from a text that ranks among split-window ICQ's highest on the persuasive/argumentative pole (2.7, cf. figure 5.4b). It is a non-adversarial, rather supportive text – a motivated discussion on plans for the weekend.¹⁰⁴

- (16) <5> So i dont know if we should still save \$\$\$ – Which i dont have – or if i need to rush out and get her something with \$\$ (that idont have)
- <5>
- <5> Thanks
- <5>
- <5> i know
- <5>
- <5> but i told her we should just celebrate v day some other time ..?
- <5>
- <5> kinda – we didnt really know if tim and val could go til today/last nite cause they had to get out of p[lay practice.
- <5>
- <E> yea I see what you are saying. your in a bad bind man did you how did that go over or was she cool with that cause if she was then I would be like hell yea o i c.
- <E>
- <5> yeah i think we'll do that – but that may be like celebrating – pause thought- sweet to what? oh – yeah and put the me/her stuff on hold?
- <5>

104 Example (16) contains several extensively overlapping turns, which the logging software fails to record in logical sequence. The video clip of the example, however, shows the turns of <5> and <E> dialogically juxtaposed.

<E> I would do that hell man you and your gf and a couple of friends in the [mountains] that would be sweet. LOL. going up to the [mountains] for the weekend over vday.

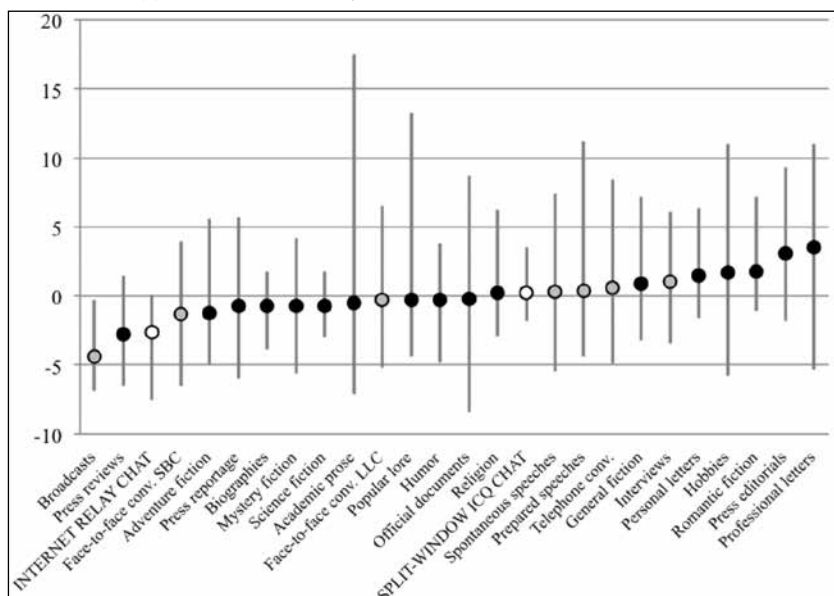
Split-window ICQ chat text 4 (UCOW)

Common for the ICQ chat texts with high scores on Dimension 4 is that they, like example (16), contain an element of advice-giving or a discussion of alternative options. The discourse resembles overt expression of persuasion/argumentation in that propositions are modulated by necessity and prediction modals, even though the propositions here are more of an encouraging nature, than of an overtly persuasive or argumentative kind. In example (16), expression of such moderate persuasion/argumentation is carried out by four prediction modals (*would*, *ll*), two necessity modals¹⁰⁵ (*shoulf*, *should*), four conditional adverbial subordinators (*if*) and a few split auxiliaries. A distribution of this density is not typical of the ICQ texts, but it illustrates the ways in which several ICQ texts deviate from most IRC texts: the ICQ chatters discuss matters from their shared real-life context, which occasionally brings about supportive or challenging argumentation, whereas the IRC chatters rarely exchange views in animated ways.

On this fourth dimension of variation, the spread of scores is extensive in most genres; see figure 5.4b. The texts of the conversational writing genres are only moderately spread. Nevertheless, the diverging ranges of the two conversational writing genres offer a measure of support to the conclusions drawn in the discussion of examples (15) and (16); the IRC texts indeed trail down into the non-argumentative domain, whereas the split-window ICQ chats, to some degree, extend into the persuasive/argumentative domain.

105 Only core modals are counted, not semi-modals; see section 4.2.

Figure 5.4b: Spread of scores along Dimension 4 for all genres (capitalization denotes conversational writing). Dimension 4: “Overt Expression of Persuasion/Argumentation” (adaptation of Biber 1988: 172–177 and 122–125, supplemented with new genres).



Two genres outscore Internet relay chat in being non-argumentative; press reviews and broadcasts. In press reviews, opinions are expressed as if they were *the* correct view, and therefore the genre largely lacks markers of *overt* argumentation (Biber 1995: 162, also unexpectedly noted for direct mail letters by Connor & Upton 2003). Live broadcasts, on the other hand, contain inherently non-opinionated, non-persuasive discourse, near-void of the linguistic features on Dimension 4 (Biber 1988). Example (17) from broadcasts serves to illustrate the non-argumentative low extreme: a broadcast from the launching of a submarine, simply reporting a current and immediate progress of events.

- (17) Her Majesty's speaking to him now #
the end of [dhi:] line #
of presentations #
on the Admiralty side #
(- - music)
now Her Majesty #

prepares #
to come onto [dhi:] the dais #

Broadcasts LLC 10: text 7a

Along the other end of the scale, several genres surpass split-window ICQ chat in overt expression of persuasion/argumentation, press editorials and professional letters being the most marked types of argumentative discourse (Biber 1988). As the professional letters are unavailable for sampling, below is a brief passage from press editorials to illustrate the dense argumentative force that is largely lacking in the conversational writing corpora. The example is part of a political appeal containing one infinitive¹⁰⁶ (*to drive*), three necessity modals (*should*) and two split auxiliary constructions (*should not despair*, *should not encourage*), all in less than 50 words.

- (18) He should not despair of keeping a large part of his copper revenue. O'Brien has praised the valour of Katanga soldiers. Tshombe should not encourage them to drive the point home. Instead of putting up a desperate resistance he should spend an hour reading the Nigerian Constitution.

Press editorials LOB B: text 2

Observing the spread of scores along Dimension 4 in figure 5.4b, we might again, in passing, remark on the extension of academic prose, in this case into the persuasive/argumentative top (cf. Biber 1988: 175). Despite the genre's moderate mean and non-persuasive intention, a host of academic texts exceed most other genres in persuasive/argumentative force. On the next dimension, however, academic prose recovers its prototypically formal and untainted status by assuming, together with official documents, a position distinct from all other genres, a position diametrically opposed to conversational writing.

5.2.5 Dimension 5: Abstract/Impersonal versus Non-Abstract/Non-Impersonal Information

The labels of Biber's dimensions of variation have altered slightly over the years as contrastive genres and perspectives have brought new light to them. Dimension 5, for instance, was labeled "Abstract versus Non-Abstract Information" in Biber (1988), "Abstract versus Non-abstract Style" in Biber (1995) and Conrad & Biber (2001b), and "Impersonal versus non-impersonal style" in Biber et al. (1998), although in essence the variant labels point to one

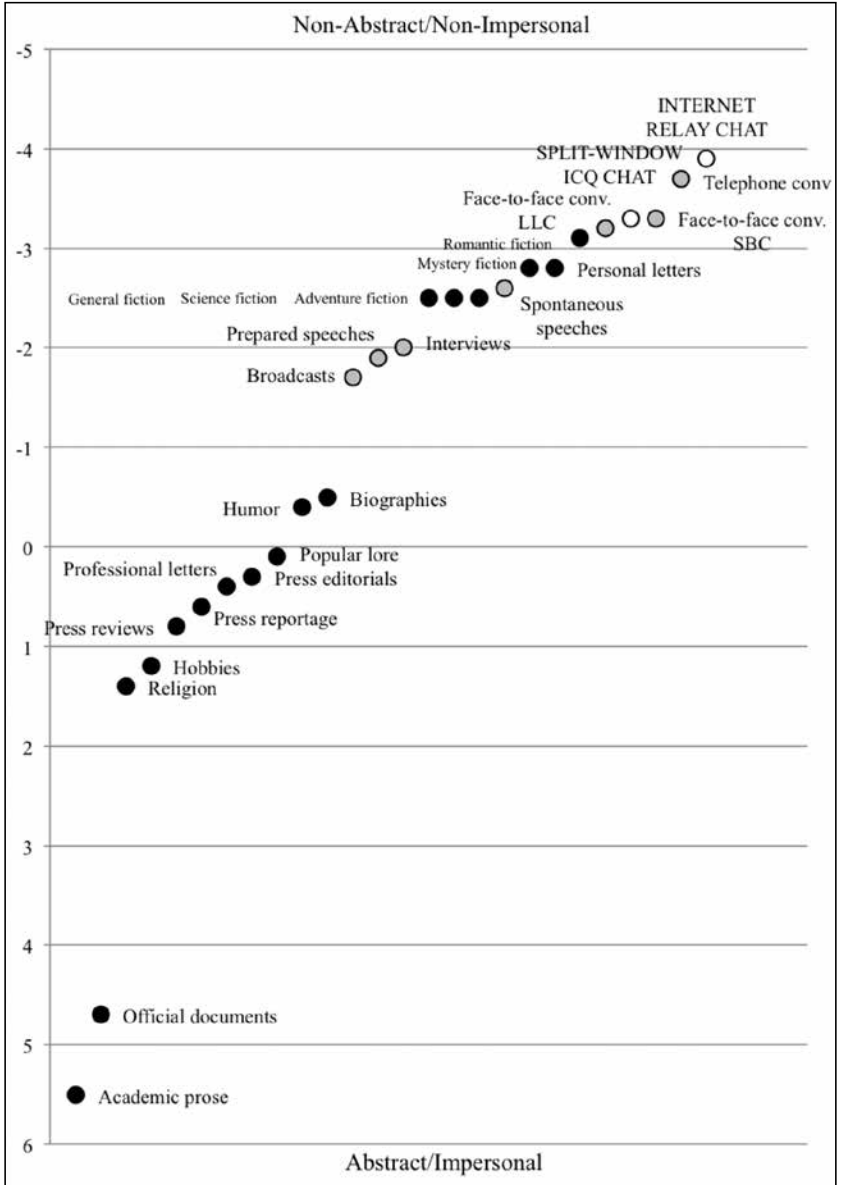
106 Only to-infinitives are included in the calculation of the Dimension 4 scores.

and the same dimension of variation: that between academic prose on the abstract end and telephone conversations on the non-abstract end. New in the Dimension 5 plot here, figure 5.5a, are the genres of conversational writing and face-to-face-conversations from SBC, which essentially prove all these variant dimension labels adequate, but shed further light on the non-abstract, non-impersonal, i.e. the rather concrete and personal, end of the scale and reevaluate informality.

Dimension 5 has only features with positive weights, although as before, the absence of the features yields negative scores. Figure 5.5a plots the mean dimension scores along Dimension 5. The genres with high positive dimension scores, academic prose and official documents, are found in the bottom left corner, and a number of other written genres cluster in intermediate position in the continuum. Conversational writing and face-to-face conversations SBC reside in the upper right corner (with no significant internal difference; see tables 5.2 and 5.3), closely interspersed with the other conversational genres and followed by personal letters, spontaneous speeches and the fiction genres.

The linguistic features with a bearing on Dimension 5 were given in table 5.4 above (from Biber 1988: 102–103). A high concentration of the features brings about the formal and complex technical style typically found in academic prose: conjuncts (e.g. *alternatively, conversely, moreover*), passives, past participial clauses (e.g. *based on the current rate, the value...*), WHIZ deletions (e.g. *the value Ø based on the current rate*) and certain adverbial subordinators (e.g. *since, while, whereby*). These co-occurring forms are devices whereby authors present information with reduced emphasis on the agent, forms that either demote the agent to a *by*-phrase or elide the agent altogether, and instead typically give prominence to a non-animate referent or an abstract concept. In academic prose and official documents, conjuncts and adverbial subordinators mark logical relations among clauses and serve to make complex reasoning explicit. Example (19) illustrates the frequent use of passives and complex clause constructions for dense integration of information in academic prose. No human agent is detectable in the discourse.

Figure 5.5a: Mean scores on Dimension 5 for all genres (capitalization denotes conversational writing). Dimension 5: “Abstract/Impersonal versus Non-Abstract/Non-Impersonal Information” (adapted from Biber 1988: 152).



- (19) The relations estimated were between the rates of change of the 'wage drift', the level of 'excess profit', the level of 'excess demand' and the rate of change in productivity. It may be pointed out that in our model productivity makes a significant contribution to the explanation of the spread between earnings and wage rates, when all variables are expressed as levels, but ceases to be a significant factor in our least squares computation in which variables are subjected to a first-difference transformation.

Academic prose LOB J: text 44

Dimension 5 is the third dimension, after Dimensions 1 and 3, to reflect a literate/oral dichotomy, i.e. a pattern emerges here, in which genres of traditional writing mainly plot in the lower, abstract end of the figure and speech in the upper, non-abstract end. The pattern closely resembles Dimensions 1 and 3 in that, apart from speech, only fiction genres and letters score in the vicinity of conversational writing. Judging from Dimensions 1, 3 and 5, therefore, we can determine that conversational writing adheres to the oral domain of the polarity, where the mean of one of its genres even takes the over-all lead (on Dimension 1, split-window ICQ, and on Dimensions 3 and 5, IRC). On Dimension 5, the absence of linguistic features with highly integrative textual functions entails the presence of features with concrete referents and active agents (although the latter features are not tagged specifically in Biber's 1988 study, and therefore not counted). Face-to-face conversations and conversational writing typically deal with matters of immediate, current relevance to participants. Agents are animate or tangible objects, and the topics discussed are concrete, as in example (20) from SBC of the interaction among the same interlocutors as in example (8), chapter 4, who are cooking a meal together.

- (20) Pete: They aren't particularly stringy.
Marilyn: Oh.
Then just snap em.
Roy: That probably looks like a three-person salad bowl,
Pete: I'll just and put em,
and put them...
Roy: hunh?
Marilyn: Man that's a big hunk of fish.
Pete: Where do you want em put.
Marilyn: Shit,
it's a huge...
Pete: Are they just going on that,
or,

Marilyn: Uh,
you wanna put em in a colander,
and then wash em?
There's a colander.

Face-to-face conversations SBC text 3

Example (20) contains no features with a bearing on Dimension 5;¹⁰⁷ neither does example (21) from IRC, below, with interlocutors discussing an image file transfer, a concrete albeit virtual object.

(21) <River> woohoo,
<Genie500> Laughing Out Loud
<River> my hair is almost as long as yours
<Genie500> now ya know who to look for honking across the street
<River> yep
<Genie500> really?? lol
<River> well just in the back
<Genie500> Laughing Out Loud
<Genie500> and what color is yours??
<lookingforagirl> blue
<Genie500> oh river just a sec I gotta turn something off for
you to send okay
<River> this one is from 95 without the glasses .
<River> ok
...
<Genie500> okay try again
...
<River> but the hair is almost the same now as then
<River> plus a wee bit more grey in it
<Genie500> Laughing Out Loud ok

Internet relay chat text 4a (UCOW)

Examples (20) and (21) are texts from informal settings with highly personal content and interactive, loosely integrated discourse, i.e. typical texts with extremely high negative scores on Dimension 5. Similarly, the chatters in the split-window ICQ chats discuss non-abstract and non-impersonal, i.e. personal, matters. Example (22) is a passage of a very personal kind, which contains none of the features that carry weight on Dimension 5.

107 The verb *put* in the construction *want them put* does not qualify as an agentless passive since Biber's (1988: 228) algorithm for agentless passives only detects those preceded by a form of the verb BE+noun/pronoun/adverb.

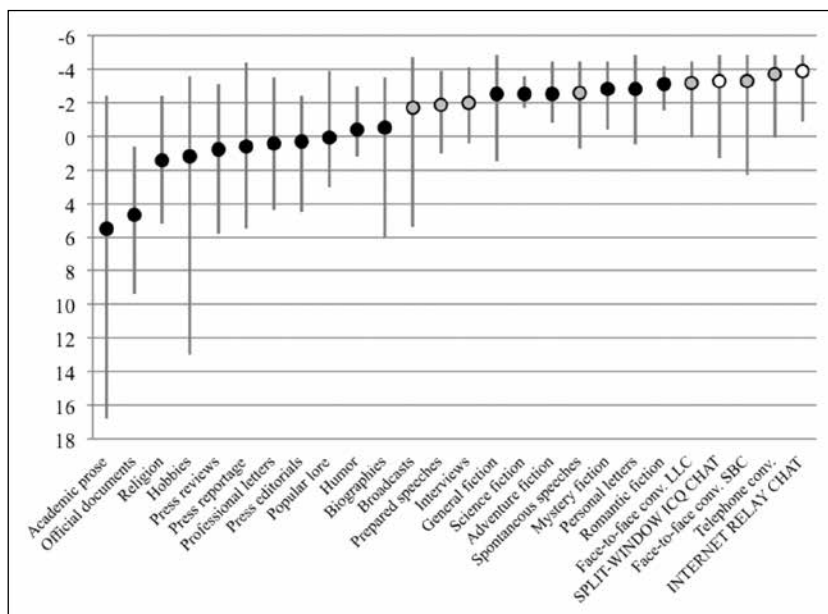
- (22) <6> i know...he's like one day "omg i like you sooo much etc.." and the next he's all pissy
 <6> he has like PMS
 <6> yeah...i do...but sometimes i can't take when he's in a bad mood..
 <6> like idk i'm one of those scarcastic girls...i can't help it i make fun of him when i get the chance...and he gets so mad at me for that
 <F> owch...do you like him?
 <6> but he knows i'm just messin
 <6> idk...
 <F> some guys take that stuff personally
 <6> correct
 <F> yeah i mean when he takes stuff so seriously, and he's in the bad mood all the time..not to mention he's a distance away, correct?
 <6> i know...like i see him...even when i don't plan to cuz he's best friends w/ my cuz...so the distance isn't even an issue..cuz me and my cuz r together like all teh time

Split-window ICQ chat text 5 (UCOW)

In figure 5.5b, the texts exemplified in (20) and (21) rank among the “top” texts at the non-abstract/non-impersonal end, which for the genres under discussion could be renamed the concrete/personal end. Judging from the spread of scores, however, texts from several genres touch the same non-abstract/non-impersonal end; 14 out of the 26 genres have texts with a minimum dimension score below (here “above”) -4 (i.e. a sum that is four standard deviations below the mean of all of Biber’s genres). Most notably, no genre has a text with a dimension score “above” the “top” IRC texts, but five genres reach the same “top” score of -4.8 (viz. general fiction, personal letters, split-window ICQ chat, face-to-face conversations SBC and telephone conversations). No other dimension shows an equally uniform distinct end to dimension scores, which naturally raises the intricate question whether there is a distinct far end to being non-abstract/non-impersonal (i.e. concrete and personal) – an initially mind-boggling question that, however, has a simple answer. Whenever a text displays no single occurrence of any of the features with a bearing on the dimension, the text will receive the “top” score. Two IRC texts, two split-window ICQ chat texts and one SBC text display such a distribution, along with an unknown number of texts from general fiction, personal letters and telephone conversations.

With regard to Dimension 5, we can conclude that conversational writing most closely resembles the genres of face-to-face and telephone conversation, but also that no text of conversational writing surpasses the most informal texts of face-to-face or telephone conversations – all four genres have texts that reach the same distinct non-abstract, non-impersonal end.

Figure 5.5b: Spread of scores along Dimension 5 for all genres (capitalization denotes conversational writing). Dimension 5: “Abstract/Impersonal versus Non-Abstract/Non-Impersonal Information” (adaptation of Biber 1988: 172–177 and 122–125, supplemented with new genres).



5.2.6 Dimension 6: On-Line Informational Elaboration

Dimensions 6 and 7 have few linguistic features with important loadings and are thus difficult to interpret. Accordingly, most studies applying Biber’s 1988 methodology have not considered these dimensions; in fact, Biber’s (1988) study discards Dimension 7 on theoretical grounds on an a priori basis, finding its factorial structure too weak for further exploration. Dimension 6, nevertheless, will be considered here, for the sake of a complete analysis, as even tentative results might be worthwhile to study in the exploration of conversational writing.

Dimension 6 has only features with positive weights; see table 5.4: three types of dependent clause (THAT complement clauses on verbs and adjectives and THAT relative clauses on object position) and demonstratives (*that*, *this*, *these*, *those* preceding nominals, not to be confused with demonstrative pronouns, which load on Dimension 1). Texts with high scores of the features are informationally elaborate, yet display relatively unplanned discourse, i.e. a type of discourse produced under real-time constraints, and the dimension is therefore

labeled “On-Line Informational Elaboration.” By contrast, texts lacking the same features are regarded as containing no on-line informational elaboration. The positions on Dimension 6 of the genres studied are shown in figure 5.6a.

Genres with high dimension scores are the less involved (cf. Dimension 1), i.e. less interactive, spoken genres spontaneous speeches, interviews and prepared speeches. They recurrently have an informational focus and often convey the speaker’s attitude or beliefs, but are produced in strict real time.¹⁰⁸ Example (23) is part of a spontaneous oration given in the House of Commons, exemplifying real-time informational elaboration.

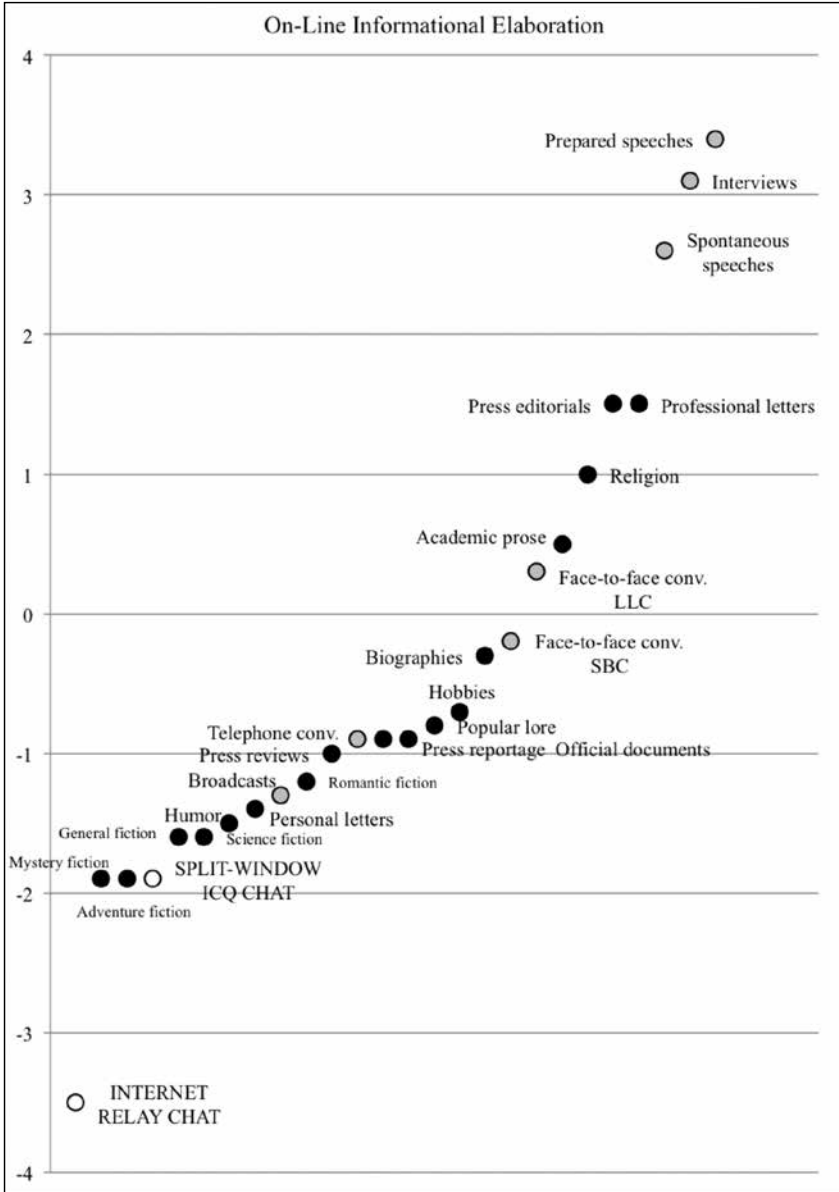
- (23) I do not think that it would be helpful # [pause]
to to engage #
in sort of in name calling #
against the opponents # [pause]
of the Concorde project #
and certainly #
neither I #
nor my right honourable friend #
intend to follow the honourable gentleman in that regard #
Mr Speaker #
will my honourable friend accept #
that many people in this House #
think that Concorde is going to be a gigantic financial disaster # [pause]
will he ensure that in any cuts in public expenditure #
education and social services take priority over this huge pit into
which money is being poured #

Spontaneous speeches LLC 11: text 4

The pauses in example (23), here retained from LLC’s original prosodic transcription, reflect the speaker’s planning time required to further elaborate the subject matter. THAT complement clauses are used ad hoc on verbs to add pieces of information (*I do not think that...*, *will my honourable friend accept that...*, *many people in this House think that...*, *will he ensure that...*). In this way, information is tacked on as the speaker progresses, rather than integrated tightly into the text (Biber 1988: 157). In combination with demonstrative determiners, which are typically thought to be informal (*that* regard, *this* House, *this* huge pit), the tacked-on information renders evident the production constraints of time and situation. In unplanned discourse, like that in example (23), demonstrative determiners are often preferred to articles (Ochs 1979).

108 Prepared speeches in LLC retain “some spontaneity in not being read from a script,” “therefore allowing for improvisation” (Greenbaum & Svartvik 1990: 12).

Figure 5.6a: Mean scores on Dimension 6 for all genres (capitalization denotes conversational writing). Dimension 6: "On-Line Informational Elaboration" (adapted from Biber 1988: 155).



As can be seen in figure 5.6a, face-to-face conversations from both SBC and LLC are unmarked with respect to the features on this dimension (the genres with signs of on-line informational elaboration appear in the upper right end of the scale in figure 5.6a). On the other hand, figure 5.6b reveals that both genres of face-to-face conversations show an extensive spread of scores, ranging from no informational elaboration to highly on-line informational. On-line informational elaboration is typically found in personal communication where speakers elaborate on information while at the same time indicating their stance on subject matters. That is to say, it occurs where the linguistic features counted on Dimension 6 enable the encoding of attitude towards propositions, as in example (24) from a business conversation between board members.

- (24) Phil: ...Uh I would prefer that,
 that you were there on one hand,
 because I think that it would be most expedient.
- Phil: But I think,
 ..what was ..felt,
 was that at this point,
 rather than ha-
 than create
 ...I don't really f- find it to be,
 ..you know,
 ..a ...confrontation,
 by any means,
 but,
 ..
- Brad: Mhm.
- Phil: I just think,
 ..they wanna be able to just kind of ...figure out,
 I think our board eh,
 ...quite frankly we have more ...problems to resolve interior, than we
 do ..outside of it

Face-to-face conversations SBC text 10

On-line informational elaboration is often found in communication such as example (24) where a speaker's turn is fairly long, i.e. more monologic than in typical face-to-face interactions. The text from which example (24) derives scores high on Dimension 6 (6.9), this brief example indicating two THAT complements on verbs¹⁰⁹ (*would prefer that, that...*, *I think that...*) and one demonstrative

109 Biber's (1988: 230) algorithms for "THAT verb complements" detects instances of *that* preceded by a tone unit boundary, which is the case for the second *that* in *would*

(*this point*). Here again, pauses enable the speaker's mental redrafting and THAT complements are added on ad hoc to elaborate the discourse. Face-to-face conversations, on average, contain more on-line informational elaboration than do telephone conversations and conversational writing.

What about in conversational writing, then? Internet relay chat and ICQ chat are indeed both carried out in real time, on-line. Firstly, Biber's term "on-line" was coined before the advent of CMC among the general public, and must be interpreted as "live" in current situational analyses. Secondly, before we address this question we must recall what was learned in the analysis of Dimensions 1 and 3 about the nature of conversational writing; computer chat is not informational, and it is not explicit/elaborated. Rather, the discourse displayed in computer chat is highly involved and interactive, with abundant situation-dependent reference. From Dimension 6 we learn that hardly any live informational elaboration takes place in conversational writing, or rather, informational elaboration is not carried out live in conversational writing. We saw in examples (3), (9) and (15), above, indications of how the interlocutors in IRC are mainly concerned with finding conversational partners, that greetings abound and conversations in public channels rarely evolve beyond superficiality. The IRC participants rarely share personal information in their typically brief turns, producing fewer private verbs (*think, know, feel*, etc.) than for instance face-to-face conversationalists. Decisive for the low scores on Dimension 6, however, is the fact that both modes of conversational writing allow interlocutors to *edit* their contributions, in IRC before sending, and in split-window ICQ by real-time erasure and replacement, which minimizes the need to add complement clauses ad hoc. With this in mind, the low dimension scores of conversational writing on Dimension 6 come as no surprise. Example (25) illustrates a few carefully edited turns in IRC (judging from turn length, word length and complex nominal constructions), a passage in which a participant (_oups) is asking for help with his/her business assignment.

- (25) <_oups> without getting into details... what would be best... taking a depreciation allowance and use your retained profit, or taking a bank loan, with favourable interest, and inviting new shareholders (its Ltd)
- <AdamSxy35> can i take both?
- <TurKizi> anyone?
- <_oups> well yeah...but that would be a disadvantage..
- <livinboy> americans rule

prefer that, that... The first instance in the example is not counted, as it is not preceded by a public, private, or suasive verb, a seem/appear, or any other identifying item.

<AdamSxy35> oups why dont you try a business chat room on yahoo?
 <_oups> hm...well do they have that..
 <AdamSxy35> it works for me when i cant fall asleep ;)
 <CityWoman> My guess would be to take a bank loan and show positive
 cash flow and reserves.. it is always a good idea to use some-
 one elses money.
 <disingV> 27/f/single
 <Zerj> 15/m/canada with webcam and netmeeting – MSG me me
 to chat
 <peluchecamote> hello
 <AdamSxy35> can i have a loan CW?
 <^sarah^xx^> me back again
 <CityWoman> 24% AdamSxy35.. ha!
 <CityWoman> MY positive cash flow.. !
 <Jivie> AdamSxy35, can you share it? :p
 <AdamSxy35> i get better rates from guido on the street corner ;)
 <Jivie> haha
 <CityWoman> Yah but Guido breaks legs.. ha!

Internet relay chat text 5b (UCOW)

Example (25) has no instance of the features counted on Dimension 6. The example derives from a text with an overall dimension score of -3.5 and is thus typical for Internet relay chat with regard to the dimension. The text in question, Internet relay chat text 5b, is further typical of IRC in that it scores high on Dimension 1, meaning it is not informational, and it scores fairly low on Dimension 3, meaning it is not particularly elaborated in reference. The seemingly elaborated turns of this IRC example are thus the result of careful editing, prior to posting, rather than real-time elaboration. No THAT clauses are added on to objects, or brought in as complements to verbs and adjectives, and no demonstratives are used to replace articles.

While participants in IRC may take advantage of editing options, before sending, to elaborate their turns, interlocutors in split-window ICQ can edit their contributions at any time in their full semi-window. In addition to the text format, several of the texts in the split-window ICQ component of UCOW are found as video clips of the chat screens (as mentioned in section 3.3), which render chatters' redrafting explicit. Upon studying these, it becomes evident that ICQ chatters indeed use their editing options frequently, but mostly to erase passages, to replace strings of text and to correct their spelling, and very rarely to elaborate their propositions. The text format split-window ICQ corpus consists of the resulting texts after the chat session ended, and if there were any elaboration by complement THAT clauses, for instance, it would be evident in these. As can be

inferred from the dimension score of split-window ICQ chat on Dimension 6, however, such evidence is very rarely found. Example (26) from ICQ contains one demonstrative (*this*), but no other features that mark added-on informational elaboration. From the clip of the interaction displayed in example (26), it is evident that both interlocutors frequently scroll back to edit their spelling, but not to elaborate their turns.

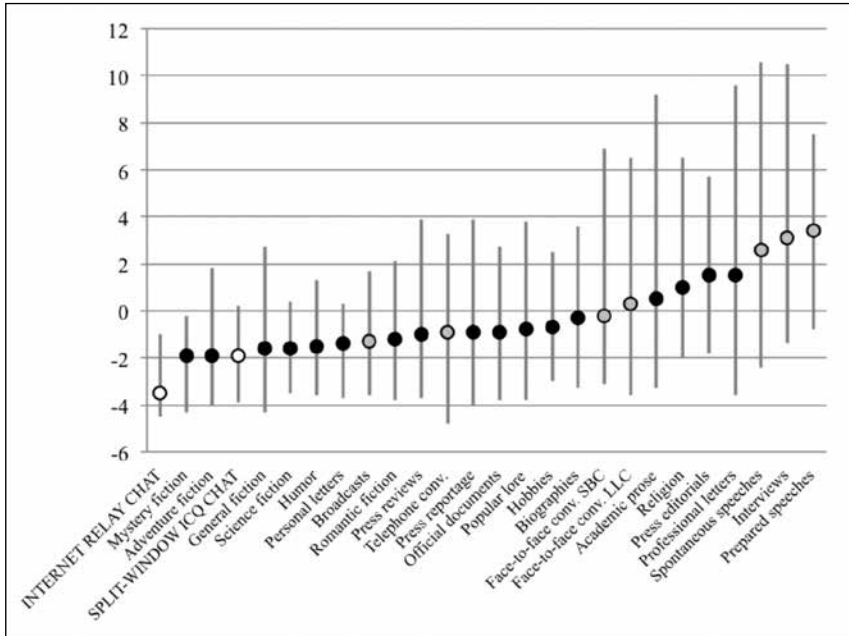
- (26) <7> hahahh did you get the lift on your truck
<7> oh yeah
<G> na, i did that when i was eating cinamon toast crunch this morning, i think im just going to get rimes and tires, i already lifted the from of my truck 3 inches
<G>
<7> if you go off roading with it why would you put rimes on it
<G> so i can have bigger tires, im not getting like 22" or nothing
<7> ok
<G> im getting 15" american racing rims, and 35 in tires, like the rimes on mike crowells jeep, or jake mitchells truck, 1100, i just don't know if i feel like paying for that cause i have the money but i want to take off a month during summer and yeah.....bye
<7> how much
<7> yeah later

Split-window ICQ chat text 6 (UCOW)

The low score of split-window ICQ chat on Dimension 6 is thus the result of an over-all lack of elaboration, though not necessarily a lack of editing. ICQ chatters simply edit their texts by scrolling back and replacing letters, words and phrases, not by elaborating their propositions post hoc.

Biber (1995: 167) tentatively labels the negative end of Dimension 6 “Edited or Not Informational.” With regard to conversational writing, both parts of the label are corroborated, even though “edited” in the live conversational writing context bears little resemblance to published traditional writing. In IRC, the rare instances of elaboration are composed prior to sending, and in split-window ICQ through live editing, and in neither case is informational elaboration added on at the end of existing turns. Figure 5.6b, finally, displays the spread of scores on Dimension 6, uncovering considerable ranges of variation in most genres. Evidently, features of on-line informational elaboration (THAT dependent clauses and demonstratives) are produced, or tolerated, in most genres, although less so in conversational writing than in conversational speech.

Figure 5.6b: Spread of scores along Dimension 6 for all genres (capitalization denotes conversational writing). Dimension 6: “On-Line Informational Elaboration” (adaptation of Biber 1988: 172–177 and 122–125, supplemented with new genres).



5.3 Chapter summary

In the present chapter, one of the primary aims of the study was accomplished; the conversational writing genres Internet relay chat and split-window ICQ chat were positioned on Biber’s (1988) dimensions of linguistic variation, alongside Biber’s multiple written and spoken genres from LOB and LLC. The focus of the chapter was to elucidate the nature of conversational writing by discussing the lexico-grammatical features that contribute to the positions of the conversational writing genres, neighboring genres and contrastive genres, on the dimensions. The distribution and functions of the features were explored, inter alia, by way of contrasting numerous textual examples from the genres. This chapter dealt with each dimension on its own terms; in the next chapter, a more overarching approach will be taken, as the results from the full investigation now may be brought together and discussed in combination.

Chapter 6. Discussion

6.1 Introductory remarks

Throughout this study, texts from the conversational writing genres have been contrasted with texts from spoken conversations, and written texts, to elucidate their qualitative, functional aspects as well as their lexico-grammatical patterns. Chapter 4, for instance, explored the interpersonal functions of modals in conversational writing and speech, and the similar lexical density in both types of communication, indicating their close relationship, and detailed the incidence in oral conversations of the salient features found in conversational writing. Ultimately, in chapter 5, the conversational writing genres and the genre of face-to-face conversations from SBC were positioned on Biber's (1988) dimensions of linguistic variation, alongside the written and spoken genres studied by Biber, and the dimension scores of Collot's (1991) genre of BBS conferencing were presented. The purpose of the present chapter is to bring together and discuss the results of the full investigation. The chapter starts out on a quantitative note and proceeds towards increasingly qualitative, multifaceted assessments.

Firstly (in section 6.2), the two hypotheses underlying the study (stated in section 1.2) are revisited quantitatively to begin to determine the relative degrees of orality in conversational writing and asynchronous CMC. This is first done by relating the positions of the conversational writing genres, and Collot's (1991) genre of APMC, to the oral conversational genres on Biber's (1988) dimensions. Secondly (in section 6.3), the overall picture afforded by all dimensions in chapter 5 is closely examined to achieve multidimensional characterizations of the conversational writing genres and the genre of APMC. The multidimensional characterizations provide the requisite input for determining the most prevalent "text types" (Biber 1989, 1995) in the CMC genres, which informs the indispensable, qualitative assessment of the results in relation to the hypotheses (Biber's notion of text types is introduced in the same section). The chapter proceeds (in section 6.4) to revisit the four research questions posed at the beginning of the study (section 1.2), the first three of which have been addressed throughout, to identify and discuss the answers to these. Among other things, the section provides a summary of the findings from the comparisons of conversational writing to writing and speech. The fourth research question, as to whether conversational writing constitutes a modality

of its own, is then addressed and answered. Finally, the working definition of conversational writing (offered in section 1.1) is revisited, in order to find out whether the definition needs to be elaborated on the basis of the findings in the full study. The last section (6.5) sums up the chapter.

6.2 Hypotheses revisited quantitatively

This section revisits the hypotheses stated in section 1.2 from a quantitative viewpoint and discusses the relationships found in the present study between the CMC genres and the spoken conversational genres. The quantitative findings will then be complemented with gradually more qualitative assessments in section 6.3, before any final conclusions can be reached regarding the hypotheses.

In chapter 1, the synchronicity of communication was presumed to contribute decisively to the linguistic character of a genre. Genres with similar synchronicity of communication were predicted to display textual similarities, despite being communicated in different media (e.g. in a medium of CMC or through the medium of speech). Conversational writing was thus expected to display similarities with oral conversation, as both involve dialogs carried out in real time. It was also suggested that the CMC genres, representing asynchronous, synchronous and supersynchronous CMC, would display different degrees of orality. The degree of orality in conversational writing was defined as the degree of linguistic correspondence to oral conversations (face-to-face and telephone conversations). The two hypotheses stated in section 1.2 are the following:

- Synchronous conversational writing displays a higher degree of orality than asynchronous CMC
- Supersynchronous conversational writing displays a higher degree of orality than synchronous conversational writing

To test the two hypotheses quantitatively, the discussion here utilizes the positions of the conversational genres on Biber's dimensions (see figures 5.1a through 5.6a in chapter 5) and the dimension scores of the genre of asynchronous CMC (presented in table 5.5). Five conversational genres are plotted on the dimensions in chapter 5, namely Biber's (1988) two genres "face-to-face conversations LLC" and "telephone conversations," and the three genres introduced in this study: "face-to-face conversations SBC" and the conversational writing genres "Internet relay chat" and "split-window ICQ chat." The dimension scores of the genre of asynchronous CMC (given in table 5.5) are those of "BBS conferencing" (studied

by Collot 1991 and originally labeled “ELC other”). The genre was not plotted on the dimensions but will nevertheless be considered here.¹¹⁰

As a genre’s degree of orality is crucially informed in this study by the genre’s proximity to oral conversations on Biber’s dimensions, it should be productive, over and above a visual inspection of the chapter 5 graphs, to measure the distances between the relevant genres. On Dimension 1, for instance, as seen in figure 5.1a, the two conversational writing genres both range in the vicinity of oral conversations, with texts displaying intense personal involvement between the interlocutors. The section accordingly explores the linguistic features that by their frequent occurrence contribute to the high scores of conversations on the dimension (first and second person pronouns, present tense verbs, direct WH-questions, etc.). From the visualization of the five conversational genres on Dimension 1 (figure 5.1a) it may be inferred that split-window ICQ chat approximates the three oral conversational genres slightly more than does Internet relay chat. By measuring the distance between the genres, in standard deviation units on the dimension, it is possible to begin to assess this quantitatively. Table 6.1 indicates the distance between each conversational writing genre and the oral conversational genres on all of the six dimensions. The table also presents the figures for BBS conferencing relative to the oral conversational genres.¹¹¹

110 The dimension scores of BBS conferencing in table 5.5 are based on the standard scores (called FDS, “feature deviation scores”) of “ELC other” reported in Collot (1991: 69–70), and are not those erroneously arrived at in Collot (1991: 77–79). Corrected dimension scores for the genre were computed in this study; see section 5.1.

111 The distances in table 6.1 are given in absolute values to enable the comparison of totals. Thus, the difference between the Dimension 1 score of IRC (25.6) and face-to-face SBC (43.7), for instance, is indicated as the positive value 18.1, i.e. as the interval of 18.1 standard deviation units on the dimension. See table 6.2 for the mean dimension scores (“mean”) of the conversational genres (derived from table 5.1 and from Appendix VIII) and table 5.5 for those of BBS conferencing.

Table 6.1: Distance of the three CMC genres to oral conversations measured as standard deviation units on each dimension (absolute values)

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6	total
Split-window ICQ to face-to-face SBC	3.5	1.6	1.7	1.5	0.0	1.7	10.0
Split-window ICQ to face-to-face LLC	11.9	1.6	0.2	0.5	0.1	2.2	16.5
Split-window ICQ to telephone conv.	10.0	0.1	1.1	0.4	0.4	1.0	13.0
total	25.4	3.3	3.0	2.4	0.5	4.9	39.5
IRC to face-to-face SBC	18.1	3.6	2.3	1.3	0.6	3.3	29.2
IRC to face-to-face LLC	9.7	3.6	0.8	2.3	0.7	3.8	20.9
IRC to telephone conv.	11.6	2.1	0.5	3.2	0.2	2.6	20.2
total	39.4	9.3	3.6	6.8	1.5	9.7	70.3
BBS conferencing to face-to-face SBC	18.4	1.7	2.8	3.4	8.0	2.0	36.3
BBS conferencing to face-to-face LLC	10.0	1.7	4.3	2.4	7.9	1.5	27.8
BBS conferencing to telephone conv.	11.9	0.2	5.6	1.5	8.4	2.7	30.3
total	40.3	3.6	12.7	7.3	24.3	6.2	94.4

As seen in table 6.1, the predicted relationship between the conversational writing genres and oral conversations on Dimension 1 appears to hold true; split-window ICQ (with a distance “total” of 25.4 units; see the leftmost column) is indeed closer to the oral conversational genres than is IRC (with a distance “total” of 39.4 units). BBS conferencing, whose Dimension 1 score is nearly identical to IRC’s (see tables 5.5 and 5.1), naturally displays a total on Dimension 1 in table 6.1 similar to that of IRC, but on Dimensions 3 and 5 it deviates markedly from the conversational genres. Judging from the totals, split-window ICQ (SSCMC) is closest to oral conversations throughout the six dimensions, and except on Dimensions 2 and 6, BBS conferencing (ACMC) is most distant from oral conversations. IRC communication (SCMC) typically ranks in the interval between these two. The totals in the rightmost column are indicative of the pattern throughout; split-window ICQ is closest to oral conversations, IRC is intermediate, and BBS conferencing is the least oral genre.¹¹² Put differently (see the bulleted hypotheses above), synchronous conversational writing (IRC) displays a higher degree of orality than asynchronous CMC (BBS conferencing) but is surpassed by supersynchronous conversational writing (split-window ICQ chat),

112 The totals in the rightmost column of table 6.1 are the sums of the standard deviation units separating the dimension scores of the relevant genres. The totals in table 6.1 are provided only to enable the surveying of all the relationships at once and must not be confused with dimension scores, as dimension scores cannot be summed across dimensions to provide an overview of the character of genres.

which displays the highest degree of orality. Judging from this calculation, both hypotheses appear to be supported.

A conscientious, statistical analysis of the relationship between the relevant genres on the dimensions, however, needs to take into account not just the crude distances between genres, but statistically valid measurements. By calculating the t-values obtaining between the genres, the variation in the data is taken into account (cf. tables 5.1b through 5.6b) as well as the number of texts in each genre (cf. table 3.1 and Appendix I). In other words, the t-value indicates the distance between the genres' mean dimension scores after accounting for these factors. T-values were obtained by using the two equations below, in which x is a conversational writing genre, y is an oral conversational genre, and n is the number of texts in the genre. The equations take into account for each genre its mean dimension score (mean) as well as its standard error of the mean (SEM), based on the standard deviation of texts. Table 6.2 presents the relevant statistics for the genres under consideration (in normal font), as well as the results of the calculation (in bold). Unfortunately, the genre of ACMC must be left out of the account here as the requisite data is unavailable for BBS conferencing (Collot 1991) and, as a result, the first hypothesis cannot be statistically tested.

$$t = \frac{\text{mean}_x - \text{mean}_y}{\sqrt{\text{SEM}_x^2 + \text{SEM}_y^2}} \qquad \text{SEM} = \frac{\text{std dev}}{\sqrt{n}}$$

Table 6.2: Distance of the conversational writing genres to oral conversations indicated as t-values on each dimension (in bold). (The input for obtaining the t-values is given in normal font)

		Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6
Split-window ICQ	vs. face-to-face SBC , t=	0.6	-1.7	-2.6	1.5	0.0	-2.2
	vs. face-to-face LLC , t=	2.9	-2.9	-0.5	0.8	-0.2	-4.4
	vs. telephone conv. , t=	2.3	-0.1	1.5	-0.5	0.7	-1.8
Internet relay chat	vs. face-to-face SBC , t=	-4.0	-3.9	-2.5	-1.1	-1.0	-4.3
	vs. face-to-face LLC , t=	-3.7	-6.7	-1.0	-2.8	-1.8	-8.2
	vs. telephone conv. , t=	-3.9	-3.4	0.5	-3.2	-0.5	-5.0
Split-window ICQ	mean	47.2	-2.2	-4.1	0.2	-3.3	-1.9
	std. dev.	13.3	1.5	1.5	1.7	1.7	1.3
	SEM	3.8	0.4	0.4	0.5	0.5	0.4
Internet relay chat	mean	25.6	-4.2	-4.7	-2.6	-3.9	-3.5
	std. dev.	7.1	1.4	2.5	2.4	1.2	1.0
	SEM	2.3	0.4	0.8	0.7	0.4	0.3

		Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6
Face-to-face SBC	mean	43.7	-0.6	-2.4	-1.3	-3.3	-0.2
	std. dev.	14.9	3.0	2.0	3.3	1.9	2.6
	SEM	4.0	0.8	0.5	0.9	0.5	0.7
Face-to-face LLC	mean	35.3	-0.6	-3.9	-0.3	-3.2	0.3
	std. dev.	9.1	2.0	2.1	2.4	1.1	2.2
	SEM	1.4	0.3	0.3	0.4	0.2	0.3
Telephone conv.	mean	37.2	-2.1	-5.2	0.6	-3.7	-0.9
	std. dev.	9.9	2.2	2.9	3.6	1.2	2.1
	SEM	1.9	0.4	0.6	0.7	0.2	0.4

The figures given in bold in table 6.2, nevertheless, are viable for addressing the second hypothesis, by which conversational writing in SSCMC should display a higher degree of orality than in SCMC. In principle, the lower the t-value (in absolute value, i.e. ignoring incidental minus signs in the comparison), the less likely is a significant difference between the genres compared. In table 6.2, the t-values for the relationship between split-window ICQ chat (SSCMC) and the spoken conversational genres are indeed generally lower than those obtaining between Internet relay chat (SCMC) and the spoken conversational genres. On Dimension 1, for instance, the t-values for split-window ICQ chat are 0.6, 2.9 and 2.3 compared to the oral conversational genres, respectively, whereas those for Internet relay chat range around 4. Similarly, on Dimensions 2 and 6, the t-values for the relationship between split-window ICQ and the oral conversational genres are also all lower than those between Internet relay chat and the latter. On Dimensions 4 and 5, the same general impression comes through, even though IRC is closer than ICQ to face-to-face SBC on Dimension 4, and to telephone conversations on Dimension 5. For the conclusive interpretation of the t-values, however, table 6.2 needs to be paired with the statistical significance of the results, i.e. with p-values (p). Table 6.3 presents the p-values for the relationship between conversational writing and conversational speech (i.e. for the same genres).¹¹³

113 The stepwise presentation of the t-tests here (via t-values) serves two purposes, both of which pertain to the replicability of the present study: 1) The t-value calculation explains how it was possible to carry out the tests (to obtain p-values) even without access to the dimension scores of Biber's (1988) individual texts. As explained, this was done via the computation of the standard error of the mean (SEM) of Biber's texts; the formulae and table 6.2 serve to clarify the procedure and the data involved. (By contrast, the same calculation was not feasible for the APMC genre, as

Table 6.3: Results from t-tests among the conversational writing genres and the conversational spoken genres. Values for probability (p), with values <.05 in bold. (Values from comparisons with SBC are repeated from table 5.3; “n.s” means “not significant.” Remaining p-values have been multiplicity adjusted)

		Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6
Split-window ICQ (SSCMC)	vs. face-to-face SBC, p=	0.5333	0.1081	n.s.	n.s.	n.s.	0.0420
	vs. face-to-face LLC, p=	0.2736	0.2592	1.0000	1.0000	1.0000	0.0275
	vs. telephone conv., p=	0.6300	0.9221	1.0000	1.0000	1.0000	1.0000
Internet relay chat (SCMC)	vs. face-to-face SBC, p=	0.0008	0.0009	n.s.	n.s.	n.s.	0.0004
	vs. face-to-face LLC, p=	0.1078	0.0030	1.0000	0.3312	1.0000	0.0029
	vs. telephone conv., p=	0.0828	0.1659	1.0000	0.2160	1.0000	0.0189

The p-values in table 6.3 specify for the corresponding t-values given in bold in table 6.2 the probability of obtaining the distances (measured in t-values) by chance if there were no difference between the dimension scores of the genres. Table 6.3 shows that a few more of the values for IRC than for ICQ are statistically significant, i.e. $p < .05$ (in bold). The p-values for IRC on Dimensions 6 all indicate a significant difference from the oral genres, as do two of IRC’s p-values on Dimension 2 and one on Dimension 1; i.e. they indicate that the dimension scores of IRC are significantly different from the oral conversational genres in question. As it is the proximity of genres that is at issue here, however, high p-values are also informative for the interpretation of the results; the preponderance of high p-values (in normal script) for the relationship between split-window ICQ and oral conversations shows that split-window ICQ chat is not significantly distinct from oral conversations; on Dimensions 1 through 5 no difference obtains; the only indications of a discrepancy are found on Dimension 6. The wealth of high p-values for IRC, moreover, indicates roughly the same relationship; IRC is generally not distinct from the oral conversational genres, either. On Dimensions 3 and 5, all the conversational genres in effect coincide (“ $p=1.000$ ” meaning that no measurable difference was found) and on Dimension 4, no statistical difference obtains between the written and the oral conversations. The statistical tests thus establish that on most dimensions, supersynchronous CMC, as represented by split-window ICQ chat, is lexico-grammatically more similar to spoken conversations than is synchronous CMC, as represented by Internet relay chat. On

no standard deviation is specified in Collot 1991.) 2) The t-values in table 6.2 show discernible trends for the data, which might be of interest in future studies (even if the corresponding multiplicity adjusted p-values in table 6.3 indicate no difference for the genres in question).

the first five dimensions, split-window ICQ is an inherently “oral” genre, whereas IRC is marginally less “oral.”

On Dimension 6, labeled “On-line Informational Elaboration,” both conversational writing genres deviate from face-to-face conversations (see the p-values in table 6.3). This finding mirrors the discussion of Dimension 6, in section 5.2.6, in which the sparsity in conversational writing of dependent clauses (THAT complement clauses on verbs and adjectives and THAT relative clauses on object position) and demonstratives (*that, this, these, those* preceding nominals) were found to yield noticeably low dimension scores for the conversational writing genres. Chatters’ turns contain extremely few elaborations added on “live” by way of complement clauses, as in for instance spontaneous speeches, but also contain fewer complement clauses and demonstratives preceding nominals than most of the written genres. On the other hand, while IRC largely lacks markers of on-line informational elaboration, split-window ICQ and telephone conversations remain similar even on this dimension (as no measurable difference was found between the latter two). In the assessment of the orality of conversational writing here, however, we must be cautious not to attach too much importance to Dimension 6. As noted in section 5.2.6, the dimension has few linguistic features with important loadings (Biber 1988) and most studies, including Biber’s more recent ones, have consequently left it out of account. The discourse observations with respect to Dimension 6 in section 5.2.6 are thus only tentative, and in the further discussion, section 6.3 below, the sixth dimension is eventually phased out.

Before proceeding, a few remarks to sum up the present section are in order. In the quantitative assessment of the hypotheses here, two main findings have emerged. Firstly, the dimension scores of the ACMC genre generally position the genre at a greater distance than the conversational writing genres from the oral conversational genres on Biber’s (1988) dimensions. This finding lends support to the first hypothesis, by which asynchronous CMC should be less “oral” than synchronous CMC (or conversational writing at large). Secondly, in the statistical tests of the positions of the conversational writing genres on the dimensions, the supersynchronous genre (split-window ICQ) was found to be substantially correspondent to oral conversations, whereas the synchronous genre (IRC) appeared to be marginally less “oral.” In other words, the statistical tests offer a measure of evidence supporting the second hypothesis, that the supersynchronous genre displays a higher degree of orality than the synchronous genre. In the next section, with gradually more qualitative assessments, we will find out whether these initial propositions hold. The section will show that, by looking beyond genre boundaries

to find similarities and differences across texts, a complementary approach to discourse variation is feasible, one which provides illuminating results.

6.3 From genres to text types

What is the linguistic nature of conversational writing? The question was raised as the first of four research questions in section 1.2, laying out the aim and scope of the study. Chapters 4 and 5 have provided ample textual examples to illustrate typical messages exchanged over the computer networks. Discussions of these recurrently recognized that the texts from both conversational writing genres resemble oral conversational texts to a great degree, syntactically and lexicogrammatically, and deviate notably from traditional written genres in many respects. The most salient features in conversational writing (frequent first and second person pronouns, direct WH-questions, analytic negation, demonstrative and indefinite pronouns, present tense verbs, predicative adjectives and contractions, and infrequent prepositional phrases), explored in chapter 4, were found to be decisive contributors to the oral character of the textual chats. Most of the features were then revisited in chapter 5 in the discussion of Dimension 1, distinguishing involved texts from informational texts. On other dimensions, conversational writing was set apart from most written genres by, for instance, a marked paucity of certain linguistic features, as on Dimension 5 on which the non-abstract chats are diametrically opposed to the genres of stereotypical abstract writing (official documents and academic prose). At this point, the multitude of findings can be interrelated and the discussion of the hypotheses (stated in section 1.2 and repeated in section 6.2) brought forward. First, the discussion here reflects on the overall picture afforded by all dimensions (see also chapter 5 for dimension graphs and concomitant in-depth descriptions). Next, the dimension score patterns across the first five dimensions will be traced to identify the “text types” (Biber 1989, 1995) of the conversational writing texts and the ACMC genre, which, in turn, will enable the conclusive assessment of the hypotheses.

The following slightly simplistic characterization of conversational writing can be made on the basis of the genre means on Biber’s dimensions explored in chapter 5 (parentheses indicating Dimension numbers). Conversational writing is involved (1), non-narrative (2), situation-dependent (3), non-argumentative (4), non-abstract (5) discourse, containing very little real-time informational elaboration (6). Oral conversations are also typically involved (1), situation-dependent (3), non-argumentative (4) and non-abstract (5), but are slightly more narrative (2) and contain more real-time informational elaboration (6) than

chats. The multidimensional characterization of the ACMC genre will be traced shortly; the discussion here first zooms in on the conversational genres.

In section 6.2, the results from statistical tests of the relations between conversational writing and conversational speech were presented. It was seen there that, except on Dimension 6, split-window ICQ is not distinct from oral conversations, and that IRC is only marginally less “oral” than the former. The discrepancies between IRC and the spoken conversational genres all appeared on Dimensions 1, 2 and 6 (see p-values in table 6.3). The findings on Dimension 6 were discussed in section 6.2 with regard to both conversational writing genres; the remaining discrepancies (i.e. those between IRC and oral conversations) are touched upon here, although as will be seen, the similarities across the conversational genres outweigh the differences.

On Dimension 1, the five conversational genres contrasted in this study (IRC, face-to-face and telephone conversations from LLC, face-to-face conversations from SBC and split-window ICQ chat) all reside, in the order given, at the involved end of the scale. The Dimension 1 graph (figure 5.1a) visualizes their distinctive positions, as the dimension separates the texts with involved production from those with informational production. *Prima facie*, split-window ICQ chat appears to be more “involved” than the oral conversational genres, i.e. to display a degree of orality beyond all of theirs (if the positive end of the dimension is taken to be the oral end). On closer inspection, however, split-window ICQ is not significantly different from either face-to-face or telephone conversations (as $p > .05$ in table 6.3). Instead, split-window ICQ is greatly akin to these, which also rules out the possibility theorized in chapter 1 of supersynchronous chats exceeding oral conversations in orality, a relationship that would have called for a redefinition of orality here. In other words, the definition of orality employed in this study (the similarity to oral conversations) still holds.

On Dimension 1, split-window ICQ, like the oral conversational genres, displays abundant markers of involvement (private verbs, THAT deletion, contractions, etc.; see section 5.2.1), reflecting real-time production circumstances. In fact, both conversational writing genres are lexico-grammatically akin to the oral conversations on this dimension, even though the IRC discourse is more moderately involved than the split-window ICQ conversations and also statistically different from these ($p < .05$ in table 5.3), as well as from face-to-face conversations SBC ($p < .05$ in table 6.3). On Dimension 2, “Narrative versus Non-Narrative Concerns,” IRC displays non-narrative discourse similar to that of telephone conversations (with few past tense verbs, few third person pronouns, etc.; see section 5.2.2) and has fewer narrative features than the face-to-face conversational

genres (as $<p.05$ in table 6.3). The latter range in intermediate position on the dimension, unmarked for narrativity, and are thus not particularly concerned with narration, either. (As Dimension 2 is not associated with a literate-oral polarity, the position of IRC on the dimension is unproblematic for the definition of orality; only on Dimensions 1, 3 and 5 can a genre “exceed” oral conversations, i.e. surpass them at the oral end, and this not the case, statistically, for any conversational writing genre.)

Apart from the few discrepancies mentioned (pertaining to IRC), the overall examination of Dimensions 1 through 5 in chapter 5 yields solid results for both conversational writing genres as regards their similarity to oral conversations, not least on Dimensions 1, 3 and 5, the three dimensions seen to “identify sharp distinctions between ‘oral’ and ‘literate’ registers” (Biber 2008: 843, also noted in e.g. Biber & Finegan 2001). On Dimension 3, “Explicit/Elaborated versus Situation-Dependent Reference,” neither of the chat genres is different from oral conversations (cf. table 6.3); conversational writing, like the oral conversational genres, displays discourse with frequent markers of situation-dependent reference (e.g. time adverbials) and a sparsity of elaborating devices (such as WH relative clauses); see section 5.2.3. On Dimension 4, the conversational genres are all generally unmarked for overt expression of persuasion/argumentation and, even though certain split-window ICQ texts contain more opinionated discourse than most IRC texts, no statistical difference obtains between the conversational genres on the dimension (cf. table 6.3). On Dimension 5 (as seen in figure 5.5a), the conversational writing genres both practically coincide with the oral conversational genres on the non-abstract/non-impersonal end of the dimension. Neither split-window ICQ nor IRC is statistically different from the oral conversations; rather, all conversational genres display a decisive paucity of markers of abstract information (e.g. conjuncts and agentless passives).

As touched upon in the previous section, Biber and other linguists carrying out post-Biber (1988) MD analyses have, over the years, paid diminishing attention to Dimension 6, identified in Biber (1988). As early as 1989, Biber ignores the dimension, asserting that “five major dimensions have been identified in English” (1989: 7). Biber’s (2008) account of multidimensional approaches mentions the sixth dimension, but also elaborates only on the first five, seeing that the sixth dimension “has few salient linguistic features” (2008: 836). Moreover, Biber (2008) notes that Dimensions 2 and 4 “have no systematic relationship to speech and writing” (2008: 843). Even though all dimensions must be considered for the full picture (as described in section 1.2), Biber (2008) argues that Dimensions 1, 3 and 5 have been seen to most clearly set the oral genres, especially

conversations (“stereotypical speech” 2008: 843), apart from the written genres. In the present study, however, conversational *writing* is observed to intermingle with oral conversations on most dimensions, and most notably on Dimensions 1, 3 and 5. Conversational writing and conversational speech have been found to be closely related functionally, irrespective of their genre classifications, as ways for writers/speakers and readers/listeners to interact personally in the immediate present (synchronously and/or supersynchronously), generally with the purpose of furthering interpersonal relationships. The multitude of textual properties found to be common to all the conversational genres makes their affiliations with the written or spoken medium rather irrelevant; instead, it is the immediacy of the situation, the synchronicity, the presence of a responsive audience, and the attendant social practices that determine the nature of the discourse (as found in section 4.2). Texts are not confined to genre boundaries; rather, texts may display similar linguistic characteristics across genres.

In his 1989 and 1995 studies, Biber offers an apt, complementary view of textual variation, bringing the 1988 study forward from defining the genres in situational/functional terms to defining the “text types” with maximally similar linguistic properties (Biber 1989, 1995). While genres are determined on the basis of external criteria such as the purpose of the author/speaker and the production circumstances, text types are groupings of texts that are similar in their linguistic form (with respect to dimension characteristics), irrespective of their genre classifications. Text types thus cut across genre boundaries, offering variationists a complementary way “to dissect the textual space of a language” (Biber 1995: 320). A single text type might include texts from several different genres. The text type “scientific exposition” for instance, marked as very informational (integrated) on Dimension 1, non-narrative on Dimension 2, elaborated on Dimension 3, etc., includes texts from academic prose, official documents, and a few more genres, all texts sharing the same linguistic characteristics. Conversely, texts from a single genre can be distributed across several text types; academic prose, for instance, is represented across four text types (3, 4, 6 and 8). Table 6.4 summarizes the eight text types identified among the Biber (1988) texts, types detailed in Biber (1989, 1995), indicating for each text type genres in which it occurs and its multidimensional characterization.

The text types in the English language (table 6.4) were identified empirically in Biber (1989) by way of a multivariate statistical procedure called cluster analysis. As input to the analysis, Biber used the dimension scores of the Biber (1988) texts. The analysis grouped texts with maximally similar dimension scores on Dimensions 1 through 5 into clusters, assigning every text to some cluster. Biber found

each cluster to represent one text type, and assigned each text type an interpretive label; thus, cluster 1 represents text type 1, which he labeled “intimate interpersonal interaction”; cluster 2 represents text type 2, which he labeled “informational interaction,” etc. For each text type, Biber traced the typical dimension characteristics of texts, which resulted in the multidimensional characterizations given in table 6.4. The multidimensional characterizations represent the most central texts of the cluster (those closest to the centroid of the cluster).¹¹⁴ In what follows, we will identify the text type(s) of conversational writing by relating its texts

Table 6.4: Summary of English text types (adapted from Biber 1995: 328–331)

Text type	Found in genres	Multidimensional characterization
1. Intimate interpersonal	face-to-face conversations telephone conversations	Dimension 1: Extremely involved Dimension 2: Unmarked Dimension 3: Situated Dimension 4: Unmarked Dimension 5: Non-abstract
2. Informational interaction	face-to-face conversations telephone conversations interviews spontaneous speeches personal letters broadcasts professional letters general fiction	Dimension 1: Very involved Dimension 2: Unmarked Dimension 3: Situated Dimension 4: Unmarked Dimension 5: Non-abstract
3. “Scientific” exposition	academic prose official documents biographies press reviews hobbies press reportage	Dimension 1: Very informational Dimension 2: Non-narrative Dimension 3: Elaborated Dimension 4: Non-persuasive Dimension 5: Extremely abstract

114 In table 6.4, the genres in each text type are listed according to the centrality of their texts (cf. Biber 1989), i.e. from genres with more central texts in the cluster to genres with more peripheral texts (except for in text type 1, in which the texts of both genres are essentially equally central in the cluster).

Text type	Found in genres	Multidimensional characterization
4. Learned exposition	academic prose press reportage official documents press reviews popular lore biographies professional letters hobbies religion press editorials	Dimension 1: Extremely informational Dimension 2: Non-narrative Dimension 3: Very elaborated Dimension 4: Non-persuasive Dimension 5: Moderately abstract
5. Imaginative narrative	general fiction romantic fiction mystery fiction adventure fiction prepared speeches interviews science fiction popular lore biographies personal letters religion	Dimension 1: Moderately involved Dimension 2: Extremely narrative Dimension 3: Situated Dimension 4: Unmarked Dimension 5: Non-abstract
6. General reported exposition	press reportage press editorials academic prose general fiction religion humor biographies press reviews hobbies broadcasts prepared speeches adventure fiction science fiction mystery fiction popular lore professional letters	Dimension 1: Informational Dimension 2: Moderately narrative Dimension 3: Unmarked Dimension 4: Unmarked Dimension 5: Unmarked

Text type	Found in genres	Multidimensional characterization
	official documents romantic fiction	
7. Situated reportage	broadcasts science fiction mystery fiction hobbies	Dimension 1: Unmarked Dimension 2: Non-narrative Dimension 3: Extremely situated Dimension 4: Very non-persuasive Dimension 5: Non-abstract
8. Involved persuasion	interviews spontaneous speeches academic prose popular lore professional letters hobbies religion press editorials personal letters prepared speeches telephone conversations broadcasts humor general fiction	Dimension 1: Unmarked Dimension 2: Non-narrative Dimension 3: Moderately elaborated Dimension 4: Very persuasive Dimension 5: Unmarked

to the text types (i.e. clusters) identified in Biber (1989, 1995). The consideration of text types here is intended to elucidate not just the character of conversational writing, but also, eventually, that of asynchronous BBS conferencing (Collot 1991), all in order to inform the assessment of the hypotheses.

Biber (1989: 15) explains the cluster patterns by first illustrating the clusters formed by combining the Dimension 1 and 3 scores of individual texts into a graph, in which Dimension 1 constitutes the x-axis and Dimension 3 the y-axis. The dimension scores for the individual text on both dimensions are plotted in their point of intersection in the graph, each numbered with the text's text type. To begin to identify the text type(s) of the conversational writing texts in the present study, their dimension scores were plotted onto this graph, which assigned most of the split-window ICQ chats positions in the text type 1 cluster and most of the IRC texts plots in the text type 2 cluster of texts. The dimension scores of the face-to-face conversations from SBC were also plotted onto the graph, expectedly yielding a position for most of the texts in the text type 1

cluster.¹¹⁵ Finally, the mean dimension 1 and 3 scores of Collot's (1991) genre of ACMC, named "BBS conferencing" in the present study (table 5.5), were plotted onto Biber's (1989: 15) graph of cluster distributions, giving the genre a position amidst the text type 2 texts, albeit more distant from the cluster centroid than most of the IRC texts.¹¹⁶

Throughout this study, conversational writing has, implicitly or explicitly, been suspected to be maximally similar to the texts of face-to-face and telephone conversations, which all appear in text types 1 and 2 (except for one outlier telephone conversation in text type 8). By first relating the texts on Dimensions 1 and 3, and next, analyzing the dimension score distribution of the conversational writing texts with respect to Dimension 2 (ICQ texts relatively unmarked, IRC texts non-narrative), Dimension 4 (most texts relatively unmarked for persuasion/argumentation) and Dimension 5 (most texts non-abstract), it is possible to conclude that most of the split-window ICQ texts indeed adhere to text type 1 and that most of the Internet relay chats fall under text type 2 (as no other text type suits their multidimensional characterization better; see table 6.4). Furthermore, most of the face-to-face conversations from SBC follow the same text type 1 pattern as split-window ICQ across the five dimensions.¹¹⁷

That split-window ICQ chatting, like most of the face-to-face conversations from SBC, represents "intimate interpersonal interaction" (text type 1) has been exemplified recurrently throughout the present study (prior to this identification

115 The plotting of texts discussed here was carried out by hand onto Biber's (1989: 15) graph of cluster distributions with respect to Dimensions 1 and 3. The resulting graph is not produced here, as no numeral dimension score values are available for Biber's (1988) individual texts, rendering the production of a new graph for print here unfeasible.

116 As the dimension scores of Collot's (1991) individual texts are unavailable, the plotting and multidimensional characterization of BBS conferencing here are based on the mean dimension scores of the genre (table 5.5) rather than on the scores of individual texts.

117 Carrying out a new comprehensive cluster analysis of Biber's (1988) texts and those investigated in the present study was not feasible, as the dimension scores for Biber's (1988) individual texts are unavailable. (Regardless, it is unlikely that adding a few dozen texts to Biber's 481 texts in the analysis would much alter the cluster patterns even if this were done.) Instead, to inform and substantiate the qualitative account given here, the dimension scores of texts in the present study, and those of the ACMC genre, were related to the cluster centroids of Biber's (1988, 1989, 1995) texts via a computation of the Euclidean distance obtaining between the scores and the centroids. The results of the computation are provided in Appendix IX.

of the text type). The split-window ICQ texts display, for instance, abundant private verbs, contractions, first and second person pronouns and present tense verbs, and a sparsity of nouns, prepositional phrases, etc., associated with extremely high scores on Dimension 1; frequent time and place adverbials, and few WH relative clauses, contributing to low scores on Dimension 3; and particularly infrequent conjuncts, agentless passives, etc., yielding low scores on Dimension 5. Labeling Internet relay chatting as “informational interaction” (text type 2), however, is slightly counter-intuitive.

The IRC texts have been more difficult than the split-window ICQ chats to characterize throughout. A few of the textual examples (e.g. examples 2, in section 3.2, and 4, in section 4.2) show interlocutors engaged in fairly involved, interpersonal interaction, but such passages are usually brief in the flow of IRC communication (i.e. extend over a restricted number of turns). Owing to the multitude of participants, conversational turns are more often interrupted. Typically, newly arrived participants’ greetings and questions break up adjacency pairs (as in example 21 in section 4.4), fragmenting the discourse. Herring (2013b: 252) notes that such disrupted adjacency makes multiparty chat systems “noisy communication environments.” As mentioned in connection with example (3) in section 5.2.1, IRC conversations (i.e. those involving two or more interlocutors in coherent exchanges) are frequently interspersed with turns consisting of one mere keystroke (? , ; , 2), compliments, phatic expressions and other attention-attracting tropes (*hey, gret thanks cheeky, grrrr u*), which eventually cause coherent conversations to wane. In this “noisy” environment, chatters are less inclined than those in split-window ICQ to produce extremely involved discourse, judging by the slightly fewer “involved” features on Dimension 1 (e.g. present tense verbs, analytic negation, first person pronouns) that surface in the IRC texts as compared to the split-window ICQ chats. Nevertheless, the IRC texts assume a “very involved” position on Dimension 1, which renders the text-type label “informational interaction” for IRC communication a misnomer.

In a study of web chat, that is, chat rooms similar to the channels of IRC, Sveningsson (2001: 58) observes that the communication resembles “multiparty telephone conversations (telephone chat lines), of the kind that used to be called ‘Heta Linjen’ (the Hot Line) in Sweden.” Sveningsson explains:

These multiparty telephone conversations should not be confused with what is referred to as hotlines today, where the main purpose seems to be phone sex. The former type [i.e. the Hot Line] consisted in telephone numbers that had no subscriber, to which people could call free of charge. The knowledge of those numbers was spread through personal communication between young people, and can indeed be seen as one of many strategies

to avoid the governance of adults and organizations, since these media provided a free and un-moderated space, in which adults had little insight. (Sveningsson 2001: 58)

Hot Line communication, popular among young Swedes in the 1980s, involved multiparty telephone conversation between several individuals, all with the intention of finding new friends. The discourse greatly resembled the web chat, and IRC, interaction of today in that it typically contained “a jumble of voices shouting ‘Hello?’, ‘Hello?’, ‘Hello?’, ‘Who are you?’, ‘What’s your name?’, ‘Hello?’, and so on” (2001: 58). Sveningsson also notes that “callers often exchanged personal telephone numbers at an early stage, to be able to call each other up and have a more coherent conversation” (2001: 58), which in IRC is accomplished by initiating the direct client-to-client protocol (see section 4.5). Speaking from personal experience, the present author agrees with Sveningsson’s description of the telephone chat lines and their resemblance to online chat rooms/channels. The discourse in public Internet relay chat channels is indeed as disjointed and ephemeral as was that on the telephone chat lines, even though both also contain passages of coherent conversation.

The participants competing for attention in IRC all value extreme brevity and quick responses. The profusion of greetings and attention-attracting tropes, however, brings about “a low signal to noise ratio” (Herring 1996b: 105) in the discourse, meaning that the verbal “flurry” of the multiparty chat does not transmit a great deal of useful information (Mann & Stewart 2000: 184). In the words of Crystal (2001), it rather resembles “a cocktail party in which everyone is talking at once – except that it is worse, because every guest can ‘hear’ every conversation equally, and every guest needs to keep talking in order to prove to others that they are still involved in the interchange” (2001: 159). Put simply, IRC discourse is disjointed multiparty conversational writing. The characterization of IRC as “informational interaction” (text type 2) here is thus misrepresentative, but apparently inevitable given the multidimensional distribution of the texts’ linguistic features. Certainly, the multidimensional character of text type 2 embraces the IRC texts more closely than does any other text type, but upon including Internet relay chats among the texts in text type 2, altering the text type label would perhaps be justified, which necessarily requires further research (outside the scope of the present study).¹¹⁸

Asynchronous CMC discourse for social interaction, such as that in Collo’s (1991) genre of BBS conferencing (“ELC other”; see *inter alia* sections 4.1, 4.2

118 To facilitate further research, the dimension scores of individual texts of the corpora annotated in the present study are given in Appendix X.

and 5.1), on the other hand, appears to more justly conform to the original text type 2 label, although more for its resemblance to the non-conversational texts in text type 2 than any resemblance to Internet relay chat or oral conversations. The dimension scores of the genre indicate that its discourse, like that of IRC, is moderately involved on Dimension 1 and, like ICQ, non-narrative on Dimension 2, but unlike conversational writing, the asynchronous discourse is unmarked for situation-dependent reference on Dimension 3, persuasive/argumentative on Dimension 4 and remarkably abstract on Dimension 5. Despite these differences on Dimensions 3 through 5, the overall distance between the dimension scores of the BBS conferencing genre and the cluster centroids of Biber's texts objectively indicate that the genre adheres to the text type 2 cluster (specified in Biber 1995 and in Appendix IX here), albeit more peripherally so than most of the IRC texts (the APMC genre being closer to text types 3 through 8 than are most of the conversational writing texts). At the same time, the dimension scores of the asynchronous genre are more distant from the text type 1 cluster than are those of most IRC texts.

While the correspondents in BBS conferencing may formulate long, thought-out verbalizations, the IRC and ICQ interlocutors, like oral conversationalists, are subject to time constraints and produce brief, impromptu turns analogous to those in unplanned, spoken interaction. Biber (1995: 328) points out that the principal difference between text type 2 and text type 1 texts "relates to the primary purpose of the interaction: to convey information in text type 2 and to maintain the interpersonal relationship in text type 1." Collot (1991) notes that "one of the primary purposes for participating in a BBS is to seek and impart information" (1991: 86) and that the APMC texts are easily compared to interviews, as well as to letters: "In much the same way as personal and professional correspondents, participants in [BBS conferencing] share neither the same physical nor the same temporal context" (1991: 89). The latter fact, to be sure, sets the APMC texts apart from the synchronous and supersynchronous CMC texts; Collot (1991) observes that participants' separation in time and space "may be at the root of the resemblance between the [BBS conferencing corpus] on the one hand, and personal and professional letters on the other" (1991: 90). As the APMC texts are produced for asynchronous delivery, the authors rely less on the immediate situation and instead, like letter-writers, produce for instance more WH relatives in subject position and more nominalizations (on Dimension 3) than in conversational writing. They usually have the required time to study incoming messages and to carefully prepare their argumentation and thus include e.g. more conditional adverbial subordinators: *if*, *unless* (on Dimension

4). The most striking difference between the ACMC genre and the conversational writing genres here, however, is the far more abstract discourse produced in the asynchronous genre, giving it a score on Dimension 5 on a par with official documents. Abstract/impersonal content is composed by way of e.g. conjuncts (*furthermore*, *moreover*, *nevertheless*, etc.), adverbial subordinators (such as *since*, *while* and *whereas*) and BY passives (Biber 1988), features largely lacking in conversational writing as well as in conversational speech. On the whole, it is on two of the three dimensions separating literate and oral genres, Dimensions 3 and 5, that the asynchronous genre deviates most from the conversational genres (cf. table 6.1). Korsgaard Sorensen (1993) finds CMC texts to reflect the time interval between exchanges; when exchanges “occur at longer intervals in time,” they display features of “prototypical written interaction” (1993: 406). That the production of asynchronous computer-mediated texts admits of thought-out, carefully composed verbalizations is evident not just in the more elaborate references and the frequent features of abstract discourse, but also in the word length and TTR of the ACMC texts, both more similar to those of writing than of speech (as seen in section 4.3) and in the high lexical density of ACMC texts (although the latter is not one of Biber’s features). The genre of asynchronous CMC, in sum, is deemed to consist of texts that, unlike the conversational writing texts, are peripheral to oral conversations.

In conclusion, rating from their convergence to the multidimensional characterizations for the text types in Biber (1989, 1995), the conversational writing texts and the asynchronous genre studied in this analysis divide up. The split-window ICQ chat texts (SSCMC) most closely resemble oral conversations of text type 1, “intimate interpersonal interaction,” and the Internet relay chat texts (SCMC), despite their defying easy classification, adhere to text type 2. However, judging from the incidence of oral conversations in both text types, and the qualitative, functional assessments here, IRC communication is no less “oral” than split-window ICQ chat, only less extremely involved (intimate). BBS conferencing (ACMC), on the other hand, more peripherally than the IRC texts in multidimensional character, but more justly in terms of its function, conforms to the text type 2 label “informational interaction.” The asynchronous genre shares more functional and lexico-grammatical characteristics with the non-conversational genres in text type 2 than does Internet relay chat and may, consequently, be regarded as more distant from face-to-face and telephone conversations than is conversational writing, i.e. as not an “oral” genre in the terms of this study. The discussion in this section has thus complemented the quantitative assessment in the previous section and necessitated a refinement of the proposition

enounced there regarding the hypotheses. The findings, in combination, support the first hypothesis, i.e. that conversational writing displays a higher degree of orality than the genre of asynchronous CMC, but do not provide evidence for confirming the second hypothesis, that split-window ICQ chatting should be more “oral” than IRC communication. Rather, the conversational writing texts, in both SSCMC and SCMC, are all closely related to oral conversations; the split-window ICQ chats are merely more intimately interpersonal in character than the Internet relay chats.

6.4 Research questions revisited

In section 1.1 of the present study, a working definition of conversational writing was formulated, by which conversational writing is written communication 1) for social interaction 2) which requires the simultaneous presence (physical or virtual) of producer and recipient, 3) in which interlocutors expect immediate feedback (i.e. within seconds) and 4) during which the discourse may be reconfigured by the participants while under construction (e.g. as interlocutors are able to influence each other’s line of thought). None of the findings that have emerged in the study have occasioned an alteration of these premises (other than that the conversational writing discourse investigated here is exclusively computer-mediated and not conveyed via note-passing, also regarded as conversational writing in section 1.1). In the present section, a selective overview of the results will be given with the primary aim of discussing and summarizing some of the answers provided to the research questions posed in section 1.2, and the secondary aim of finding out whether the definition of conversational writing needs to be elaborated on account of the discussion.

Four research questions were posed at the outset of this study (the first three of which have been addressed and the fourth will be addressed shortly):

- What is the linguistic nature of conversational writing and the genres studied here, IRC and split-window ICQ chat?
- How does conversational writing carried out in SCMC and SSCMC, respectively, relate to writing and speech?
- How do the genres of SCMC, SSCMC and ACMC relate to oral conversations on Biber’s (1988) dimensions?
- Does conversational writing carried out in SCMC and SSCMC constitute a modality of its own?

It stands to reason that any attempt to summarize the answers to the research questions in one brief section is bound to be slightly simplistic and might fail to

reflect the complexity of the results (thus, for the comprehensive answers readers are referred to chapters 4 and 5, as well as to the entirety of the present chapter). Nevertheless, the task is taken on here; firstly with regard to the first research question, as to what conclusions about the nature of the conversational writing genres may be drawn on the basis of the full investigation. (The nature of conversational writing is first discussed in functional linguistic terms, before the more specific, text-linguistic findings are summarized in bullet points.)

Conversational writing is not a homogeneous entity; rather, just like oral conversations, it occurs in endless constellations of contexts, for a variety of purposes, on infinite numbers of topics, between any two or more people. A major difference between conversational writing and face-to-face conversations, however, is the limited *shared* context in the former, a context even more limited than in telephone conversations. While face-to-face interlocutors share the physical surrounding (and with all senses available to them take in the visible objects, background sounds, scents, temperature, etc., as well as the non-verbal cues signaled by the conversational partner) and telephone conversationalists share the audible surroundings (and are able to perceive clues as to each other's sentiments and e.g. turn-yielding signals, such as changes in vocal pitch), conversational writers (in the CMC media studied here) are confined to the interface shared on their screens, mostly to the text conveyed. This limited semiotic field (Halliday 1985a, 2004, Martin 2001a; cf. section 2.4) naturally impinges on interlocutors' language, but not as much as one might first imagine. In this study, conversational writing texts have been found to be remarkably similar to transcribed oral conversations (with prosodic annotations removed in the latter), while at the same time inherently different from most other written genres. The split-window ICQ chats, however, have been found to be of a more close, interpersonal character than the Internet relay chat texts (cf. section 6.3).

The social relationships formed in the conversational writing interface (i.e. the semiotic tenor of the interaction; cf. section 2.4) largely depend on what the particular chat client allows (in terms of number of participants). In IRC, multiple individuals (dozens at once in the IRC corpus studied) in remotely separated localities convene in the virtual rooms. Some chatters frequent the same channels, keeping the same nickname, which means that close relationships may form between regulars (Mar 2000), although these are more often maintained in the direct client-to-client protocol than in the public channel. Most IRC chatters in public channels, however, are not previous acquaintances and rarely meet in real life. As a result, they have little at stake if they are not appreciated or accepted in the chats, especially in the public channels, as they can simply leave the channel

without embarrassment (Mar 2000). At the same time, chatters in ICQ more often use the medium to complement or extend real-life interactions and discuss matters from their occasionally shared real-life context, expecting to be held responsible for views expressed. The results presented in section 5.2.4, for instance, show that IRC chatters in public channels rarely exchange views in animated ways, whereas the semiotic mode in ICQ occasionally brings about extended supportive or challenging argumentation.

The ICQ chatters in the corpus studied, unlike most IRC chatters, have met, and regularly meet (met), face-to-face. At the time of the recording, the ICQ chat client was designed to handle only a few (two to three) participants chatting at a time, and was mostly used for communication between those with a previous real-life relationship (via an editable, personal list of friends). The semiotic mode of split-window ICQ chat, at least for the participants in the present study, is therefore different from that in IRC, i.e. the language exchanged plays different roles for participants. In public IRC (or in web chat), the interface is used to initiate or maintain mostly fleeting relationships, and in ICQ (or, in IM) it is used to maintain or further existing relationships. These fundamental properties of the two conversational writing genres studied, their different semiotic tenor and mode, contributing to a typically lower signal-to-noise ratio in multiparty IRC (section 6.3), more than the genres' respective synchronicity of communication (synchronous vs. supersynchronous) have been found to have a bearing on the results in this study. Whereas the public IRC chats are disjointed, superficial, rapid-fire exchanges between multiple (i.e. >2) participants, the ICQ discourse, just like most of the spoken conversational material, consists of intimate, personal, conversations, ranging from adversarial to affective, between two or three previously acquainted individuals. Differences in the lexico-grammatical make-up of the IRC and ICQ texts are thus due more to the situational, client-imposed, cultural and semiotic factors associated with the respective genres than to the sheer difference in synchronicity (synchronous vs. supersynchronous communication) between the two. More precisely, although a potential supersynchronicity effect is vaguely discernible in the more common use of the inserts "response forms" and "hesitators" in the split-window ICQ chats (as seen in section 4.6), no all-round supersynchronicity effect is evidenced in the material that would liken the supersynchronous chats, more than the synchronous, to spoken conversations.¹¹⁹

119 Inserts of the categories mentioned are not included among Biber's (1988) features and, consequently, have no bearing on the quantitative results of the MD analysis. Upon inclusion of conversational writing genres in future MD analyses, the

Rather, the marginally “higher” degree of orality in split-window ICQ found in section 6.2 (in the quantitative assessment) and the adherence of split-window ICQ chat to text type 1 (“intimate interpersonal interaction”; see section 6.3) are more likely due to the similarity in semiotic tenor and mode of the supersynchronous chats and the oral conversations. That is not to say that a supersynchronicity effect is ruled out. Observing and establishing the existence of such an effect, however, would require not only access to high-quality video recordings of the supersynchronous chats for an analysis of overlapping sequences (which was not the case here, as mentioned in section 4.6), but also that the synchronous material (i.e. the control group material) be acquired from channels, or chat clients, that allow only two or three participants, participants who also preferably are previous acquaintances, to make for maximally comparable situational settings. In hindsight, such a research design might have been preferable.

That said, there still remains a substantial synchronicity effect evident in conversational writing, which likens conversational writing to conversational speech, and distinguishes both from the medium of writing, as well as from ACMC. That conversational writing resembles oral conversations to a great degree is due to the related synchronicities of the two; conversational writing and conversational speech are both carried out in real time, which gives rise to a number of linguistic features typical of immediate, interpersonal interaction, while at the same time restricting the number of linguistic traits typically associated with edited asynchronous writing, or the elaborated writing produced for one-way communication (cf. table 1.1). The following lists survey the relationships found in the present study between conversational writing (SCMC as represented by IRC and SSCMC as represented by split-window ICQ chat), writing and speech, at the level of medium (cf. figure 1.2 and the results in chapters 4 and 5 and in the present chapter), i.e. they sum up some of the answers to the first and second research questions.

Compared to writing, conversational writing (SCMC and SSCMC) has

- lower lexical density
- shorter clause length
- shorter word length
- more explicitly involved, interpersonal content, as reflected in e.g. more frequent use of first and second person pronouns, present tense verbs, direct WH-questions and contractions

consideration of various inserts (cf. Biber et al. 1999: 1082ff) is recommended, as well as the consideration of lexical density and emotives (cf. sections 4.3 and 4.6).

- fewer prepositional phrases, reflecting very limited clausal elaboration
- more situation-dependent reference, as reflected in e.g. frequent use of time adverbials and infrequent use of WH-relative constructions
- inherently non-abstract, non-impersonal content (unlike writing), with e.g. few conjuncts and passive constructions
- less informational elaboration by way of e.g. THAT verb complements and THAT relatives in object position
- more analytic than synthetic negation (whereas the opposite obtains in writing)
- paralinguistic features encoded in the script, e.g. graphic imagery, repeated question marks, uppercase words and repeated letters for acoustic effects etc. (rare or absent in writing)
- more exophoric reference to the extra-linguistic context, for instance to the shared virtual room and to web content and files shared
- emotives (unlike writing), signaling the interlocutor's sentiment or the sentiment in which a message is to be received (including ironic or tongue-in-cheek intention)

In addition to the traits above, Internet relay chat (SCMC) in comparison to writing has

- fewer third person pronouns, reflecting hardly any third person reference to participants in the same virtual room other than by way of nicknames

In addition to the traits noted in the 12 bullet points for conversational writing, split-window ICQ (SSCMC) in comparison to writing has

- more possibility and prediction modals, signaling a high degree of involvement and sensitivity to developing relationships
- more predicative adjectives, most of which reflect evaluative and/or supportive discourse content

Compared to speech, conversational writing (SCMC and SSCMC) has

- lexical density similar to that of face-to-face conversations
- slightly shorter clause length
- slightly shorter word length
- similar explicitly involved, interpersonal content as oral conversations, reflected in e.g. similar or more frequent use of first and second person pronouns and present tense verbs
- fewer prepositional phrases, reflecting very limited clausal elaboration

- situation-dependent reference similar to that of oral conversations, reflected in e.g. frequent use of time adverbials and infrequent use of WH-relative constructions
- similar non-abstract, non-impersonal content as in oral conversations, with e.g. few conjuncts and passive constructions
- less real-time informational elaboration by way of e.g. THAT verb complements and THAT relatives in object position
- a similar ratio of analytic to synthetic negation (with more of the former)
- graphemic script (unlike speech), which enables the encoding of paralinguistic features, e.g. graphic imagery, repeated exclamation marks and uppercase words
- nearly similar reference to extra-linguistic content (even though conversational writing lacks options for accompanying the exophoric reference with e.g. glances, pointing and actions in a shared physical space)
- more limited paralinguistic means for expressing emotions, attitudes and sentiments

In addition to the 12 traits above, Internet relay chat (SCMC) in comparison to speech has

- fewer third person pronouns, reflecting hardly any third person reference to participants in the same virtual room other than by way of nicknames

In addition to the traits noted in the 12 bullet points for conversational writing (compared to speech), split-window ICQ (SSCMC) in comparison to speech has

- slightly more possibility and prediction modals, signaling a high degree of involvement and sensitivity to developing relationships
- more predicative adjectives, most of which reflect evaluative and/or supportive discourse content

The third research question, as to how the genres of SCMC, SSCMC and ACMC studied relate to oral conversations on Biber's dimensions, was addressed in the previous chapter and above in the present chapter. The requisite data for addressing the question was illustrated in the dimension graphs of chapter 5 and discussed, and a conclusive analysis of the data was provided in section 6.3. First, the analysis in this chapter (section 6.2) appeared to support the initial hypotheses of the study, whereby the CMC genres were assumed to display a declining degree of orality (i.e. similarity to oral conversations) in the order SSCMC > SCMC > ACMC. Secondly, however, the analysis in this chapter (section 6.3) approached the dimension scores of the conversational genres and ACMC on Biber's (1988) dimensions of linguistic variation from a different perspective. By moving away from

genre boundaries to find similarities in text types across the genres (cf. Biber 1989, 1995), invoking the multidimensional characterization of texts, it was possible to determine, in a complementary way, the relationships between the supersynchronous and synchronous texts, the APMC genre, and oral conversations. The results showed that both conversational writing genres contain texts of the same text types as most oral conversations. Whereas the split-window ICQ texts belong to the text type “intimate interpersonal interaction” (text type 1), the IRC texts are more difficult to analyze but nevertheless belong to text type 2, “informational interaction,” which for IRC entails just as “oral,” but less intimate, interaction.¹²⁰ Collot’s (1991) genre of APMC, however, was deemed by its more peripheral relationship to the multidimensional characterizations of text types 1 and 2, to be more distant than conversational writing from face-to-face and telephone conversations.

The fourth research question, as to whether conversational writing constitutes a modality of its own, brings us back to figure 1.2, illustrating the working relationship between modalities, media and genres/modes in the present study. Modalities are means of production/reception of linguistic content, of which three are regularly recognized in linguistics: speech (language conveyed via acoustic signals), writing (language encoded/decoded in written or typed characters) and sign language (language encoded/decoded in manual and non-manual signs). It was seen in section 1.1, and *passim*, that linguists have recurrently characterized computer-mediated communication as a hybrid variety of communication that occupies the middle ground between the first two modalities (leaving sign language out of the account, as in the present study) or as a variety different from all three. Figure 1.2 illustrates the latter relationship, not assuming *a priori* a hybrid status for conversational writing (SCMC and SSCMC), but nevertheless a tentative status as a fourth modality (APMC being subsumed under the written modality). It is now time to address the question raised in connection with the figure, in the light of insights gained in the study.

Conversational writing has been put to the test repeatedly throughout this study, in chapter 4 relating the genres of SCMC and SSCMC to the media of writing and speech, and in chapter 5 more expressly contrasting the conversational writing genres to the written and spoken genres. Discussions of textual examples, illustrating structural patterns (e.g. a similar low lexical density) and recurring lexico-grammatical features, have consistently found a great similarity

120 In the discussion of the text type of the Internet relay chats, the disjointed character of the discourse was elaborated on and further research into text type 2 was eventually suggested, such that its label might encompass texts of the IRC kind.

between the conversational writing texts and oral conversational texts, and a noticeable dissimilarity between the former and texts of traditional writing. From the text-linguistic point of view, then, conversational writing and oral conversations are strikingly similar. Bear in mind, however, that in the consideration of spoken texts here, the prosodic mark-up was either removed (cf. section 3.4) or largely ignored. Variation in intonation (pitch), stress (loudness), pauses, pace and rhythm, as well as other vocal traits of speech, have thus not been taken into account in the comparisons. In speaking, such vocal traits and, furthermore, gestures, facial expressions, shrugs and glances, as well as, for instance, conventions of body posture, all critically contribute to the encoding of messages with the interlocutor's intention, stance and attitude. In conversational writing, chatters tend to exploit paralinguistic devices (repeated letters, capitals, exclamation marks, emotives etc., the latter arguably linguistic, as seen in section 4.5) to substitute for the lack of acoustic means and facial expressions. However, even though these devices add some expressiveness to messages, they are not capable of carrying all the nuances typically encoded in speech (cf. Crystal 2004b). Conversational writing is rather inherently different from speech, especially from face-to-face conversations, in this respect.

Conversational writing differs from oral conversations in several more respects. For one thing, its production and reception is slower than in speech, owing to the relatively slow pace of typing. For another, written conversations persist for some time on the screen (typically in a scrollable window), or optionally in a log file, whereas spoken conversations are genuinely ephemeral. The textual persistence makes, for instance, response elicitors (e.g. *Okay?*, *Pardon?*, *What?*) rare in conversational writing, even though they are common in conversational speech (as seen in section 4.6). The textual persistence of conversational writing, furthermore, enables interlocutors in IRC to participate in several conversations at once, in one or multiple windows – a peculiarity unparalleled in conversational speech (Crystal 2004b). What is more, IRC messages differ from speech in that they are delivered upon their completion, whereas in speech, and in split-window ICQ chat, messages are decoded while under construction. This makes interlocutors in IRC unable to signal their understanding or puzzlement, or any feedback equivalent to a nod or a backchannel (such as *mhm*), while the conversational partner is composing their turn, even though such signals (response forms; see section 4.6) do appear in IRC upon the interlocutor's receipt of the sender's full turn.

In sum, it is evident that conversational writing is dependent on the encoding and decoding of typed characters and that the set of keys on the keyboard determines what kind of information can be conveyed (just as traditional writing is

confined within the bounds of the alphabet). This kinship to writing, however, is challenged when the synchronicity of the communication is taken into account; for, whereas traditional writing is communicated one way or in asynchronous exchanges, produced in one context and received later in another, conversational writing is communicated in real time, synchronously or supersynchronously, in two-way exchanges (cf. table 1.1). Authors of traditional writing, accordingly, rarely refer exophorically to the extra-linguistic context or situation in which their text is produced. Rather, cohesion in writing must be lexicalized and the state of things made explicit (cf. sections 2.2 and 4.5). In conversational writing, by contrast, exophoric reference is possible in principle (when interlocutors share, for instance, audio files, images, web content and the like). The two factors mentioned, the synchronicity of communication and the degree of shared context (the co-spatiality), in fact, appear to be fundamental for the distinction between writing and conversational writing. To begin to relate the two factors, and their influence on texts, the factors are here combined into a matrix, figure 6.1, in which the x-axis divides the genres studied by their synchronicity of communication and the y-axis determines their position in terms of degree of shared context (“no” shared context, “limited” degree of shared context and “high” degree of shared context). Just as in the figures of chapter 5, the written genres in figure 6.1 are represented by black bullets (for a list of the genres, see Appendix I), spoken genres by gray, and the conversational writing genres by white bullets.

Figure 6.1: Matrix combining the degree of shared context and the synchronicity of communication in the genres studied.

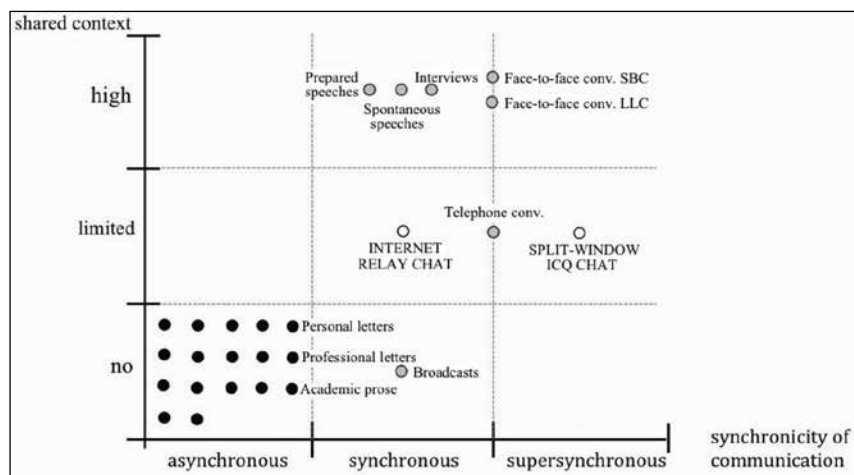


Figure 6.1 clearly illustrates the divide between the written and spoken genres, and the intermediate position of the conversational writing genres. All genres of traditional writing studied reside in the leftmost bottom sector, as they represent asynchronous communication in which the producer and recipient do not share the same context. At the same time, most of the spoken genres (prepared and spontaneous speeches, interviews and face-to-face-conversations) reside in the upper third of the matrix, characterized by a high degree of shared context, except for broadcasts, in which the situation of production and reception are separate contexts, and telephone conversations, in which interlocutors share only the auditory context.¹²¹ The spoken and conversational writing genres are all represented in the synchronous and supersynchronous parts, with face-to-face and telephone conversations symbolically positioned on the dividing line between the synchronicities, as oral conversations may contain limited stretches of complete overlap (cf. sections 1.2 and 1.3).¹²² Internet relay chat is exclusively synchronous communication, whereas split-window ICQ chat admits extensively overlapping, supersynchronous, turns, which assigns the genre a position in the rightmost third of the matrix in figure 6.1, to set it apart from the limited supersynchronicity in oral conversations (cf. table 1.1).

As mentioned, participants in computer-mediated conversational writing share a limited semiotic field. The field is defined by features of the software window, the ongoing interaction and the surrounding information shared on participants' screens (e.g. web content and shared files). Although the chatters' discourse is heavily influenced by the synchronicity of their interaction and chatters' occasional sense of shared context, it is nevertheless restricted to the

121 The genre of broadcasts is extremely diverse and eludes simple classification in figure 6.1, as it may contain, for instance, synchronous texts of all three degrees of shared context (depending on the studio setting and persons involved, etc.). The position opted for to denote the genre here indicates a live broadcast in which the producer of the broadcast discourse and the final recipient are in different locations, as is the case in, for instance, a radio news broadcast. The vast majority of the LLC texts in the genre derive from radio broadcasts (Greenbaum & Svartvik 1990).

122 The "interviews" genre contains interviews, public conversations and debates (Biber 1988, 1995: 87) and may, like conversations, contain limited stretches of overlapping speech, motivating the same position as face-to-face conversation. Its position in the synchronous sector here, however, serves to illustrate that the typical turns produced in the genre are very long (rather monologic) and that the genre contains significantly fewer overlaps than do the face-to-face and telephone conversations.

modality of writing, reduced to the characters on the keyboard. Telephone conversationalists also share a limited context, confined as they are to the auditory context, but their communication shares the richness of the face-to-face genres in that interlocutors are able to use prosody to convey meaning.¹²³ The semiotic richness of face-to-face communication, its high degree of shared context, is not easily recreated in computer chatting. Conversational writing, despite its resemblance to oral conversations, is, after all, still writing.

On the other hand, conversational writing texts differ markedly from traditional writing. For one, traditional writing is bound to static and permanent representations, prototypically on sheets of paper. For another, the asynchronous character of traditional writing enables authors to carefully plan, redraft and edit their texts and to compose elaborate constructions. As there is no recipient simultaneously present (as in conversational writing), there is no pressure on authors to communicate rapidly. Moreover, most traditional writing consists of one-way texts with the character of a monolog, whereas chatted texts by default are dialogs, or even “polylogues” (conversations between ≥ 3 people, Kerbrat-Orecchioni 2004). Despite all these differences, authors of traditional writing and the writers in computer-mediated communication all rely on the same means of representation for the production/reception of language, the *graphemes*, in themselves abstractions of phonemes. All in all, this reliance on the same means of representation makes conversational writing a variety subsumed under the modality of writing, although owing to its resemblance to oral conversations, the variety is pulled a long way in the direction of conversations. Figure 6.2 sums up and illustrates the relationships found in the present study between the modalities writing and speech, their respective media and the genres investigated. The genres of traditional writing are the same 17 genres represented by black bullets in figure 6.1 (as well as in chapter 5; see Appendix I for a list of these), and, again, the spoken genres are represented by gray and the conversational writing genres by white bullets.

123 Neither of the axes in figure 6.1 represents a continuum; rather, the genres are represented in sectors to which they conceptually belong (cf. table 1.1). If the vertical axis were a continuum, telephone conversations might be positioned above conversational writing, but this would entail that the matrix indicates variable degrees of shared context in the traditional writing genres as well, which it is not intended to do.

Figure 6.2: Relationships found between modalities, media and the genres investigated.

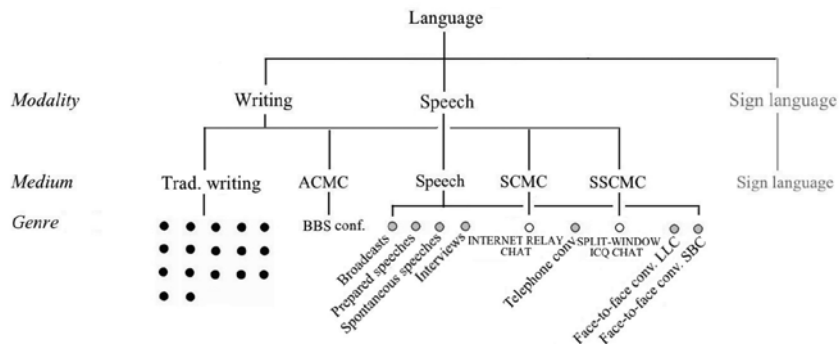


Figure 6.2 thus illustrates the answer to the fourth research question, as to whether conversational writing (SCMC and SSCMC) constitutes a modality of its own (raised in section 1.2 in connection with figure 1.2). The findings in the present study have not evidenced that conversational writing is a new modality. Rather, the qualitative assessment of conversational writing in the present section, based on the combined results, indicates that the genres of conversational writing share the same modality as traditional writing and ACMC (the modality of writing), while at the same time being extreme offshoots of traditional writing. In fact, conversational writing is so spoken-like that its genres functionally, structurally and lexico-grammatically most fittingly are represented among the oral conversational genres in figure 6.2. The interspersing of the conversational writing genres among the spoken genres is justified, among other things, by the adherence of the conversational writing texts to Biber's (1989, 1995) text types 1 and 2 (cf. section 6.3). The discussion in section 6.2, and the combined in-depth analyses of the genres' positions on Biber's (1988) dimensions in chapter 5, also lead to the same conclusion – Internet relay chat and split-window ICQ texts are about as “oral” as oral conversations, despite being conveyed via graphemes. Figure 6.2 thus expressly evolves the hybrid character categorizations of CMC made in previous studies (cf. section 1.1) by illustrating how the written and spoken genres intertwine in linguistic space. The figure, of course, is highly synoptic; the relative positions of the conversational writing genres – indeed all genres – are most accurately specified in a continuum of multiple dimensions, those explored in chapter 5 and above.

A number of authors and linguistic scholars who pioneered the exploration of computer-mediated communication asserted that, while displaying some similarities with traditional notions of spoken and written discourse, the linguistic

character of online interaction is something entirely new and unique (e.g. Ferrara et al. 1991, Davis & Brewer 1997, Crystal 2001). Crystal (2001) went so far as to propound that linguistic accounts after the advent of the Internet must comprise four discourse varieties (cf. modalities): written language, spoken language, sign language and now computer-mediated language (as mentioned in section 1.2). Later, investigators attempted to defuse the initial insinuations of a fourth modality (or “fourth medium,” in Crystal’s 2001: 238 terms) by reverting to explanations of genre/activity variation, pointing to written character aspects and users’ creative adaptation of written language to the confined semiotic field(s) of CMC (e.g. Hård af Segerstad 2002). Regardless of their perspectives, most scholars agree that the electronic media “facilitate and constrain our ability to communicate in ways that are fundamentally different from those found in other semiotic situations” (Crystal 2004b: 68). As the novelty of the CMC media is wearing off, however, the adduction of explanations relating to genre/activity variation appears increasingly level-headed and progressive. Genres of speech, by all means, range from the most written-like genres, such as news broadcasts or prepared speeches (cued by written props or manuscripts), to intimate face-to-face conversations. Similarly, written genres, as seen here, range from the most prototypical, information-dense, elaborated pieces of text to the most oral-like pieces of conversational writing. At the same time, there is no simple dichotomy of writing vs. speech at the levels of medium and genre (see figure 6.2); rather, written and spoken genres intertwine. As early as 1988, Biber demonstrated this interspersing of written and spoken genres in linguistic space on his dimensions of linguistic variation (Biber 1988). The present study has only extended the range of written genres on the same dimensions and added to the complexity of linguistic variation in the English language. In future accounts of the complete textual variation in a language, conversational writing texts, with all their peculiarities, cannot be ignored.

The present section has revisited, discussed and summarized some of the answers to the four research questions addressed in this study. Before concluding the section, the promised review of the definition of conversational writing will be tended to. At the beginning of this section, the tentative definition (from section 1.1) was reiterated, by which conversational writing is written communication 1) for social interaction 2) which requires the simultaneous presence (physical or virtual) of producer and recipient, 3) in which interlocutors expect immediate feedback (i.e. within seconds) and 4) during which the discourse may be reconfigured by the participants while under construction (e.g. as interlocutors are able to influence each other’s line of thought). The discussion in this

section was not intended to address the definition per se, nor did it give reason to question the four criteria for identifying conversational writing. The discussion, nevertheless, prompted a consideration of the situational features of the individual conversational writing genres, parameters in light of which conversational writing genres may be seen to differ from each other. In the consideration, at the beginning of the section, it was found that computer-mediated conversational writing, besides points 1–4, is also written communication in which participants adopt oral linguistic strategies that reflect the semiotic field (e.g. participants' degree of shared context, potential reference to web content and shared files), the semiotic tenor (i.e. the parameters of participants' social relationship, previous acquaintance and future off-line relations) and the semiotic mode (e.g. the configuration of the conversational writing client, its options for selecting the number and identity of participants, public and private space etc.) of their interaction. Conversational writing genres may be identified on the basis of the four-fold definition and described, and possibly grouped, by means of the parameters (the parameters are found in the parentheses here; the semiotic elements field, tenor and mode, of course, are at play in any communicative act). In the present study, the conversational writing genres have both essentially been found to be lexico-grammatically correspondent to oral conversations. Functionally, the dialogic configuration of split-window ICQ chat serves better for participants to further personal, real-life relationships, whereas the extremely fragmented, polylogic structure of IRC, rarely found in typical face-to-face and telephone conversations, invites more cursory acquaintance. Structural differences notwithstanding, both chat genres are inherently conversational; just as oral conversations, they involve real-time communication between interlocutors sharing features of the same situation and the ability to immediately affect each other's contributions to the discourse.

It should be borne in mind that the four-fold definition of conversational writing indeed draws on the properties of the two conversational writing genres studied here, but equally well applies to, for instance, web chat and recent IM applications such as Facebook chat (when used for SCMC). A close functional, structural and/or lexico-grammatical examination of more genres (for instance, textual conversations in virtual worlds, with a higher degree of shared context, or the occasionally two-way synchronous communication in SMS, e-mail and Twitter; cf. table 1.1) is likely to occasion refinements of the definition. The definition offered, all the same, may serve as a starting point for the identification of conversational writing in existing and emergent modes of CMC and telecommunications. Conversational writing is likely to stay relevant well into the future, but

also to evolve along unforeseeable paths. In this constant state of flux, the field of CMC linguistics rarely allows long-standing definitions, but continues to afford ample opportunity for scholars to explore emergent modes against the backdrop of previously described communications.

6.5 Chapter summary

The purpose of the present chapter has been to synthesize and discuss the results of the full investigation. After a few introductory remarks, the chapter opened with the consideration of quantitative results pertaining to the two hypotheses underlying the study. By relating the positions of the conversational writing genres and the genre of ACMC to the oral conversational genres on Biber's (1988) dimensions, it was possible to begin the assessment of the orality in conversational writing and asynchronous CMC. Initially, both hypotheses appeared to be supported; more precisely, the quantitative findings indicated that the highest degree of orality (i.e. lexico-grammatical similarity to oral conversations) was observed in supersynchronous conversational writing, followed by that in synchronous conversational writing, and showed that the asynchronous CMC genre was the least oral of the three CMC genres investigated. The chapter proceeded, however, via a close examination of the overall picture afforded by all dimensions, to assess the multidimensional character of the three CMC genres to identify the genres' most prevalent text types (Biber 1989, 1995), in order to address the hypotheses further. The texts of the conversational writing genres studied, like most oral conversational texts in Biber's studies, were found to belong to text types 1, "intimate interpersonal interaction," and 2, "informational interaction," whereas Collo's (1991) genre of BBS conferencing adhered to the latter, but more for its similarity to, for instance, personal letters than to conversations. The quantitative and qualitative assessments in combination thus supported the first and rejected the second hypothesis; that is, in short, the study has found conversational writing to be more "oral" than asynchronous CMC, but the SCMC genre to be no less "oral" than the SSCMC genre. Next, the chapter revisited the research questions posed at the beginning of the study, and addressed throughout, to find and synthesize the answers provided to these. By way of a semiotic analysis of the communication in the material studied, it was possible, among other things, to relate the higher degree of orality initially found in split-window ICQ than in IRC to the more similar semiotic tenor and mode of the former and oral conversations, rather than to any supersynchronicity effect. Although no supersynchronicity effect was evidenced in the material, a substantial synchronicity effect was found, which likens the conversational writing texts, of both chat genres, to

oral conversations. A selective bullet-point overview of the relationships found between conversational writing, writing and speech was then provided, before the last research question, as to whether conversational writing constitutes a modality of its own, was addressed and answered. In short, conversational writing was found to rely on the modality of writing but to convey discourse most closely akin to oral conversations. Finally, the definition of conversational writing was revisited in light of the findings. The final chapter, below, offers a concluding summary of the full study and some suggestions for further research.

Chapter 7. Conclusion

7.1 Summary of the study

Since Biber's (1988) comprehensive multidimensional analysis of variation in spoken and written English, a wealth of linguistic multidimensional studies have been carried out by a great number of scholars (cf. section 2.2). Some have applied the approach to a single genre or a few genres, with a synchronic or a diachronic perspective, to investigate patterns in, for instance, child and adult language, variation in interdisciplinary texts and historical shifts in the language of women and men. Others have analyzed patterns of variation in languages other than English, and a few have made cross-linguistic comparisons of variation in several languages. Yet, to the present author's knowledge, prior to this study, no multidimensional linguistic analysis of English computer-mediated conversational writing has been presented. The investigation at hand has attempted to fill this gap.

A conversational writing corpus, UCOW, introduced in section 1.2 and described in sections 3.1–3.3, was compiled and annotated for this multidimensional analysis, as was a subset of the Santa Barbara Corpus (SBC) of face-to-face conversations. The foremost aim of the analysis was to situate the two UCOW corpus components, Internet relay chat and split-window ICQ chat, and incidentally also the SBC subset genre, on Biber's (1988) dimensions of variation among the previously positioned genres of writing and speech (from LOB and LLC; see Biber 1988). The investigation was motivated, among other things, by an anticipated similarity between conversational writing and conversational speech (as the status of conversational writing proposed in previous research was that of a hybrid between speech and writing) and a desire to elucidate the relationship between conversational writing and traditional writing. Chapter 1 set the stage for the study by presenting the hypotheses to be tested and the research questions to be answered; chapter 2 provided a background of previous research into writing, speech and CMC and presented the interfaces of the synchronous (IRC) and supersynchronous (ICQ) chat clients; and chapter 3 described the material to be investigated and the methodology for obtaining the data required (inter alia, the frequencies of the 67 linguistic features) for the genres to be contrasted with the genres of writing and speech.

In chapters 4 and 5, the results of the empirical investigation were presented. The primary purpose of chapter 4 was to exemplify and discuss the distribution of the salient linguistic features in conversational writing, those identified to be

most frequent compared to the mean of Biber's (1988) written and spoken genres, as well as previously understudied features, such as lexical density, inserts and emotives. Whenever possible, functional comparisons of feature distributions were made across the genres and media, in order to relate conversational writing to writing and speech, as well as to Collo's (1991) genre of APMC. The results consistently showed a similar distribution of features in conversational writing and conversational speech, and a different distribution in writing, recurrently also in APMC. In chapter 5, the dimension scores of the conversational writing genres were plotted on Biber's (1988) dimensions, revealing, on most dimensions, positions in the vicinity of oral conversations. On Dimension 1, conversational writing and conversational speech both display involved, interactive and occasionally affective discourse. A majority of the most salient linguistic features identified in conversational writing in chapter 4 (e.g. first and second person pronouns, direct WH-questions, analytic negation, demonstrative pronouns and present tense verbs) contribute to this similarity. On Dimension 2, conversational writing is slightly less narrative than conversational speech, although neither is particularly concerned with narration. Dimension 3 evidences, for conversational writing and conversational speech alike, a discourse with abundant situation-dependent reference, Dimension 4 generally little overt expression of persuasion/argumentation in either, and Dimension 5, for both, a discourse with typically non-abstract/non-impersonal information. Dimension 6 indicates a slightly more complex relationship between conversational writing and the spoken and written genres.

In chapter 6, the ample and multifaceted results were brought together and discussed. The primary purposes of the chapter were to interrelate the results in order to test the hypotheses of the study, quantitatively and qualitatively, and to sum up the answers provided to the research questions. Via statistical calculations, relating the conversational writing genres to the oral conversational genres on Biber's (1988) dimensions, the degree of orality in the former was assessed quantitatively (i.e. the genres' proximity to oral conversations), initially showing a slightly higher degree for split-window ICQ than for IRC. Collo's (1991) genre of APMC was also assessed and was observed to be the least oral of the three CMC genres. The chapter then analyzed and discussed the multidimensional character of the three CMC genres, to identify their most prevalent text types (Biber 1989, 1995) in order to assess the degrees of orality qualitatively. In brief, the combined quantitative and qualitative assessments evidenced no higher degree of orality in split-window ICQ chat than in IRC, but a higher degree of orality in both conversational writing genres than in the APMC genre. The orality in

conversational writing was then examined by way of a semiotic analysis, which found the “higher” degree of orality initially observed in split-window ICQ to be due to a semiotic structure more similar to that in intimate, oral conversations. The similar semiotic structure of split-window ICQ and such conversations was seen as a more decisive contributor to the greater lexico-grammatical correspondence between split-window ICQ and oral conversations than was its supersynchronicity.

More qualitative, contrastive discussions of conversational writing and the modalities of writing and speech followed upon these findings. Traits that liken conversational writing to writing were brought to light (such as the persistence of the graphemic script in both) as well as traits that distinguish conversational writing from conversational speech (such as the inability in the former to convey a particular tone of voice). The degree of shared context and the synchronicity of communication in the media were also contrasted and discussed in order to determine the potential status of conversational writing as a modality of its own, alongside the modalities of writing and speech. The discussion yielded no support for the formulation of a new modality; rather, conversational writing may be regarded as the most oral-like form of writing, just as, for instance, broadcast and prepared speeches, cued by props or manuscripts, may be regarded as some of the most written-like forms of speech. Genres of writing and speech simply intersperse in linguistic space. The most accurate and fine-grained representation of the relationship between genres across writing and speech, moreover, is a multidimensional one, as illustrated in Biber (1988) and, for the conversational writing genres, in chapter 5 here.

In an extensive special issue on “computer-mediated conversation” in the online scholarly journal *Language@Internet* (volumes 7 and 8), a collection of original research articles from several disciplines (including conversation analysis, interactional sociolinguistics and pragmatics, spanning more than a decade of research) is presented, articles that all contribute significantly to the field of CMC linguistics, exploring the conversationality in various modes of CMC. Introducing the collection, Herring (2011b) acknowledges the unique contributions of the articles, but notes that only a few of them directly assess the relative degrees of conversationality across different modes and that “no single set of methods is employed, or questions asked, across the collection that would make the results of the individual studies directly comparable with one another” (2011b: 7; as seen in section 1.1 here). Herring proceeds to suggest additional studies, especially systematic comparisons of several modes using “a common set of methods” (2011b:

7) and incidentally calls for studies that compare “CMC with spoken and/or written genres (cf. Collot & Belmore, 1996; Ko, 1996; Yates, 1996)” (ibid.).

The present study has responded to Herring’s (2011b) call for research by providing a description of prototypical conversational writing, a description methodologically comparable to Collot’s (1991) and Collot & Belmore’s (1996) description of APMC, partially comparable to Yates’ (1993, 1996) study of APMC, and complementary to e.g. Ko’s (1994, 1996) and Freiermuth’s (2003) studies of SCMC. Biber’s (1988) dimensions have here enabled the systematic comparison of two conversational writing genres, and BBS conferencing (the latter from Collot 1991), relative to a range of written and spoken genres and particularly provided a methodology to elucidate the CMC genres’ relative degrees of conversationality (here called orality). Via multidimensional characterizations, among other things, it was possible to explore just what it means for the chatted texts to be conversational, i.e. to assess their similarity to oral conversations. In sum, just as there was a gap in variationist linguistics as regards a description of synchronous and super-synchronous conversational writing genres, there was a gap in CMC linguistics as regards a systematic variationist analysis of the same genres. While filling the first gap, the present study also incidentally filled the second, an effort which, taken as a whole, might be regarded as the major contribution of the study.

Biber’s (1988) dimensions, for a time, may constitute the gauge for lexico-grammatical descriptions of computer-mediated conversational writing genres and other CMC genres, although eventually, of course, the currency and universality of Biber’s (1988) genres (from LOB and LLC) and features may be called into question, motivating a new comprehensive multifeature/multidimensional analysis of the English language. Until then, Biber’s (1988) approach, as employed here, is one of the more rigorous ways to systematically compare existing and emergent genres of CMC. For future purposes, the present study has underscored the importance of including conversational writing genres, and other genres of CMC, in any analysis of the full variation in the English language, and suggested the consideration of lexical density, inserts and emotives in such analyses.

The conversational writing carried out in split-window ICQ chat is arguably the most intimate, “oral” (or, in Herring’s 2011b terms, “conversational”) form of writing ever documented. In fact, the corpus of split-window ICQ chat presented in this study is believed to document a unique stage in the history of English, in which written texts, functionally and lexico-grammatically, were closer than they ever have been to extremely involved, oral conversational texts. In conclusion, the linguistic documentation of this distinctive genre is another significant contribution of this study, to the fields of variation and CMC studies alike.

7.2 Suggestions for further research

In the decade that has passed since the recording of UCOW, CMC has evolved in several directions. Computer-mediated conversational writing, for instance, appears to have developed along at least three *prima facie* discernible trajectories, those involving privatization (i.e. a popular move from public chat channels, such as in IRC, to the private conversations in IM, such as Facebook chat); desynchronization (by which supersynchronous modes have become synchronous, such as present-day ICQ, and by which synchronous modes are gradually supplanted by asynchronous ones, which enable users to receive messages at a time of their convenience); and specialization/topicalization (which makes public chat increasingly used for particular events, such as web chat with public officials after their televised appearance, or for public or commercial services, such as library and travel agent chat services). The latter developments suggest that synchronous conversational writing today is found in a range of contexts, both private and public, that may have given rise to several genres of conversational texts, or possibly to sub-genres. To explore and contrast the linguistic properties of current and emerging genres/sub-genres of conversational writing would be an intriguing area of research.

At the same time, a few asynchronous modes of CMC and telecommunications are increasingly used for two-way synchronous communication. Mobile texting, for instance, is occasionally used for interaction resembling conversational writing as defined here. Software seamlessly incorporated into text messaging functions in mobiles, such as iMessage (really an IM service with push technology), indicates in the message window (in iMessage by three dots) that a user is keying in a message, making users aware of each other's simultaneous participation in the communication. It would be interesting to investigate whether, and if so, how, the language in the synchronous exchanges of such communication differs from that in the asynchronous exchanges between the same users, whether the synchronous sequences, for instance, give rise to more backchannels.

The present study has far from exhausted the topic of conversational writing. Rather, in the flux of developments, CMC and telecommunications continue to give rise to ample reconfigurations of linguistic material, to texts that may be explored from the variationist's and the CMC scholar's perspectives alike. Emerging texts need to be closely surveyed and analyzed in order for linguists to effectively contribute to the collaborative scholarly effort of elucidating the workings of human interaction. It will be fascinating to continue the pursuit.

Appendices

Appendix I. Texts used in Biber's (1988) study

Texts used in Biber's (1988) study of language variation. The corpus totals approximately 960,000 words for 481 texts (Biber 1988: 67, 209–210, 1995: 87).

	Genre	Texts used	Number of texts	Approx. number of words
Speech	Face-to-face conv.	texts 1.1–1.14 and 3.1–3.6 from LLC	44	115,000
	Telephone conv.	texts 7.1–7.3, 8.1–8.4 and 9.1–9.3 from LLC	27	32,000
	Interviews ¹	texts 5.1–5.3, 5.5–5.7, 6.1, 6.3, 6.4a, 6.5 and 6.6 from LLC	22	48,000
	Broadcasts	texts 10.1–10.7 and part of text 10.8 from LLC	18	38,000
	Spont. speeches ²	texts 11.1–11.5 from LLC	16	26,000
	Prepared speeches	texts 12.1–12.6 from LLC	14	31,000
Writing	Press reportage	all texts in LOB category A	44	88,000
	Press editorials	all texts in LOB category B	27	54,000
	Press reviews	all texts in LOB category C	17	34,000
	Religion	all texts in LOB category D	17	34,000
	Hobbies	the first 30,000 words (texts 1–14) LOB category E	14	30,000
	Popular lore	the first 30,000 words (texts 1–14) LOB category F	14	30,000
	Biographies	the first 30,000 words (texts 1–14) LOB category G	14	30,000
	Official documents	texts 1–6, 13–14 and 25–30 from LOB category H	14	28,000
	Academic prose	all texts in LOB category J	80	160,000
	General fiction	all texts in LOB category K	29	58,000
	Mystery fiction	the first 30,000 words (texts 1–14) LOB category L	13	26,000

124 "Interviews" denotes public conversations, debates and interviews (Biber 1988, 1995: 87).

125 In his description of the sampling procedure, Biber (1988: 210) indicates that spontaneous speeches were divided into 15 texts, which would yield a total of 480 texts. Later accounts, however, maintain that there was a total of 481 texts (e.g. Biber 1995: 87, Conrad & Biber 2001: 111), which explains why the figure from Biber (1988: 67) is retained here.

Genre	Texts used	Number of texts	Approx. number of words
Science fiction	all texts in LOB category M	6	12,000
Adventure fiction	the first 30,000 words (texts 1–14) LOB category N	13	26,000
Romantic fiction	the first 30,000 words (texts 1–14) LOB category P	13	26,000
Humor	all texts in LOB category R	9	18,000
Personal letters	written to friends/relatives, collected by D. Biber	6	6,000
Professional letters	on administrative matters, collected by W. Grabe	10	10,000
total		481	960,000

Appendix II. Descriptive statistics for genres studied

The frequencies in tables 1–7 are all normalized to text lengths of 1,000 tokens (except for type/token-ratio and word length); see section 3.2.

Table 1: *Descriptive statistics for Internet relay chat*

	Linguistic feature	Mean	Min. value	Max. value	Range	Std. deviation
1	past tense verbs	12.0	3.0	19.3	16.3	6.8
2	perfect aspect verbs	2.5	1.0	6.1	5.1	1.5
3	present tense verbs	147.2	135.4	170.0	34.7	11.4
4	place adverbials	2.5	1.0	4.1	3.1	1.2
5	time adverbials	8.4	4.0	13.0	9.0	3.2
6	first person pronouns	56.9	31.3	79.6	48.3	13.3
7	second person pronouns	50.4	26.3	70.1	43.8	13.4
8	third person pronouns	10.3	0.0	15.3	15.3	4.9
9	pronoun IT	12.3	6.2	17.1	10.9	3.1
10	demonstrative pronouns	6.6	3.0	11.2	8.2	3.0
11	indefinite pronouns	11.7	2.0	22.2	20.2	5.5
12	DO as pro-verb	4.2	0.0	10.4	10.4	3.2
13	direct WH-questions	3.5	1.0	8.3	7.3	2.9
14	nominalizations	4.1	0.0	12.1	12.1	4.3
15	gerunds	1.2	0.0	4.0	4.0	1.3
16	nouns	144.5	105.1	176.8	71.7	27.7
17	agentless passives	1.8	0.0	4.1	4.1	1.7
18	BY passives	0.2	0.0	2.0	2.0	0.6
19	BE as main verb	22.4	17.0	31.6	14.6	4.7
20	existential THERE	1.2	0.0	3.1	3.1	1.0
21	THAT verb complements	0.3	0.0	2.0	2.0	0.7
22	THAT adj. complements	0.0	0.0	0.0	0.0	0.0
23	WH clauses	2.0	0.0	8.1	8.1	2.4
24	infinitives	12.1	6.0	20.2	14.2	5.5
25	present participial clauses	0.0	0.0	0.0	0.0	0.0
26	past participial clauses	0.0	0.0	0.0	0.0	0.0
27	past prt. WHIZ deletions	0.2	0.0	1.0	1.0	0.4
28	present prt. WHIZ deletions	0.0	0.0	0.0	0.0	0.0
29	THAT relatives: subj. position	0.0	0.0	0.0	0.0	0.0
30	THAT relatives: obj. position	0.4	0.0	3.0	3.0	1.0
31	WH relatives: subj. position	0.3	0.0	3.0	3.0	1.0

Linguistic feature	Mean	Min. value	Max. value	Range	Std. deviation
32 WH relatives: obj. position	0.0	0.0	0.0	0.0	0.0
33 WH relatives: pied pipes	0.0	0.0	0.0	0.0	0.0
34 sentence relatives	0.1	0.0	1.0	1.0	0.3
35 adv. subordinator – cause	0.4	0.0	2.0	2.0	0.7
36 adv. sub. – concession	0.7	0.0	2.1	2.1	0.7
37 adv. sub. – condition	2.1	0.0	5.0	5.0	1.5
38 adv. sub. – other	0.4	0.0	2.0	2.0	0.7
39 prepositional phrases	47.0	33.3	65.5	32.2	10.1
40 attributive adjectives	49.8	30.0	62.6	32.6	11.0
41 predicative adjectives	8.4	5.0	13.0	8.0	2.9
42 adverbs	79.9	51.7	109.1	57.4	17.0
43 type/token ratio	54.9	48.4	60.2	11.9	4.4
44 word length	4.0	3.8	4.4	0.6	0.2
45 conjuncts	0.0	0.0	0.0	0.0	0.0
46 downtoners	1.7	0.0	4.1	4.1	1.5
47 hedges	0.6	0.0	2.1	2.1	0.7
48 amplifiers	2.1	0.0	4.2	4.2	1.4
49 emphatics	7.8	3.1	14.4	11.2	3.6
50 discourse particles	3.3	1.0	6.1	5.1	1.9
51 demonstratives	3.4	1.0	6.1	5.1	1.5
52 possibility modals	6.3	2.1	15.2	13.1	3.5
53 necessity modals	1.8	0.0	4.1	4.1	1.6
54 prediction modals	6.1	1.0	14.1	13.1	3.9
55 public verbs	2.1	1.0	4.2	3.2	1.1
56 private verbs	17.9	6.2	25.0	18.8	6.3
57 suasive verbs	1.2	0.0	2.1	2.1	0.8
58 SEEM/APPEAR	0.1	0.0	1.0	1.0	0.3
59 contractions	30.8	19.5	40.0	20.6	7.5
60 THAT deletion	3.3	0.0	7.1	7.1	2.3
61 stranded prepositions	3.1	0.0	7.3	7.3	2.0
62 split infinitives	0.0	0.0	0.0	0.0	0.0
63 split auxiliaries	2.0	0.0	5.1	5.1	1.7
64 phrasal coordination	4.3	0.0	11.2	11.2	3.4
65 non-phrasal coordination	2.7	0.0	6.2	6.2	2.2
66 synthetic negation	2.3	0.0	6.0	6.0	2.2
67 analytic negation	13.1	3.1	24.0	20.9	6.0

Table 2: Descriptive statistics for split-window ICQ chat

	Linguistic feature	Mean	Min. value	Max. value	Range	Std. deviation
1	past tense verbs	34.4	20.0	65.4	45.4	14.5
2	perfect aspect verbs	2.7	0.0	7.0	7.0	2.2
3	present tense verbs	168.5	130.9	204.0	73.1	25.7
4	place adverbials	2.6	0.0	9.5	9.5	2.7
5	time adverbials	5.0	1.9	8.7	6.8	1.8
6	first person pronouns	88.9	73.9	111.4	37.5	12.2
7	second person pronouns	45.0	29.6	77.6	48.0	13.5
8	third person pronouns	23.6	0.0	38.1	38.1	12.2
9	pronoun IT	19.9	9.2	28.0	18.8	6.1
10	demonstrative pronouns	16.4	9.2	27.3	18.1	5.2
11	indefinite pronouns	6.0	0.0	13.8	13.8	4.3
12	DO as pro-verb	8.2	0.0	14.5	14.5	3.8
13	direct WH-questions	3.9	1.5	8.7	7.2	2.3
14	nominalizations	3.6	0.0	10.7	10.7	3.8
15	gerunds	0.4	0.0	3.5	3.5	1.1
16	nouns	135.1	99.2	195.7	96.5	32.0
17	agentless passives	1.1	0.0	2.9	2.9	1.0
18	BY passives	0.1	0.0	1.6	1.6	0.5
19	BE as main verb	24.7	17.4	40.0	22.6	7.3
20	existential THERE	0.5	0.0	2.1	2.1	0.7
21	THAT verb complements	1.5	0.0	4.2	4.2	1.3
22	THAT adj. complements	0.2	0.0	1.0	1.0	0.4
23	WH clauses	2.6	0.0	8.7	8.7	2.7
24	infinitives	11.9	4.6	21.8	17.2	5.9
25	present participial clauses	0.0	0.0	0.0	0.0	0.0
26	past participial clauses	0.1	0.0	1.2	1.2	0.3
27	past prt. WHIZ deletions	0.2	0.0	2.2	2.2	0.6
28	present prt. WHIZ deletions	0.4	0.0	2.6	2.6	0.9
29	THAT relatives: subj. position	0.1	0.0	0.9	0.9	0.3
30	THAT relatives: obj. position	0.4	0.0	2.1	2.1	0.7
31	WH relatives: subj. position	0.0	0.0	0.0	0.0	0.0
32	WH relatives: obj. position	0.1	0.0	0.9	0.9	0.3
33	WH relatives: pied pipes	0.0	0.0	0.0	0.0	0.0
34	sentence relatives	0.4	0.0	1.7	1.7	0.6
35	adv. subordinator – cause	4.3	0.0	11.4	11.4	3.5
36	adv. sub. – concession	1.3	0.0	3.2	3.2	1.1

	Linguistic feature	Mean	Min. value	Max. value	Range	Std. deviation
37	adv. sub. – condition	4.4	0.0	7.0	7.0	1.9
38	adv. sub. – other	0.8	0.0	2.9	2.9	1.1
39	prepositional phrases	42.0	20.0	53.7	33.7	9.7
40	attributive adjectives	30.4	22.6	49.5	26.9	8.1
41	predicative adjectives	15.3	9.7	21.8	12.1	4.3
42	adverbs	71.5	51.8	90.1	38.4	12.9
43	type/token ratio	52.0	47.0	60.0	13.0	4.1
44	word length	3.7	3.4	4.0	0.6	0.1
45	conjuncts	0.3	0.0	2.1	2.1	0.7
46	downtoners	0.5	0.0	1.7	1.7	0.7
47	hedges	2.5	0.0	7.0	7.0	2.6
48	amplifiers	1.5	0.0	6.5	6.5	2.0
49	emphatics	13.1	8.3	19.1	10.8	4.1
50	discourse particles	4.9	0.0	10.6	10.6	4.0
51	demonstratives	7.2	3.2	16.3	13.0	3.5
52	possibility modals	9.2	4.8	15.3	10.4	2.7
53	necessity modals	2.0	0.0	6.2	6.2	2.3
54	prediction modals	9.3	1.0	17.4	16.4	4.9
55	public verbs	5.2	0.0	10.6	10.6	3.1
56	private verbs	30.6	16.8	43.8	26.9	7.9
57	suasive verbs	1.5	0.0	6.1	6.1	1.9
58	SEEM/APPEAR	0.4	0.0	2.1	2.1	0.7
59	contractions	55.0	25.8	82.6	56.8	16.5
60	THAT deletion	9.9	3.5	21.2	17.7	4.9
61	stranded prepositions	2.6	0.0	6.2	6.2	1.9
62	split infinitives	0.1	0.0	1.6	1.6	0.5
63	split auxiliaries	4.2	1.5	8.3	6.8	2.3
64	phrasal coordination	2.3	0.0	4.6	4.6	1.5
65	non-phrasal coordination	6.1	0.0	19.1	19.1	5.1
66	synthetic negation	2.5	0.0	4.8	4.8	1.8
67	analytic negation	29.7	20.0	40.3	20.3	7.6

Table 3: Descriptive statistics for the SBC subset (spoken American English)

	Linguistic feature	Mean	Min. value	Max. value	Range	Std. deviation
1	past tense	36.0	1.4	67.3	65.9	21.3
2	perfect aspect verbs	5.7	0.0	14.0	14.0	5.1
3	present tense	141.6	90.9	197.5	106.6	28.2
4	place adverbials	1.4	0.0	2.8	2.8	1.3
5	time adverbials	3.3	0.0	11.2	11.2	2.7
6	first person pronouns	61.0	28.0	108.0	80.0	20.5
7	second person pronouns	36.0	7.0	75.3	68.3	16.4
8	third person pronouns	40.9	6.9	84.0	77.1	24.0
9	pronoun IT	27.0	7.0	43.1	36.0	9.8
10	demonstrative pronouns	16.0	7.0	30.8	23.8	6.5
11	indefinite pronouns	6.6	1.4	12.7	11.3	3.9
12	DO as pro-verb	6.1	2.8	18.1	15.3	4.6
13	direct WH-questions	2.7	0.0	8.3	8.3	2.3
14	nominalizations	7.4	0.0	43.5	43.5	10.9
15	gerunds	0.5	0.0	1.4	1.4	0.7
16	nouns	135.6	78.5	209.0	130.4	33.6
17	agentless passives	2.6	0.0	7.0	7.0	2.0
18	BY passives	0.0	0.0	0.0	0.0	0.0
19	BE as main verb	19.2	4.2	32.2	28.0	9.5
20	existential THERE	2.8	0.0	5.6	5.6	1.9
21	THAT verb complements	1.9	0.0	15.4	15.4	4.0
22	THAT adj. complements	0.2	0.0	1.4	1.4	0.5
23	WH clauses	2.3	0.0	5.6	5.6	1.5
24	infinitives	8.8	4.2	16.8	12.6	3.4
25	present participial clauses	0.1	0.0	1.4	1.4	0.4
26	past participial clauses	0.1	0.0	1.4	1.4	0.4
27	past prt. WHIZ deletions	0.0	0.0	0.0	0.0	0.0
28	present prt. WHIZ deletions	0.1	0.0	1.4	1.4	0.4
29	THAT relatives: subj. position	0.5	0.0	2.8	2.8	0.9
30	THAT relatives: obj. position	1.4	0.0	5.6	5.6	2.0
31	WH relatives: subj. position	0.0	0.0	0.0	0.0	0.0
32	WH relatives: obj. position	0.0	0.0	0.0	0.0	0.0
33	WH relatives: pied pipes	0.1	0.0	1.4	1.4	0.4
34	sentence relatives	0.6	0.0	2.8	2.8	0.9
35	adv. subordinator – cause	4.3	0.0	8.4	8.4	2.5
36	adv. sub. – concession	0.3	0.0	2.8	2.8	0.8

	Linguistic feature	Mean	Min. value	Max. value	Range	Std. deviation
37	adv. sub. – condition	4.1	0.0	12.6	12.6	4.1
38	adv. sub. – other	0.7	0.0	2.8	2.8	0.9
39	prepositional phrases	61.1	43.4	82.5	39.1	12.5
40	attributive adjectives	33.7	9.8	54.7	44.9	14.3
41	predicative adjectives	8.2	2.8	18.1	15.3	5.0
42	adverbs	68.0	42.0	93.8	51.9	13.9
43	type/token ratio	44.2	34.0	50.5	16.5	4.7
44	word length	4.0	3.6	4.6	1.0	0.3
45	conjuncts	0.3	0.0	2.8	2.8	0.8
46	downtoners	1.3	0.0	7.0	7.0	2.0
47	hedges	2.3	0.0	11.2	11.2	3.4
48	amplifiers	1.7	0.0	8.4	8.4	2.3
49	emphatics	12.7	2.8	21.0	18.2	5.4
50	discourse particles	7.7	0.0	16.8	16.8	4.7
51	demonstratives	9.5	5.6	12.6	7.0	2.4
52	possibility modals	8.0	1.4	16.8	15.4	4.0
53	necessity modals	1.2	0.0	4.2	4.2	1.6
54	prediction modals	6.9	0.0	16.8	16.8	4.6
55	public verbs	6.5	0.0	23.8	23.8	6.1
56	private verbs	33.6	11.2	65.9	54.7	14.8
57	suasive verbs	1.8	0.0	7.0	7.0	2.2
58	SEEM/APPEAR	0.1	0.0	1.4	1.4	0.4
59	contractions	48.5	33.5	89.0	55.5	16.8
60	THAT deletion	7.1	1.4	15.3	13.9	4.3
61	stranded prepositions	3.4	0.0	7.0	7.0	2.2
62	split infinitives	0.0	0.0	0.0	0.0	0.0
63	split auxiliaries	4.2	0.0	8.4	8.4	3.1
64	phrasal coordination	3.4	0.0	8.4	8.4	2.8
65	non-phrasal coordination	16.7	1.4	46.0	44.6	11.5
66	synthetic negation	2.2	0.0	7.0	7.0	2.1
67	analytic negation	18.9	11.2	32.0	20.8	6.5

Table 4: Descriptive statistics for Biber's corpus as a whole (Biber 1988: 77–78)

	Linguistic feature	Mean	Min. value	Max. value	Range	Std. deviation
1	past tense verbs	40.1	0.0	119.0	119.0	30.4
2	perfect aspect verbs	8.6	0.0	40.0	40.0	5.2
3	present tense verbs	77.7	12.0	182.0	170.0	34.3
4	place adverbials	3.1	0.0	24.0	24.0	3.4
5	time adverbials	5.2	0.0	24.0	24.0	3.5
6	first person pronouns	27.2	0.0	122.0	122.0	26.1
7	second person pronouns	9.9	0.0	72.0	72.0	13.8
8	third person pronouns	29.9	0.0	124.0	124.0	22.5
9	pronoun IT	10.3	0.0	47.0	47.0	7.1
10	demonstrative pronouns	4.6	0.0	30.0	30.0	4.8
11	indefinite pronouns	1.4	0.0	13.0	13.0	2.0
12	DO as pro-verb	3.0	0.0	22.0	22.0	3.5
13	direct WH-questions	0.2	0.0	4.0	4.0	0.6
14	nominalizations	19.9	0.0	71.0	71.0	14.4
15	gerunds	7.0	0.0	23.0	23.0	3.8
16	nouns	180.5	84.0	298.0	214.0	35.6
17	agentless passives	9.6	0.0	38.0	38.0	6.6
18	BY passives	0.8	0.0	8.0	8.0	1.3
19	BE as main verb	28.3	7.0	72.0	65.0	9.5
20	existential THERE	2.2	0.0	11.0	11.0	1.8
21	THAT verb complements	3.3	0.0	20.0	20.0	2.9
22	THAT adj. complements	0.3	0.0	3.0	3.0	0.6
23	WH clauses	0.6	0.0	7.0	7.0	1.0
24	infinitives	14.9	1.0	36.0	35.0	5.6
25	present participial clauses	1.0	0.0	11.0	11.0	1.7
26	past participial clauses	0.1	0.0	3.0	3.0	0.4
27	past prt. WHIZ deletions	2.5	0.0	21.0	21.0	3.1
28	present prt. WHIZ deletions	1.6	0.0	11.0	11.0	1.8
29	THAT relatives: subj. position	0.4	0.0	7.0	7.0	0.8
30	THAT relatives: obj. position	0.8	0.0	7.0	7.0	1.1
31	WH relatives: subj. position	2.1	0.0	15.0	15.0	2.0
32	WH relatives: obj. position	1.4	0.0	9.0	9.0	1.7
33	WH relatives: pied pipes	0.7	0.0	7.0	7.0	1.1
34	sentence relatives	0.1	0.0	3.0	3.0	0.4
35	adv. subordinator – cause	1.1	0.0	11.0	11.0	1.7
36	adv. sub. – concession	0.5	0.0	5.0	5.0	0.8

	Linguistic feature	Mean	Min. value	Max. value	Range	Std. deviation
37	adv. sub. – condition	2.5	0.0	13.0	13.0	2.2
38	adv. sub. – other	1.0	0.0	6.0	6.0	1.1
39	prepositional phrases	110.5	50.0	209.0	159.0	25.4
40	attributive adjectives	60.7	16.0	115.0	99.0	18.8
41	predicative adjectives	4.7	0.0	19.0	19.0	2.6
42	adverbs	65.6	22.0	125.0	103.0	17.6
43	type/token ratio	51.1	35.0	64.0	29.0	5.2
44	word length	4.5	3.7	5.3	1.6	0.4
45	conjuncts	1.2	0.0	12.0	12.0	1.6
46	downtoners	2.0	0.0	10.0	10.0	1.6
47	hedges	0.6	0.0	10.0	10.0	1.3
48	amplifiers	2.7	0.0	14.0	14.0	2.6
49	emphatics	6.3	0.0	22.0	22.0	4.2
50	discourse particles	1.2	0.0	15.0	15.0	2.3
51	demonstratives	9.9	0.0	22.0	22.0	4.2
52	possibility modals	5.8	0.0	21.0	21.0	3.5
53	necessity modals	2.1	0.0	13.0	13.0	2.1
54	prediction modals	5.6	0.0	30.0	30.0	4.2
55	public verbs	7.7	0.0	40.0	40.0	5.4
56	private verbs	18.0	1.0	54.0	53.0	10.4
57	suasive verbs	2.9	0.0	36.0	36.0	3.1
58	SEEM/APPEAR	0.8	0.0	6.0	6.0	1.0
59	contractions	13.5	0.0	89.0	89.0	18.6
60	THAT deletion	3.1	0.0	24.0	24.0	4.1
61	stranded prepositions	2.0	0.0	23.0	23.0	2.7
62	split infinitives	0.0	0.0	1.0	1.0	0.0
63	split auxiliaries	5.5	0.0	15.0	15.0	2.5
64	phrasal coordination	3.4	0.0	12.0	12.0	2.7
65	non-phrasal coordination	4.5	0.0	44.0	44.0	4.8
66	synthetic negation	1.7	0.0	8.0	8.0	1.6
67	analytic negation	8.5	0.0	32.0	32.0	6.1

Table 5: Normalized frequencies per Internet relay chat text

	Linguistic feature	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
1	past tense verbs	19.0	19.0	18.7	4.2	4.0	9.1	13.3	19.3	3.0	10.1
2	perfect aspect verbs	2.0	2.0	1.0	1.0	2.0	2.0	2.1	6.1	4.0	3.0
3	present tense verbs	138.1	161.2	145.7	137.2	170.0	152.2	137.4	144.6	135.4	149.9
4	place adverbials	3.0	4.0	3.1	3.1	1.0	1.0	4.1	1.0	2.0	3.0
5	time adverbials	13.0	7.0	9.4	7.3	4.0	12.1	12.3	7.1	4.0	8.1
6	first person pronouns	48.0	74.1	58.3	54.1	58.4	79.6	56.4	58.0	31.3	50.7
7	second person pronouns	58.1	70.1	61.4	56.1	51.3	40.3	57.4	49.9	26.3	33.4
8	third person pronouns	13.0	12.0	4.2	0.0	15.1	13.1	10.3	15.3	9.1	11.1
9	pronoun IT	11.0	13.0	12.5	6.2	10.1	17.1	15.4	10.2	14.1	13.2
10	demonstrative pronouns	3.0	9.0	6.2	4.2	5.0	3.0	7.2	11.2	6.1	11.1
11	indefinite pronouns	11.0	9.0	13.5	17.7	8.0	11.1	9.2	2.0	22.2	13.2
12	DO as pro-verb	0.0	3.0	10.4	6.2	2.0	8.1	3.1	2.0	5.1	2.0
13	direct WH-questions	1.0	5.0	2.1	8.3	1.0	4.0	1.0	1.0	3.0	8.1
14	nominalizations	0.0	1.0	5.2	5.2	5.0	1.0	1.0	0.0	12.1	10.1
15	gerunds	4.0	0.0	0.0	0.0	2.0	2.0	0.0	2.0	1.0	1.0
16	nouns	117.1	105.1	116.5	172.6	136.8	122.0	169.2	167.0	176.8	162.1
17	agentless passives	2.0	0.0	0.0	0.0	4.0	3.0	0.0	4.1	2.0	3.0
18	BY passives	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0
19	BE as main verb	17.0	25.0	27.1	20.8	19.1	18.1	25.6	31.6	19.2	20.3
20	existential THERE	2.0	1.0	0.0	1.0	2.0	0.0	0.0	3.1	1.0	2.0
21	THAT verb complements	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	2.0
22	THAT adj. complements	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
23	WH clauses	1.0	3.0	1.0	2.1	1.0	1.0	3.1	0.0	8.1	0.0
24	infinitives	11.0	6.0	6.2	10.4	12.1	20.2	9.2	7.1	20.2	18.2
25	present participial clauses	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
26	past participial clauses	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
27	past prt. WHIZ deletions	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0
28	present prt. WHIZ deletions	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
29	THAT relatives: subj. position	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
30	THAT relatives: obj. position	0.0	0.0	0.0	1.0	3.0	0.0	0.0	0.0	0.0	0.0
31	WH relatives: subj. position	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0
32	WH relatives: obj. position	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
33	WH relatives: pied pipes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
34	sentence relatives	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
35	adv. subordinator – cause	0.0	2.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0
36	adv. sub. – concession	0.0	0.0	2.1	1.0	1.0	1.0	1.0	0.0	0.0	1.0
37	adv. sub. – condition	3.0	0.0	3.1	2.1	5.0	0.0	3.1	2.0	2.0	1.0
38	adv. sub. – other	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	2.0
39	prepositional phrases	41.0	41.0	35.4	65.5	49.3	33.3	44.1	47.9	55.6	56.7
40	attributive adjectives	38.0	30.0	61.4	62.4	43.3	45.4	47.2	53.0	62.6	54.7

Linguistic feature	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
41 predicative adjectives	13.0	12.0	10.4	8.3	5.0	8.1	10.3	6.1	5.1	6.1
42 adverbs	109.1	83.1	82.2	81.1	93.6	67.5	84.1	88.6	57.6	51.7
43 type/token ratio	60.1	48.4	58.0	50.3	57.8	54.9	60.2	52.1	50.1	56.9
44 word length	4.4	3.9	4.0	3.8	3.9	3.8	4.0	4.0	4.2	4.3
45 conjuncts	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
46 downtoners	0.0	0.0	3.1	2.1	0.0	1.0	4.1	1.0	3.0	3.0
47 hedges	1.0	0.0	1.0	0.0	0.0	0.0	2.1	1.0	0.0	1.0
48 amplifiers	3.0	0.0	3.1	4.2	2.0	2.0	1.0	3.1	0.0	3.0
49 emphatics	6.0	7.0	6.2	3.1	7.0	4.0	14.4	11.2	12.1	7.1
50 discourse particles	1.0	4.0	5.2	5.2	1.0	2.0	4.1	2.0	2.0	6.1
51 demonstratives	2.0	4.0	1.0	4.2	4.0	4.0	4.1	6.1	2.0	2.0
52 possibility modals	7.0	6.0	5.2	4.2	6.0	4.0	2.1	7.1	6.1	15.2
53 necessity modals	0.0	1.0	0.0	2.1	3.0	2.0	2.1	4.1	4.0	0.0
54 prediction modals	1.0	1.0	7.3	5.2	10.1	14.1	5.1	7.1	4.0	6.1
55 public verbs	2.0	3.0	4.2	1.0	3.0	1.0	1.0	1.0	3.0	2.0
56 private verbs	16.0	25.0	22.9	6.2	11.1	17.1	23.6	19.3	24.2	13.2
57 suasive verbs	2.0	0.0	2.1	1.0	1.0	0.0	1.0	2.0	2.0	1.0
58 SEEM/APPEAR	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
59 contractions	40.0	33.0	37.5	23.9	36.2	34.3	19.5	36.7	24.2	22.3
60 THAT deletion	2.0	6.0	3.1	0.0	2.0	5.0	1.0	7.1	2.0	5.1
61 stranded prepositions	3.0	2.0	5.2	7.3	2.0	0.0	3.1	3.1	3.0	2.0
62 split infinitives	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63 split auxiliaries	2.0	0.0	1.0	2.1	2.0	3.0	0.0	5.1	1.0	4.1
64 phrasal coordination	4.0	0.0	1.0	4.2	2.0	5.0	5.1	11.2	2.0	8.1
65 non-phrasal coordination	0.0	5.0	2.1	5.2	0.0	3.0	6.2	2.0	1.0	2.0
66 synthetic negation	2.0	5.0	1.0	3.1	6.0	0.0	2.1	0.0	4.0	0.0
67 analytic negation	24.0	13.0	9.4	3.1	14.1	14.1	10.3	21.4	11.1	10.1

Table 6: Normalized frequencies per split-window ICQ chat text

Ling. feature	1	2	3	4	5	6	7	8	9	10	11	12
1 past tense verbs	26.1	26.1	27.1	27.8	29.5	65.4	20.0	61.4	42.0	26.9	32.2	27.5
2 perf. asp. verbs	1.2	2.2	7.0	1.7	2.9	0.0	2.5	6.4	3.2	0.0	3.2	1.5
3 pres. t. verbs	187.6	198.3	174.3	134.8	187.4	133.9	204.0	167.4	130.9	179.8	176.0	148.3
4 place adverbials	1.2	4.4	0.0	0.9	9.5	1.5	1.3	4.2	3.2	4.1	1.1	0.0
5 time adverbials	3.6	6.5	3.5	8.7	1.9	6.1	6.3	4.2	4.8	4.1	5.4	4.6
6 first pers. pron.	110.5	85.0	81.4	73.9	96.1	77.6	111.4	86.9	87.2	80.6	95.5	81.0
7 sec. pers. pron.	54.6	54.5	40.3	29.6	30.4	41.1	77.6	42.4	30.7	47.5	40.8	50.5
8 third pers. pron.	16.6	0.0	35.9	20.0	38.1	33.5	18.8	36.0	29.1	31.0	17.2	7.6
9 pronoun IT	26.1	21.8	28.0	24.3	21.9	16.7	13.8	23.3	11.3	24.8	17.2	9.2
10 dem. pronouns	27.3	13.1	12.3	20.9	19.0	15.2	16.3	21.2	11.3	18.6	12.9	9.2
11 indef. pronouns	1.2	0.0	6.1	4.3	5.7	6.1	13.8	12.7	8.1	4.1	8.6	1.5
12 DO as pro-verb	10.7	8.7	10.5	9.6	6.7	9.1	6.3	8.5	14.5	0.0	10.7	3.1
13 direct WH-q.	5.9	8.7	1.8	3.5	1.9	6.1	5.0	4.2	1.6	4.1	2.1	1.5
14 nominalizations	2.4	0.0	0.0	6.1	5.7	0.0	2.5	4.2	9.7	0.0	2.1	10.7
15 gerunds	0.0	0.0	0.0	3.5	1.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16 nouns	116.4	132.9	113.8	129.6	118.9	188.7	117.6	99.6	150.2	99.2	158.8	195.7
17 agentless pass.	1.2	2.2	0.9	2.6	2.9	0.0	1.3	0.0	0.0	0.0	1.1	1.5
18 BY passives	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.6	0.0	0.0	0.0
19 BE as main verb	23.8	34.9	26.3	17.4	40.0	24.4	18.8	31.8	17.8	18.6	20.4	22.9
20 exist. THERE	0.0	0.0	0.0	0.0	1.0	1.5	0.0	2.1	0.0	0.0	1.1	0.0
21 THAT v. compl.	0.0	2.2	1.8	2.6	0.0	1.5	1.3	4.2	1.6	2.1	1.1	0.0
22 THAT adj compl.	0.0	0.0	0.0	0.9	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
23 WH clauses	1.2	8.7	5.3	1.7	1.0	0.0	3.8	0.0	0.0	4.1	4.3	1.5
24 infinitives	7.1	21.8	7.0	6.1	12.4	4.6	12.5	12.7	17.8	20.7	6.4	13.8
25 pres. particip. cl.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
26 past particip. cl.	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
27 p.prt. WHIZ del.	0.0	2.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
28 pr. prt. WHIZ del	0.0	0.0	0.0	2.6	0.0	0.0	0.0	0.0	1.6	0.0	1.1	0.0
29 THAT rel: s. pos.	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
30 THAT rel: o. pos.	0.0	0.0	0.9	0.9	1.0	0.0	0.0	2.1	0.0	0.0	0.0	0.0
31 WH rel: s. pos.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
32 WH rel: o. pos.	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
33 WH rel: p. pipes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
34 sentence rel.	1.2	0.0	0.9	1.7	0.0	0.0	0.0	0.0	0.0	0.0	1.1	0.0
35 adv. sub. – cause	4.8	4.4	11.4	5.2	3.8	7.6	0.0	6.4	6.5	0.0	1.1	0.0
36 adv. sub. – conc.	1.2	0.0	1.8	1.7	1.9	0.0	1.3	0.0	3.2	2.1	2.1	0.0
37 adv. sub. – cond.	2.4	6.5	4.4	7.0	4.8	4.6	5.0	4.2	6.5	4.1	3.2	0.0
38 adv. sub. – other	2.4	0.0	1.8	0.9	2.9	1.5	0.0	0.0	0.0	0.0	0.0	0.0
39 prep. phrases	39.2	32.7	44.7	49.6	49.5	42.6	20.0	44.5	53.3	53.7	38.6	35.2
40 attributive adj.	22.6	37.0	28.0	33.9	49.5	24.4	31.3	23.3	32.3	22.7	23.6	36.7
41 predicative adj.	19.0	21.8	14.9	10.4	20.9	16.7	10.0	19.1	9.7	12.4	12.9	15.3

Ling. feature	1	2	3	4	5	6	7	8	9	10	11	12
42 adverbs	61.8	63.2	78.8	79.1	89.4	51.8	90.1	74.2	64.6	82.6	68.7	53.5
43 type/token ratio	51.0	48.6	47.5	52.3	56.6	50.6	54.8	48.5	50.8	47.0	56.1	60.0
44 word length	3.7	3.7	3.7	3.8	3.7	3.7	3.7	3.8	3.9	3.4	3.8	4.0
45 conjuncts	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.1	0.0
46 downtoners	1.2	0.0	0.0	1.7	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.5
47 hedges	5.9	4.4	4.4	7.0	4.8	0.0	0.0	0.0	0.0	0.0	2.1	1.5
48 amplifiers	2.4	0.0	1.8	0.0	3.8	0.0	0.0	0.0	6.5	0.0	2.1	1.5
49 emphatics	19.0	8.7	17.5	8.7	13.3	9.1	15.0	19.1	14.5	8.3	9.7	13.8
50 disc. particles	9.5	0.0	0.9	2.6	7.6	3.0	2.5	10.6	6.5	10.3	5.4	0.0
51 demonstratives	3.6	6.5	6.1	6.1	4.8	7.6	16.3	6.4	3.2	6.2	8.6	10.7
52 poss. modals	7.1	15.3	7.9	7.8	11.4	9.1	10.0	8.5	4.8	12.4	8.6	7.6
53 necess. modals	0.0	0.0	1.8	1.7	2.9	1.5	0.0	0.0	0.0	6.2	3.2	6.1
54 predict. modals	9.5	8.7	8.8	17.4	1.0	12.2	8.8	6.4	6.5	14.5	15.0	3.1
55 public verbs	2.4	8.7	5.3	6.1	5.7	0.0	6.3	10.6	4.8	8.3	3.2	1.5
56 private verbs	32.1	32.7	38.5	27.8	43.8	30.4	37.5	27.5	25.8	18.6	35.4	16.8
57 suasive verbs	1.2	0.0	1.8	0.9	1.0	1.5	1.3	4.2	0.0	0.0	0.0	6.1
58 SEEM/APPEAR	1.2	0.0	0.0	0.9	1.0	0.0	0.0	2.1	0.0	0.0	0.0	0.0
59 contractions	67.7	56.6	62.2	37.4	70.4	38.1	82.6	69.9	25.8	47.5	48.3	53.5
60 THAT deletion	14.3	10.9	6.1	3.5	8.6	10.7	6.3	21.2	11.3	8.3	12.9	4.6
61 stranded prep.	5.9	2.2	4.4	1.7	1.0	3.0	1.3	2.1	0.0	6.2	2.1	1.5
62 split infinitives	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.6	0.0	0.0	0.0
63 split auxiliaries	8.3	4.4	1.8	6.1	4.8	4.6	3.8	2.1	8.1	2.1	3.2	1.5
64 phrasal coord.	1.2	2.2	3.5	2.6	1.9	4.6	3.8	0.0	3.2	0.0	1.1	3.1
65 non-phr. coord.	7.1	4.4	10.5	7.0	5.7	4.6	1.3	19.1	1.6	8.3	4.3	0.0
66 synthetic neg.	4.8	2.2	0.9	3.5	2.9	1.5	1.3	4.2	0.0	0.0	4.3	4.6
67 analytic neg.	35.6	34.9	36.8	20.0	35.2	21.3	37.5	40.3	22.6	22.7	22.5	27.5

Table 7: Normalized frequencies per text in the SBC subset (spoken Am. English)

Ling. feat.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 past t. v.	9.8	47.6	16.8	28.2	37.5	66.1	27.9	65.1	20.9	50.5	67.3	1.4	22.3	42.0
2 perf. asp. v.	0.0	4.2	0.0	8.9	13.9	1.4	1.4	2.8	0.0	7.0	14.0	7.0	8.3	11.2
3 pres. t. v.	159.0	142.9	162.5	145.6	156.9	128.0	170.2	100.4	127.1	151.5	117.8	131.8	197.5	90.9
4 place adv.	2.8	1.4	0.0	0.0	0.0	2.8	2.8	2.8	2.8	2.8	0.0	1.4	0.0	0.0
5 time adv.	2.8	4.2	4.2	1.5	4.2	2.8	11.2	4.2	1.4	0.0	2.8	1.4	4.2	1.4
6 1st p. pron.	53.0	49.0	70.0	75.8	76.4	47.8	71.1	45.3	47.5	108.0	67.3	39.3	75.1	28.0
7 2nd p. pron.	75.3	33.6	35.0	37.1	34.7	30.9	26.5	31.1	51.7	39.3	37.9	14.0	50.1	7.0
8 3rd p. pron.	25.1	84.0	23.8	41.6	6.9	68.9	60.0	55.2	9.8	39.3	51.9	16.8	23.6	65.7
9 pron. IT	34.9	14.0	25.2	31.2	43.1	33.8	29.3	24.0	20.9	19.6	37.9	7.0	34.8	22.4
10 dem. pron.	11.2	30.8	23.8	13.4	13.9	12.7	7.0	15.6	22.3	15.4	15.4	14.0	20.9	7.0
11 ind. pron.	8.4	7.0	2.8	3.0	6.9	5.6	11.2	12.7	2.8	12.6	5.6	1.4	9.7	2.8
12 DO as prov.	18.1	5.6	2.8	4.5	6.9	4.2	12.6	2.8	9.8	7.0	2.8	2.8	2.8	2.8
13 dir. WH-q.	0.0	2.8	5.6	0.0	8.3	1.4	1.4	1.4	2.8	0.0	2.8	4.2	4.2	2.8
14 nominaliz.	8.4	2.8	0.0	4.5	5.6	0.0	7.0	7.1	2.8	7.0	2.8	43.5	1.4	11.2
15 gerunds	1.4	1.4	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	1.4	1.4	0.0	0.0
16 nouns	110.2	114.8	168.1	157.5	140.3	157.5	107.4	106.1	120.1	78.5	124.8	209.0	139.1	165.0
17 agentl pass.	1.4	5.6	1.4	3.0	0.0	1.4	1.4	4.2	0.0	4.2	1.4	2.8	2.8	7.0
18 BY pass.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
19 BE main v.	5.6	32.2	32.2	13.4	18.1	22.5	13.9	4.2	9.8	18.2	15.4	25.2	29.2	29.4
20 ex. THERE	4.2	1.4	4.2	0.0	4.2	2.8	0.0	2.8	1.4	4.2	4.2	5.6	4.2	0.0
21 THAT v. c.	0.0	1.4	1.4	1.5	1.4	0.0	1.4	0.0	0.0	11.2	1.4	4.2	0.0	2.8
22 THAT adj. c.	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0
23 WH cl.	1.4	2.8	2.8	1.5	2.8	1.4	4.2	1.4	1.4	5.6	4.2	1.4	0.0	1.4
24 infinitives	11.2	5.6	5.6	11.9	8.3	9.8	9.8	11.3	7.0	16.8	7.0	8.4	5.6	4.2
25 pr. part. cl.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0
26 p. part. cl.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.4
27 p. WHIZ d.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
28 pr. WHIZ d.	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
29 THAT rel. s.	0.0	0.0	0.0	0.0	0.0	1.4	1.4	1.4	0.0	0.0	0.0	0.0	0.0	2.8
30 THAT rel. o.	1.4	0.0	1.4	0.0	2.8	0.0	5.6	0.0	0.0	5.6	0.0	1.4	0.0	1.4
31 WH rel. s.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
32 WH rel. o.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
33 WH rel. pp	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
34 sent. rel.	0.0	0.0	0.0	1.5	0.0	0.0	2.8	1.4	1.4	0.0	0.0	1.4	0.0	0.0
35 adv. sub. c.	7.0	5.6	4.2	7.4	4.2	4.2	8.4	7.1	2.8	1.4	1.4	2.8	4.2	0.0
36 adv. s. con.	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.8	0.0
37 adv. s. cond.	2.8	0.0	5.6	4.5	2.8	0.0	12.6	2.8	1.4	4.2	5.6	12.6	2.8	0.0
38 adv. s. other	0.0	1.4	0.0	1.5	0.0	0.0	0.0	0.0	1.4	0.0	1.4	1.4	0.0	2.8
39 prep. phr.	64.2	43.4	46.2	65.4	56.9	52.0	53.0	75.0	71.2	53.3	56.1	81.3	54.2	82.5
40 attr. adj.	39.1	44.8	53.2	34.2	30.6	49.2	26.5	18.4	9.8	15.4	23.8	54.7	44.5	28.0

Ling. feat.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
41 pred.adj.	5.6	16.8	12.6	10.4	5.6	5.6	4.2	4.2	2.8	7.0	2.8	12.6	18.1	7.0
42 adverbs	86.5	93.8	86.8	62.4	72.2	68.9	65.6	67.9	60.1	58.9	72.9	58.9	55.6	42.0
43 type/token	35.5	48.3	41.5	48.5	43.0	45.8	44.8	43.8	34.0	43.3	47.5	46.3	46.3	50.5
44 word length	3.6	4.1	4.0	3.8	3.9	3.8	3.9	3.8	3.7	3.8	3.8	4.6	3.9	4.6
45 conjuncts	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	2.8	0.0	0.0
46 downtoners	2.8	0.0	0.0	0.0	2.8	1.4	0.0	2.8	0.0	1.4	0.0	0.0	7.0	0.0
47 hedges	11.2	0.0	1.4	1.5	1.4	1.4	1.4	8.5	0.0	4.2	0.0	1.4	0.0	0.0
48 amplifiers	1.4	2.8	0.0	0.0	4.2	1.4	0.0	0.0	0.0	8.4	1.4	2.8	1.4	0.0
49 emphatics	18.1	14.0	9.8	14.9	16.7	11.3	11.2	5.7	9.8	21.0	18.2	7.0	18.1	2.8
50 disc.part.	8.4	5.6	14.0	7.4	12.5	5.6	4.2	5.7	2.8	11.2	16.8	4.2	9.7	0.0
51 demonstr.	7.0	12.6	9.8	11.9	12.5	9.8	5.6	9.9	12.6	11.2	9.8	7.0	7.0	7.0
52 poss.mod.	11.2	8.4	16.8	7.4	4.2	4.2	7.0	4.2	8.4	1.4	12.6	7.0	11.1	8.4
53 nec.mod.	0.0	0.0	2.8	0.0	0.0	0.0	1.4	0.0	0.0	0.0	4.2	4.2	2.8	1.4
54 pred.mod.	2.8	5.6	16.8	10.4	0.0	2.8	2.8	5.7	11.2	5.6	7.0	4.2	11.1	11.2
55 pub.verbs	1.4	7.0	0.0	1.5	2.8	5.6	7.0	2.8	5.6	23.8	12.6	11.2	4.2	5.6
56 priv.verbs	55.8	26.6	23.8	38.6	31.9	39.4	34.9	38.2	16.8	65.9	28.1	18.2	40.3	11.2
57 suas.verbs	1.4	2.8	2.8	0.0	0.0	0.0	7.0	0.0	0.0	4.2	2.8	0.0	0.0	4.2
58 SEEM/APP.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0
59 contr.	33.5	64.4	56.0	35.7	68.1	49.2	58.6	33.9	40.5	36.5	44.9	35.1	89.0	33.6
60 THAT del.	8.4	14.0	9.8	8.9	5.6	4.2	2.8	5.7	2.8	9.8	8.4	1.4	15.3	2.8
61 str.prep.	0.0	2.8	2.8	1.5	2.8	4.2	4.2	5.7	5.6	7.0	0.0	4.2	5.6	1.4
62 split inf.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
63 split aux.	5.6	1.4	2.8	5.9	0.0	1.4	5.6	0.0	8.4	5.6	4.2	1.4	8.3	8.4
64 phr.coord.	2.8	7.0	4.2	4.5	8.3	2.8	2.8	1.4	0.0	2.8	1.4	1.4	0.0	8.4
65 n-p.coord.	46.0	16.8	5.6	14.9	15.3	21.1	25.1	28.3	15.4	12.6	19.6	5.6	1.4	5.6
66 synth.neg.	1.4	1.4	1.4	4.5	0.0	2.8	0.0	0.0	2.8	1.4	5.6	7.0	1.4	1.4
67 analyt.neg.	19.5	22.4	18.2	13.4	29.2	16.9	22.3	11.3	15.4	15.4	25.2	12.6	32.0	11.2

Appendix III. Raw frequencies of linguistic features

Tables 1a–3a present the raw frequencies per text of the linguistic features in the corpora investigated (for type/token ratio and word length, see Appendix II). The length of each text is shown in tables 1b–3b.

Table 1a: Raw frequencies per Internet relay chat text

Linguistic feature	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
1 past tense verbs	19	19	18	4	4	9	13	19	3	10
2 perfect aspect verbs	2	2	1	1	2	2	2	6	4	3
3 present tense verbs	138	161	140	132	169	151	134	142	134	148
4 place adverbials	3	4	3	3	1	1	4	1	2	3
5 time adverbials	13	7	9	7	4	12	12	7	4	8
6 first person pronouns	48	74	56	52	58	79	55	57	31	50
7 second person pronouns	58	70	59	54	51	40	56	49	26	33
8 third person pronouns	13	12	4	0	15	13	10	15	9	11
9 pronoun IT	11	13	12	6	10	17	15	10	14	13
10 demonstrative pronouns	3	9	6	4	5	3	7	11	6	11
11 indefinite pronouns	11	9	13	17	8	11	9	2	22	13
12 DO as pro-verb	0	3	10	6	2	8	3	2	5	2
13 direct WH-questions	1	5	2	8	1	4	1	1	3	8
14 nominalizations	0	1	5	5	5	1	1	0	12	10
15 gerunds	4	0	0	0	2	2	0	2	1	1
16 nouns	117	105	112	166	136	121	165	164	175	160
17 agentless passives	2	0	0	0	4	3	0	4	2	3
18 BY passives	0	0	0	0	0	0	0	0	0	2
19 BE as main verb	17	25	26	20	19	18	25	31	19	20
20 existential THERE	2	1	0	1	2	0	0	3	1	2
21 THAT verb complements	0	0	0	0	0	0	1	0	0	2
22 THAT adj. complements	0	0	0	0	0	0	0	0	0	0
23 WH clauses	1	3	1	2	1	1	3	0	8	0
24 infinitives	11	6	6	10	12	20	9	7	20	18
25 present participial clauses	0	0	0	0	0	0	0	0	0	0
26 past participial clauses	0	0	0	0	0	0	0	0	0	0
27 past prt. WHIZ deletions	0	0	0	0	1	0	0	0	1	0
28 present prt. WHIZ deletions	0	0	0	0	0	0	0	0	0	0
29 THAT relatives: subj. position	0	0	0	0	0	0	0	0	0	0
30 THAT relatives: obj. position	0	0	0	1	3	0	0	0	0	0
31 WH relatives: subj. position	0	0	0	0	0	0	0	0	3	0
32 WH relatives: obj. position	0	0	0	0	0	0	0	0	0	0

	Linguistic feature	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
33	WH relatives: pied pipes	0	0	0	0	0	0	0	0	0	0
34	sentence relatives	0	0	0	0	0	0	1	0	0	0
35	adv. subordinator – cause	0	2	0	1	0	0	0	0	0	1
36	adv. sub. – concession	0	0	2	1	1	1	1	0	0	1
37	adv. sub. – condition	3	0	3	2	5	0	3	2	2	1
38	adv. sub. – other	0	1	0	0	0	0	0	0	1	2
39	prepositional phrases	41	41	34	63	49	33	43	47	55	56
40	attributive adjectives	38	30	59	60	43	45	46	52	62	54
41	predicative adjectives	13	12	10	8	5	8	10	6	5	6
42	adverbs	109	83	79	78	93	67	82	87	57	51
45	conjuncts	0	0	0	0	0	0	0	0	0	0
46	downtoners	0	0	3	2	0	1	4	1	3	3
47	hedges	1	0	1	0	0	0	2	1	0	1
48	amplifiers	3	0	3	4	2	2	1	3	0	3
49	emphatics	6	7	6	3	7	4	14	11	12	7
50	discourse particles	1	4	5	5	1	2	4	2	2	6
51	demonstratives	2	4	1	4	4	4	4	6	2	2
52	possibility modals	7	6	5	4	6	4	2	7	6	15
53	necessity modals	0	1	0	2	3	2	2	4	4	0
54	prediction modals	1	1	7	5	10	14	5	7	4	6
55	public verbs	2	3	4	1	3	1	1	1	3	2
56	private verbs	16	25	22	6	11	17	23	19	24	13
57	suasive verbs	2	0	2	1	1	0	1	2	2	1
58	SEEM/APPEAR	0	0	0	0	1	0	0	0	0	0
59	contractions	40	33	36	23	36	34	19	36	24	22
60	THAT deletion	2	6	3	0	2	5	1	7	2	5
61	stranded prepositions	3	2	5	7	2	0	3	3	3	2
62	split infinitives	0	0	0	0	0	0	0	0	0	0
63	split auxiliaries	2	0	1	2	2	3	0	5	1	4
64	phrasal coordination	4	0	1	4	2	5	5	11	2	8
65	non-phrasal coordination	0	5	2	5	0	3	6	2	1	2
66	synthetic negation	2	5	1	3	6	0	2	0	4	0
67	analytic negation	24	13	9	3	14	14	10	21	11	10

Table 1b: Length of the Internet relay chat texts

	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
Text length	999	999	961	962	994	992	975	982	990	987

Table 2a: Raw frequencies per split-window ICQ chat text

	Ling. feature	1	2	3	4	5	6	7	8	9	10	11	12
1	past tense verbs	22	12	31	32	31	43	16	29	26	13	30	18
2	perf. asp. verbs	1	1	8	2	3	0	2	3	2	0	3	1
3	pres. t. verbs	158	91	199	155	197	88	163	79	81	87	164	97
4	place adverbials	1	2	0	1	10	1	1	2	2	2	1	0
5	time adverbials	3	3	4	10	2	4	5	2	3	2	5	3
6	first pers. pron.	93	39	93	85	101	51	89	41	54	39	89	53
7	sec. pers. pron.	46	25	46	34	32	27	62	20	19	23	38	33
8	third pers. pron.	14	0	41	23	40	22	15	17	18	15	16	5
9	pronoun IT	22	10	32	28	23	11	11	11	7	12	16	6
10	dem. pronouns	23	6	14	24	20	10	13	10	7	9	12	6
11	indef. pronouns	1	0	7	5	6	4	11	6	5	2	8	1
12	DO as pro-verb	9	4	12	11	7	6	5	4	9	0	10	2
13	direct WH-q.	5	4	2	4	2	4	4	2	1	2	2	1
14	nominalizations	2	0	0	7	6	0	2	2	6	0	2	7
15	gerunds	0	0	0	4	2	0	0	0	0	0	0	0
16	nouns	98	61	130	149	125	124	94	47	93	48	148	128
17	agentless pass.	1	1	1	3	3	0	1	0	0	0	1	1
18	BY passives	0	0	0	0	0	0	0	0	1	0	0	0
19	BE as main verb	20	16	30	20	42	16	15	15	11	9	19	15
20	exist. THERE	0	0	0	0	1	1	0	1	0	0	1	0
21	THAT v. compl.	0	1	2	3	0	1	1	2	1	1	1	0
22	THAT adj compl.	0	0	0	1	1	0	0	0	0	0	0	0
23	WH clauses	1	4	6	2	1	0	3	0	0	2	4	1
24	infinitives	6	10	8	7	13	3	10	6	11	10	6	9
25	pres. particip. cl.	0	0	0	0	0	0	0	0	0	0	0	0
26	past particip. cl.	1	0	0	0	0	0	0	0	0	0	0	0
27	p. prt. WHIZ del.	0	1	0	0	0	0	0	0	0	0	0	0
28	pr. prt. WHIZ del	0	0	0	3	0	0	0	0	1	0	1	0
29	THAT rel: s. pos.	0	0	0	1	0	0	0	0	0	0	0	0
30	THAT rel: o. pos.	0	0	1	1	1	0	0	1	0	0	0	0
31	WH rel: s. pos.	0	0	0	0	0	0	0	0	0	0	0	0
32	WH rel: o. pos.	0	0	0	1	0	0	0	0	0	0	0	0
33	WH rel: p. pipes	0	0	0	0	0	0	0	0	0	0	0	0
34	sentence rel.	1	0	1	2	0	0	0	0	0	0	1	0
35	adv. sub. – cause	4	2	13	6	4	5	0	3	4	0	1	0
36	adv. sub. – conc.	1	0	2	2	2	0	1	0	2	1	2	0

Ling. feature	1	2	3	4	5	6	7	8	9	10	11	12
37 adv. sub. – cond.	2	3	5	8	5	3	4	2	4	2	3	0
38 adv. sub. – other	2	0	2	1	3	1	0	0	0	0	0	0
39 prep. phrases	33	15	51	57	52	28	16	21	33	26	36	23
40 attributive adj.	19	17	32	39	52	16	25	11	20	11	22	24
41 predicative adj.	16	10	17	12	22	11	8	9	6	6	12	10
42 adverbs	52	29	90	91	94	34	72	35	40	40	64	35
45 conjuncts	1	0	0	0	0	0	0	0	0	0	2	0
46 downtoners	1	0	0	2	1	0	0	0	0	0	0	1
47 hedges	5	2	5	8	5	0	0	0	0	0	2	1
48 amplifiers	2	0	2	0	4	0	0	0	4	0	2	1
49 emphatics	16	4	20	10	14	6	12	9	9	4	9	9
50 disc. particles	8	0	1	3	8	2	2	5	4	5	5	0
51 demonstratives	3	3	7	7	5	5	13	3	2	3	8	7
52 poss. modals	6	7	9	9	12	6	8	4	3	6	8	5
53 necess. modals	0	0	2	2	3	1	0	0	0	3	3	4
54 predict. modals	8	4	10	20	1	8	7	3	4	7	14	2
55 public verbs	2	4	6	7	6	0	5	5	3	4	3	1
56 private verbs	27	15	44	32	46	20	30	13	16	9	33	11
57 suasive verbs	1	0	2	1	1	1	1	2	0	0	0	4
58 SEEM/APPEAR	1	0	0	1	1	0	0	1	0	0	0	0
59 contractions	57	26	71	43	74	25	66	33	16	23	45	35
60 THAT deletion	12	5	7	4	9	7	5	10	7	4	12	3
61 stranded prep.	5	1	5	2	1	2	1	1	0	3	2	1
62 split infinitives	0	0	0	0	0	0	0	0	1	0	0	0
63 split auxiliaries	7	2	2	7	5	3	3	1	5	1	3	1
64 phrasal coord.	1	1	4	3	2	3	3	0	2	0	1	2
65 non-phr. coord.	6	2	12	8	6	3	1	9	1	4	4	0
66 synthetic neg.	4	1	1	4	3	1	1	2	0	0	4	3
67 analytic neg.	30	16	42	23	37	14	30	19	14	11	21	18

Table 2b: Length of the split-window ICQ chat texts

	1	2	3	4	5	6	7	8	9	10	11	12
Text length	842	459	1142	1150	1051	657	799	472	619	484	932	654

Table 3a: Raw frequencies per text in the SBC subset (spoken American English)

Ling. feat.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 past t. v.	7	34	12	19	27	47	20	46	15	36	48	1	16	30
2 perf. asp. v.	0	3	0	6	10	1	1	2	0	5	10	5	6	8
3 pres. t. v.	114	102	116	98	113	91	122	71	91	108	84	94	142	65
4 place adv.	2	1	0	0	0	2	2	2	2	2	0	1	0	0
5 time adv.	2	3	3	1	3	2	8	3	1	0	2	1	3	1
6 1st p. pron.	38	35	50	51	55	34	51	32	34	77	48	28	54	20
7 2nd p. pron.	54	24	25	25	25	22	19	22	37	28	27	10	36	5
8 3rd p. pron.	18	60	17	28	5	49	43	39	7	28	37	12	17	47
9 pron. IT	25	10	18	21	31	24	21	17	15	14	27	5	25	16
10 dem. pron.	8	22	17	9	10	9	5	11	16	11	11	10	15	5
11 ind. pron.	6	5	2	2	5	4	8	9	2	9	4	1	7	2
12 DO as pr-v.	13	4	2	3	5	3	9	2	7	5	2	2	2	2
13 dir. WH-q.	0	2	4	0	6	1	1	1	2	0	2	3	3	2
14 nominaliz.	6	2	0	3	4	0	5	5	2	5	2	31	1	8
15 gerunds	1	1	0	0	1	0	0	0	0	0	1	1	0	0
16 nouns	79	82	120	106	101	112	77	75	86	56	89	149	100	118
17 agentl pass.	1	4	1	2	0	1	1	3	0	3	1	2	2	5
18 BY pass.	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19 BE main v.	4	23	23	9	13	16	10	3	7	13	11	18	21	21
20 ex. THERE	3	1	3	0	3	2	0	2	1	3	3	4	3	0
21 THAT v. c.	0	1	1	1	1	0	1	0	0	8	1	3	0	2
22 THAT adj c.	0	1	0	0	0	0	0	0	0	0	1	0	0	0
23 WH cl.	1	2	2	1	2	1	3	1	1	4	3	1	0	1
24 infinitives	8	4	4	8	6	7	7	8	5	12	5	6	4	3
25 pr. part. cl.	0	0	0	0	0	0	0	1	0	0	0	0	0	0
26 p. part. cl.	0	0	0	0	0	0	0	0	0	0	0	0	0	1
27 p. WHIZ d.	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28 pr. WHIZ d.	0	0	0	0	0	0	1	0	0	0	0	0	0	0
29 THAT rel: s.	0	0	0	0	0	1	1	1	0	0	0	0	0	2
30 THAT rel: o.	1	0	1	0	2	0	4	0	0	4	0	1	0	1
31 WH rel: s.	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32 WH rel: o.	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33 WH rel: pp	0	1	0	0	0	0	0	0	0	0	0	0	0	0
34 sent. rel.	0	0	0	1	0	0	2	1	1	0	0	1	0	0
35 adv. sub. c.	5	4	3	5	3	3	6	5	2	1	1	2	3	0
36 adv. s. con.	0	0	1	0	0	0	0	0	0	0	0	0	2	0

Ling. feat.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
37 adv s. cond.	2	0	4	3	2	0	9	2	1	3	4	9	2	0
38 adv s. other	0	1	0	1	0	0	0	0	1	0	1	1	0	2
39 prep. phr.	46	31	33	44	41	37	38	53	51	38	40	58	39	59
40 attr. adj.	28	32	38	23	22	35	19	13	7	11	17	39	32	20
41 pred. adj.	4	12	9	7	4	4	3	3	2	5	2	9	13	5
42 adverbs	62	67	62	42	52	49	47	48	43	42	52	42	40	30
45 conjuncts	0	0	0	0	0	1	0	0	0	0	0	2	0	0
46 downtoners	2	0	0	0	2	1	0	2	0	1	0	0	5	0
47 hedges	8	0	1	1	1	1	1	6	0	3	0	1	0	0
48 amplifiers	1	2	0	0	3	1	0	0	0	6	1	2	1	0
49 emphatics	13	10	7	10	12	8	8	4	7	15	13	5	13	2
50 disc. part.	6	4	10	5	9	4	3	4	2	8	12	3	7	0
51 demonstr.	5	9	7	8	9	7	4	7	9	8	7	5	5	5
52 poss. mod.	8	6	12	5	3	3	5	3	6	1	9	5	8	6
53 nec. mod.	0	0	2	0	0	0	1	0	0	0	3	3	2	1
54 pred. mod.	2	4	12	7	0	2	2	4	8	4	5	3	8	8
55 pub. verbs	1	5	0	1	2	4	5	2	4	17	9	8	3	4
56 priv. verbs	40	19	17	26	23	28	25	27	12	47	20	13	29	8
57 suas. verbs	1	2	2	0	0	0	5	0	0	3	2	0	0	3
58 SEEM/APP.	0	0	0	0	0	0	0	1	0	0	0	0	0	0
59 contr.	24	46	40	24	49	35	42	24	29	26	32	25	64	24
60 THAT del.	6	10	7	6	4	3	2	4	2	7	6	1	11	2
61 str. prep.	0	2	2	1	2	3	3	4	4	5	0	3	4	1
62 split inf.	0	0	0	0	0	0	0	0	0	0	0	0	0	0
63 split aux.	4	1	2	4	0	1	4	0	6	4	3	1	6	6
64 phr. coord.	2	5	3	3	6	2	2	1	0	2	1	1	0	6
65 n-p. coord.	33	12	4	10	11	15	18	20	11	9	14	4	1	4
66 synth. neg.	1	1	1	3	0	2	0	0	2	1	4	5	1	1
67 analyt. neg.	14	16	13	9	21	12	16	8	11	11	18	9	23	8

Table 3b: Length of the SBC subset texts (spoken American English)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Text length	717	714	714	673	720	711	717	707	716	713	713	713	719	715

Appendix IV. Examples of excluded material

Certain messages and strings of text were excluded from the conversational writing logs and the SBC subset before the texts were annotated for the features in Biber's (1988) methodology. Typical excluded instances are exemplified below.

Excluded from Internet relay chat	Example
Bracketed nickname turn indicators	<River>
Server-generated messages (session start messages, time stamps, join- and leave messages)	Session Start: Mon Mar 25 18:01:47 2002 [18:01] *** Now talking in #family *** edi-tr has joined #family *** edi-tr has quit IRC (Killed (NickServ (Nickname Enforcement)))
Channel operator interference	*** ezococx was kicked by Sp0ck (banned: spam)
Action commands (including graphic noise)	* big-dog °@o¿,,¿o@°°°°°° HELLO WELCOME TO>>> #family CHANNEL°<<< memyselfandi @°°°°°°@o¿,,¿o@ * NA_TuPaC slaps ma7ash around a large trout * SwampRocker Bunny. (Y) ... (Y).. ^gypsy^ . (Y)... (Y).. Hugs * SwampRocker Bunny . (°°) . (°°). ^gypsy^ . (°°).. (°°). Hugs * SwampRocker Bunny . ()^-°-() ()^-°-() ^gypsy^ ()^-°-() ()^-°-() Hugs * SwampRocker Bunny . ()-() ()-() ^gypsy^ ()-() ()-() Hugs
Foreign language	alguien habla espaÑol???
Excluded from split-window ICQ	Example
Bracketed nickname turn indicators	<5>
Action tropes	<i>2 points finger at you, scolding you for your actions</i> <i>B picks a flower and hands it to you</i>
Foreign language	hablamos espanol abren los libros a la paginaaaaaaaaaaaaaa tres
Excluded from face-to-face SBC	Example
Foreign language	Que es mas o menos. No es exelente, pero es mas o menos.

Appendix V. Features with a |standard score| >2.0

Table 1 lists the features with a standard score above 2.0, or below -2.0, in the genres studied, the most influential (most salient) contributors to the dimension scores of the particular genre. Section 4.4, and part of 4.2, explore the most salient features of the conversational writing genres (split-window ICQ chat and Internet relay chat) and present their distribution in writing, ACMC and speech. The procedure of standard score calculation is described in section 3.5.

Table 1: Features with a |standard score| >2.0 in the genres studied. A hyphen (-) indicates that the genre has no feature with a |standard score| >2.0

Medium	Genre	Feature	Standard score
SSCMC	Split-window ICQ chat	direct WH-questions	6.1
		predicative adjectives	4.1
		analytic negation	3.5
		prepositional phrases	-2.7
		present tense verbs	2.6
		second person pronouns	2.5
		demonstrative pronouns	2.5
		first person pronouns	2.4
		indefinite pronouns	2.3
		contractions	2.2
SCMC	Internet relay chat	direct WH-questions	5.4
		indefinite pronouns	5.2
		second person pronouns	2.9
		prepositional phrases	-2.5
Speech	Face-to-face conversations SBC	direct WH-questions	4.2
		discourse particles	2.8
		indefinite pronouns	2.6
		non-phrasal coordination	2.5
		demonstrative pronouns	2.4
		pronoun IT	2.4

Medium	Genre	Feature	Standard score
	Face-to-face conversations LLC	–	
	Telephone conversations	discourse particles	2.3
		contractions	2.2
	Interviews	–	
	Broadcasts	time adverbials	2.5
	Spontaneous speeches	non-phrasal coordination	2.2
	Prepared speeches	–	
ACMC	BBS conferencing “ELC other”	sentence relatives	5.7
		direct WH-questions	4.6
		non-phrasal coordination	2.6
		adverbial subordinators -condition	2.5
Writing	Press reportage	–	
	Press editorials	–	
	Press reviews	–	
	Religion	–	
	Hobbies	–	
	Popular lore	–	
	Biographies	–	
	Official documents	–	
	Academic prose	–	
	General fiction	–	
	Mystery fiction	–	
	Science fiction	–	
	Adventure fiction	–	
	Romantic fiction	third person pronouns	2.2
		present participial clauses	2.1
	Humor	–	
	Personal letters	THAT deletion	2.4
		hedges	2.1
	Professional letters	–	

Appendix VI. Statistical tests of salient features

Table 1 presents the values for probability (p) from t -tests of the feature distributions in SCMC, SSCMS, writing and speech for the salient features in conversational writing p discussed in chapter 4. For some of the features, or combinations of features, p -values are not available (“n.a.”) owing to the unavailability in Biber (1988) of the requisite data for the test. As regards inserts, no annotation of Biber’s (1988) texts of writing or speech was carried out; instead, the p -values for inserts given in table 1 in the comparisons to “speech” reflect for “speech” only the face-to-face conversations from SBC (as noted in section 4.6). With regard to emotives, the tests here reflect that none of the written (LOB) or spoken (LLC or SBC) texts contains emotives.

Table 1: Values for probability (p) from t -tests of features in writing, speech, SCMC and SSCMC. Significant differences in bold ($p < .05$). The p -values have not been multiplicity adjusted

	SCMC vs. Writing	SSCMC vs. Writing	SCMC vs. Speech	SSCMC vs. Speech
possibility modals	0.3916	0.0005	0.4443	0.0352
necessity modals	0.6291	0.8451	0.8459	1.0000
prediction modals	0.6291	0.0207	1.0000	0.0420
total modals	n.a.	n.a.	n.a.	n.a.
first person pronouns	<.0001	<.0001	0.3916	<.0001
second person pronouns	<.0001	<.0001	0.0002	0.0002
third person pronouns	<.0001	0.0840	<.0001	0.1618
first+second pers. pron.	n.a.	n.a.	n.a.	n.a.
total pronouns	n.a.	n.a.	n.a.	n.a.
word length	<.0001	<.0001	0.0287	<.0001
type/token ratio	0.1950	0.4985	0.0003	0.0013
direct WH-questions	0.0049	0.0001	0.0176	0.0008
analytic negation	0.0067	<.0001	0.6985	<.0001
demonstrative pronouns	0.0015	<.0001	0.0042	0.0035
indefinite pronouns	0.0002	0.0018	0.0008	0.0420
present tense verbs	<.0001	<.0001	<.0001	<.0001
predicative adjectives	0.0036	<.0001	0.0042	<.0001
contractions	<.0001	<.0001	0.0899	0.0029
prepositional phrases	<.0001	<.0001	<.0001	<.0001
inserts	n.a.	n.a.	0.0067	0.2553
emotives	0.0003	0.0035	0.0003	0.0035

Appendix VII. Word lists for the corpora studied

Table 1: Word frequency lists for the corpora studied: IRC, split-window ICQ chat and the SBC subset, by rank for the fifty most frequent types (not lemmatized). N.B. raw frequencies for full corpora

N	Internet relay chat			Split-window ICQ			SBC subset		
	Word	Freq.	%	Word	Freq.	%	Word	Freq.	%
1	I	265	2.6	I	455	4.9	I	352	3.5
2	YOU	212	2.1	TO	217	2.3	YOU	305	3.1
3	HI	202	2.0	YOU	193	2.1	THE	304	3.0
4	TO	169	1.7	THE	170	1.8	AND	289	2.9
5	A	146	1.5	THAT	162	1.7	IT	199	2.0
6	U	134	1.3	IT	145	1.5	THAT	196	2.0
7	THE	132	1.3	AND	143	1.5	TO	190	1.9
8	LOL	128	1.3	A	132	1.4	OF	163	1.6
9	IS	112	1.1	U	112	1.2	A	159	1.6
10	ME	111	1.1	LIKE	104	1.1	KNOW	138	1.4
11	AND	99	1.0	KNOW	102	1.1	HAVE	110	1.1
12	IT	86	0.9	SO	98	1.0	IN	104	1.0
13	ARE	84	0.8	WHAT	96	1.0	THEY	100	1.0
14	IN	83	0.8	ME	93	1.0	YEAH	99	1.0
15	HEY	81	0.8	IS	89	1.0	WAS	93	0.9
16	FROM	74	0.7	BUT	86	0.9	IS	87	0.9
17	HELLO	71	0.7	NOT	75	0.8	WHAT	81	0.8
18	ALL	68	0.7	DONT	69	0.7	HE	79	0.8
19	HERE	66	0.7	IN	67	0.7	LIKE	78	0.8
20	THAT	66	0.7	MY	63	0.7	SO	74	0.7
21	HOW	64	0.6	NO	62	0.7	WELL	69	0.7
22	WHAT	61	0.6	WE	62	0.7	BUT	68	0.7
23	FOR	60	0.6	YEAH	62	0.7	WE	68	0.7
24	OF	60	0.6	DO	61	0.7	DON'T	66	0.7
25	DO	58	0.6	JUST	61	0.7	THAT'S	65	0.7
26	GOOD	54	0.5	YEA	61	0.7	THIS	65	0.7
27	HAVE	54	0.5	HAVE	59	0.6	IT'S	64	0.6
28	MY	54	0.5	BE	54	0.6	NO	63	0.6
29	22	53	0.5	WAS	54	0.6	ON	63	0.6
30	ANY	53	0.5	OF	52	0.6	DO	58	0.6
31	BUT	51	0.5	SHE	52	0.6	OH	58	0.6

	Internet relay chat			Split-window ICQ			SBC subset		
N	Word	Freq.	%	Word	Freq.	%	Word	Freq.	%
32	CHAT	51	0.5	THATS	49	0.5	JUST	57	0.6
33	JUST	51	0.5	WITH	47	0.5	OR	57	0.6
34	WHERE	48	0.5	HE	46	0.5	UM	57	0.6
35	NOT	47	0.5	THIS	46	0.5	RIGHT	56	0.6
36	THERE	47	0.5	GO	45	0.5	MHM	55	0.6
37	BACK	46	0.5	IM	45	0.5	WITH	52	0.5
38	OK	46	0.5	OH	45	0.5	ONE	50	0.5
39	KNOW	45	0.4	YOUR	45	0.5	ABOUT	49	0.5
40	AM	44	0.4	FOR	44	0.5	ALL	49	0.5
41	BE	43	0.4	ARE	43	0.5	FOR	48	0.5
42	YOUR	42	0.4	ON	43	0.5	THEN	47	0.5
43	YES	41	0.4	UP	43	0.5	REALLY	45	0.5
44	NO	40	0.4	GET	42	0.4	I'M	43	0.4
45	OUT	40	0.4	WELL	42	0.4	AT	42	0.4
46	CAN	39	0.4	WOULD	42	0.4	MEAN	42	0.4
47	CHANEL	37	0.4	DID	41	0.4	TWO	42	0.4
48	WELL	37	0.4	HER	41	0.4	IF	41	0.4
49	LIKE	36	0.4	ALL	40	0.4	GET	40	0.4
50	UP	36	0.4	IF	40	0.4	NOT	40	0.4

Appendix VIII. Dimension score statistics for Biber's (1988) genres

Dimension 1: Informational versus Involved Production

Dimension 2: Narrative versus Non-Narrative Concerns

Dimension 3: Explicit/Elaborated versus Situation-Dependent Reference

Dimension 4: Overt Expression of Persuasion/Argumentation

Dimension 5: Abstract/Impersonal versus Non-Abstract/Non-Impersonal Information

Dimension 6: On-Line Informational Elaboration

Table 1: Descriptive dimension statistics for Biber's genres (Biber 1988: 122–125)

Speech (LLC)	Dimension	Mean	Minimum value	Maximum value	Range	Standard deviation
Face-to-face conversations	Dimension 1	35.3	17.7	54.1	36.4	9.1
	Dimension 2	-0.6	-4.4	4.0	8.4	2.0
	Dimension 3	-3.9	-10.5	1.6	12.1	2.1
	Dimension 4	-0.3	-5.2	6.5	11.7	2.4
	Dimension 5	-3.2	-4.5	0.1	4.6	1.1
	Dimension 6	0.3	-3.6	6.5	10.1	2.2
Telephone conversations	Dimension 1	37.2	7.2	52.9	45.8	9.9
	Dimension 2	-2.1	-4.2	4.7	8.9	2.2
	Dimension 3	-5.2	-10.1	2.3	12.5	2.9
	Dimension 4	0.6	-4.9	8.4	13.3	3.6
	Dimension 5	-3.7	-4.8	0.1	4.9	1.2
	Dimension 6	-0.9	-4.8	3.3	8.1	2.1
Interviews	Dimension 1	17.1	3.5	36.0	32.5	10.7
	Dimension 2	-1.1	-5.0	2.7	7.8	2.1
	Dimension 3	-0.4	-6.3	8.3	14.7	4.0
	Dimension 4	1.0	-3.4	6.1	9.5	2.4
	Dimension 5	-2.0	-4.1	0.4	4.5	1.3
	Dimension 6	3.1	-1.4	10.5	11.9	2.6
Broadcasts	Dimension 1	-4.3	-19.6	16.9	36.5	10.7
	Dimension 2	-3.3	-5.2	-0.6	4.6	1.2
	Dimension 3	-9.0	-15.8	-2.2	13.6	4.4
	Dimension 4	-4.4	-6.9	-0.3	6.5	2.0
	Dimension 5	-1.7	-4.7	5.4	10.0	2.8
	Dimension 6	-1.3	-3.6	1.7	5.3	1.6

Speech (LLC)	Dimension	Mean	Minimum value	Maximum value	Range	Standard deviation
Spontaneous speeches	Dimension 1	18.2	-2.6	33.1	35.7	12.3
	Dimension 2	1.3	-3.8	9.4	13.2	3.6
	Dimension 3	1.2	-5.4	9.7	15.1	4.3
	Dimension 4	0.3	-5.5	7.4	12.9	4.4
	Dimension 5	-2.6	-4.5	0.7	5.1	1.7
	Dimension 6	2.6	-2.4	10.6	13.0	4.2
Prepared speeches	Dimension 1	2.2	-7.3	14.8	22.1	6.7
	Dimension 2	0.7	-4.9	6.1	11.0	3.3
	Dimension 3	0.3	-5.6	6.1	11.6	3.6
	Dimension 4	0.4	-4.4	11.2	15.5	4.1
	Dimension 5	-1.9	-3.9	1.0	5.0	1.4
	Dimension 6	3.4	-0.8	7.5	8.3	2.8
Writing (LOB)	Dimension	Mean	Minimum value	Maximum value	Range	Standard deviation
Press reportage	Dimension 1	-15.1	-24.1	-3.1	21.0	4.5
	Dimension 2	0.4	-3.2	7.7	10.9	2.1
	Dimension 3	-0.3	-6.2	6.5	12.7	2.9
	Dimension 4	-0.7	-6.0	5.7	11.7	2.6
	Dimension 5	0.6	-4.4	5.5	9.9	2.4
	Dimension 6	-0.9	-4.0	3.9	8.0	1.8
Press editorials	Dimension 1	-10.0	-18.0	1.6	19.5	3.8
	Dimension 2	-0.8	-3.5	1.8	5.3	1.4
	Dimension 3	1.9	-2.9	5.4	8.3	2.0
	Dimension 4	3.1	-1.8	9.3	11.2	3.2
	Dimension 5	0.3	-2.4	4.5	6.9	2.0
	Dimension 6	1.5	-1.8	5.7	7.5	1.6
Press reviews	Dimension 1	-13.9	-20.5	-8.6	11.8	3.9
	Dimension 2	-1.6	-4.3	2.7	7.0	1.9
	Dimension 3	4.3	-1.8	10.3	12.2	3.7
	Dimension 4	-2.8	-6.5	1.5	8.1	2.0
	Dimension 5	0.8	-3.1	5.8	9.0	2.1
	Dimension 6	-1.0	-3.7	3.9	7.6	1.9
Religion	Dimension 1	-7.0	-17.2	16.5	33.7	8.3
	Dimension 2	-0.7	-4.4	5.5	9.9	2.7

Writing (LOB)	Dimension	Mean	Minimum value	Maximum value	Range	Standard deviation
	Dimension 3	3.7	-0.6	9.8	10.4	3.3
	Dimension 4	0.2	-2.9	6.2	9.1	2.7
	Dimension 5	1.4	-2.4	5.2	7.6	2.4
	Dimension 6	1.0	-2.0	6.5	8.4	2.4
Hobbies	Dimension 1	-10.1	-18.8	-2.0	16.9	5.0
	Dimension 2	-2.9	-4.8	1.6	6.4	1.9
	Dimension 3	0.3	-5.7	10.0	15.7	3.6
	Dimension 4	1.7	-5.8	11.0	16.8	4.6
	Dimension 5	1.2	-3.6	13.0	16.6	4.2
	Dimension 6	-0.7	-3.0	2.5	5.5	1.8
Popular lore	Dimension 1	-9.3	-24.7	9.9	34.5	11.3
	Dimension 2	-0.1	-4.7	9.2	13.9	3.7
	Dimension 3	2.3	-2.1	11.5	13.6	3.5
	Dimension 4	-0.3	-4.4	13.3	17.8	4.8
	Dimension 5	0.1	-3.9	3.0	6.9	2.3
	Dimension 6	-0.8	-3.8	3.8	7.6	1.8
Biographies	Dimension 1	-12.4	-21.4	7.5	28.9	7.5
	Dimension 2	2.1	-1.5	8.0	9.5	2.5
	Dimension 3	1.7	-2.4	8.8	11.2	3.5
	Dimension 4	-0.7	-3.9	1.8	5.7	1.6
	Dimension 5	-0.5	-3.5	6.0	9.5	2.5
	Dimension 6	-0.3	-3.3	3.6	6.9	2.2
Official documents	Dimension 1	-18.1	-26.3	-9.1	17.2	4.8
	Dimension 2	-2.9	-5.4	-1.5	3.9	1.2
	Dimension 3	7.3	2.1	13.4	11.3	3.6
	Dimension 4	-0.2	-8.4	8.7	17.1	4.1
	Dimension 5	4.7	0.6	9.4	8.8	2.4
	Dimension 6	-0.9	-3.8	2.7	6.5	2.0
Academic prose	Dimension 1	-14.9	-26.5	7.1	33.6	6.0
	Dimension 2	-2.6	-6.2	5.3	11.5	2.3
	Dimension 3	4.2	-5.8	18.6	24.3	3.6
	Dimension 4	-0.5	-7.1	17.5	24.6	4.7

Writing (LOB)	Dimension	Mean	Minimum value	Maximum value	Range	Standard deviation
	Dimension 5	5.5	-2.4	16.8	19.2	4.8
	Dimension 6	0.5	-3.3	9.2	12.5	2.7
General fiction	Dimension 1	-0.8	-19.6	22.3	41.9	9.2
	Dimension 2	5.9	1.2	15.6	14.3	3.2
	Dimension 3	-3.1	-8.2	1.0	9.2	2.3
	Dimension 4	0.9	-3.2	7.2	10.3	2.6
	Dimension 5	-2.5	-4.8	1.5	6.3	1.6
	Dimension 6	-1.6	-4.3	2.7	6.9	1.9
Mystery fiction	Dimension 1	-0.2	-15.4	12.6	28.0	8.5
	Dimension 2	6.0	0.7	10.3	9.7	3.0
	Dimension 3	-3.6	-7.2	4.8	12.0	3.4
	Dimension 4	-0.7	-5.6	4.2	9.7	3.3
	Dimension 5	-2.8	-4.5	-0.4	4.1	1.2
	Dimension 6	-1.9	-4.3	-0.2	4.1	1.3
Science fiction	Dimension 1	-6.1	-12.1	-1.7	10.4	4.6
	Dimension 2	5.9	2.4	8.7	6.3	2.5
	Dimension 3	-1.4	-6.0	3.8	9.8	3.7
	Dimension 4	-0.7	-3.0	1.8	4.8	1.7
	Dimension 5	-2.5	-3.6	-1.7	1.8	0.8
	Dimension 6	-1.6	-3.5	0.4	3.9	1.6
Adventure fiction	Dimension 1	0.0	-11.9	11.1	23.1	6.3
	Dimension 2	5.5	2.2	10.5	8.3	2.7
	Dimension 3	-3.8	-7.8	-1.6	6.2	1.7
	Dimension 4	-1.2	-5.0	5.6	10.6	2.8
	Dimension 5	-2.5	-4.5	-0.8	3.7	1.2
	Dimension 6	-1.9	-4.0	1.8	5.8	1.7
Romantic fiction	Dimension 1	4.3	-6.5	15.3	21.9	5.6
	Dimension 2	7.2	1.4	11.7	10.3	2.8
	Dimension 3	-4.1	-6.4	-1.2	5.2	1.6
	Dimension 4	1.8	-1.1	7.2	8.2	2.7
	Dimension 5	-3.1	-4.2	-1.5	2.7	0.9
	Dimension 6	-1.2	-3.8	2.1	5.9	2.2

Writing (LOB)	Dimension	Mean	Minimum value	Maximum value	Range	Standard deviation
Humor	Dimension 1	-7.8	-13.7	7.6	21.3	6.7
	Dimension 2	0.9	-2.0	3.0	5.0	1.8
	Dimension 3	-0.8	-3.5	4.2	7.7	2.6
	Dimension 4	-0.3	-4.8	3.8	8.6	2.7
	Dimension 5	-0.4	-3.0	1.2	4.2	1.4
	Dimension 6	-1.5	-3.6	1.3	4.8	1.7
Personal letters	Dimension 1	19.5	13.8	27.0	13.2	5.4
	Dimension 2	0.3	-0.9	1.7	2.6	1.0
	Dimension 3	-3.6	-6.6	-1.3	5.3	1.8
	Dimension 4	1.5	-1.6	6.4	8.0	2.6
	Dimension 5	-2.8	-4.8	0.5	5.4	1.9
	Dimension 6	-1.4	-3.7	0.3	4.0	1.6
Professional letters	Dimension 1	-3.9	-17.1	24.8	41.9	13.7
	Dimension 2	-2.2	-6.9	4.6	11.5	3.5
	Dimension 3	6.5	1.4	12.4	11.0	4.2
	Dimension 4	3.5	-5.3	11.0	16.3	4.7
	Dimension 5	0.4	-3.5	4.4	7.9	2.4
	Dimension 6	1.5	-3.6	9.6	13.2	3.6

Appendix IX. Computation of cluster affiliations

Table 1 shows the centroid scores of each cluster identified in Biber (1989, 1995) with respect to Biber's (1988) Dimensions 1 through 5 (C1 means cluster centroid 1, C2 cluster centroid 2, etc.). Tables 2–4 each present the Euclidean distances found between the texts and the cluster centroids (and those between the average dimension scores of the genre and the latter), with the resulting cluster affiliations indicated in the rightmost column. Table 5 presents the Euclidean distances found between the dimension scores of Collot's (1991) genre of BBS conferencing (i.e. the “ELC other” corpus of ACMC) and the cluster centroids. The polarity of all scores follows that in Biber (1988, 1989), rather than that in Biber (1995). See Appendix X for the dimension scores of the individual texts, and tables 5.1 and 5.5 (in chapter 5) for those of the genres.

Table 1: Cluster centroid scores (Biber 1995: 328–331)¹

Cluster centroids	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
C1	48.0	-1.0	-5.5	0.5	-4.0
C2	32.0	-0.5	-4.5	0.5	-3.0
C3	-17.0	-2.5	3.5	-2.0	9.5
C4	-20.0	-2.0	4.5	-3.0	2.0
C5	5.0	6.5	-3.5	1.0	-2.5
C6	-12.0	1.5	0.0	-1.0	-0.5
C7	0.0	-3.0	-13.5	-4.5	-3.5
C8	3.0	-2.0	2.0	4.0	-1.0

Table 2: Distances to cluster centroids of the IRC texts (and the IRC genre)

IRC text	Dist fr	Dist fr	Dist fr	Dist fr	Dist fr	Dist fr	Dist fr	Dist fr	MIN	CLUST
	C1	C2	C3	C4	C5	C6	C7	C8		
1a	31.4	16.4	39.0	40.0	17.9	31.4	18.1	19.9	16.4	2
1b	14.6	9.4	55.8	57.5	33.4	49.1	36.6	36.1	9.4	2
2a	21.2	8.1	48.1	49.5	26.2	41.1	28.7	27.7	8.1	2
2b	16.3	6.0	52.4	54.0	30.3	45.6	33.9	31.6	6.0	2
3a	33.6	17.7	35.2	36.2	12.8	27.3	17.6	14.1	12.8	5
3b	24.0	9.3	44.8	46.2	23.3	37.8	26.7	23.5	9.3	2
4a	24.8	10.4	44.5	45.8	22.7	37.3	25.0	24.3	10.4	2
4b	29.8	14.3	38.7	39.9	17.9	31.6	22.2	17.4	14.3	2
5a	19.0	5.1	48.6	50.3	26.6	42.0	32.4	27.4	5.1	2
5b	17.9	6.8	50.1	52.2	29.5	44.2	34.1	29.7	6.8	2
IRC genre	22.8	8.0	45.5	47.0	23.6	38.5	27.2	24.8	8.0	2

1 The polarity of all scores follows that in Biber (1988, 1989).

Table 3: Distances to cluster centroids of the split-window ICQ texts (and the splitwindow ICQ genre)

ICQ text	Dist fr C1	Dist fr C2	Dist fr C3	Dist fr C4	Dist fr C5	Dist fr C6	Dist fr C7	Dist fr C8	MIN	CLUST
1	19.3	34.7	84.1	86.7	62.1	78.5	67.4	63.8	19.3	1
2	13.0	28.9	79.4	81.6	56.6	73.2	61.6	58.3	13.0	1
3	10.0	25.5	75.7	77.9	53.1	69.6	58.3	54.9	10.0	1
4	5.9	11.0	61.8	63.8	38.7	55.2	44.3	40.3	5.9	1
5	3.0	15.6	66.1	68.3	43.2	59.7	48.0	45.4	3.0	1
6	9.1	8.4	58.5	60.4	36.0	52.0	41.4	37.3	8.4	2
7	4.7	20.1	70.7	72.7	47.9	64.2	52.6	49.6	4.7	1
8	13.7	29.4	80.1	82.2	56.5	73.5	62.0	59.2	13.7	1
9	15.7	3.3	52.0	53.8	29.4	45.2	34.8	30.4	3.3	2
10	5.7	13.0	63.7	65.6	40.6	56.9	45.4	42.0	5.7	1
11	4.8	11.7	62.4	64.4	39.5	55.9	44.8	41.3	4.8	1
12	29.0	13.3	39.3	40.4	17.3	31.9	23.1	17.7	13.3	2
ICQ genre	2.1	15.3	66.0	68.1	43.1	59.6	48.4	44.9	2.1	1

Table 4: Distances to cluster centroids of the SBC subset texts (and the SBC subset genre)

SBC text	Dist fr C1	Dist fr C2	Dist fr C3	Dist fr C4	Dist fr C5	Dist fr C6	Dist fr C7	Dist fr C8	MIN	CLUST
1	15.8	31.4	81.5	83.7	59.2	75.4	63.7	60.8	15.8	1
2	7.7	15.3	64.3	66.3	41.7	58.0	47.4	44.1	7.7	1
3	4.5	16.7	67.0	69.0	44.5	60.6	49.4	45.6	4.5	1
4	12.6	5.6	55.1	56.9	32.1	48.4	38.8	33.9	5.6	2
5	14.3	28.6	78.0	79.9	55.8	71.9	60.8	57.6	14.3	1
6	15.7	7.3	53.5	55.2	30.5	46.6	35.8	33.4	7.3	2
7	7.1	22.4	73.1	75.2	49.9	66.6	55.3	51.6	7.1	1
8	5.9	14.4	64.4	66.3	41.5	57.8	46.5	43.8	5.9	1
9	10.2	7.5	57.4	59.3	35.0	50.9	39.9	36.4	7.5	2
10	9.4	24.2	74.4	76.5	51.0	67.9	57.5	53.3	9.4	1
11	7.1	16.5	66.4	68.5	42.3	59.6	49.4	45.4	7.1	1
12	27.4	11.9	40.4	42.2	17.8	33.6	26.7	19.0	11.9	2
13	6.9	22.5	73.0	75.1	50.2	66.6	55.5	51.8	6.9	1
14	40.2	24.5	27.4	29.5	10.2	21.4	19.8	8.8	8.8	8
SBCgenre	5.7	12.1	62.4	64.4	39.5	55.9	45.3	41.4	5.7	1

Table 5: Distances to cluster centroids of the BBS conferencing genre (“ELC other” corpus, Collot 1991)

ACMC	Dist fr C1	Dist fr C2	Dist fr C3	Dist fr C4	Dist fr C5	Dist fr C6	Dist fr C7	Dist fr C8	MIN	CLUST
ELC other	25.1	11.6	42.9	45.9	23.6	38.0	30.7	23.2	11.6	2

Appendix X. Dimension scores for individual texts

Tables 1–3 present the dimension scores on Biber’s (1988) dimensions for the individual texts annotated in the present study.

Table 1: Dimension scores for the Internet relay chat texts (UCOW)

	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b
Dimension 1	17.3	35.9	27.7	32.6	14.4	24.6	23.9	18.6	29.7	31.6
Dimension 2	-4.2	-2.1	-5.0	-4.9	-1.9	-5.9	-4.6	-4.7	-3.0	-5.6
Dimension 3	-8.3	-6.9	-6.5	-4.7	-4.7	-4.7	-6.6	-2.2	-0.9	-1.5
Dimension 4	-4.2	-7.5	-3.9	-3.1	0.1	-0.1	-3.7	-0.7	-0.8	-2.2
Dimension 5	-4.5	-3.9	-4.8	-4.8	-3.9	-4.3	-4.8	-4.2	-3.2	-0.9
Dimension 6	-4.2	-3.8	-4.5	-2.8	-1.0	-3.8	-3.4	-3.3	-4.2	-3.5

Table 2: Dimension scores for the split-window ICQ texts (UCOW)

	1	2	3	4	5	6	7	8	9	10	11	12
Dimension 1	66.4	60.7	57.3	42.7	47.3	39.7	51.7	61.2	32.7	44.0	43.6	19.3
Dimension 2	-2.1	-3.1	-2.0	-1.9	-1.3	-2.8	-3.5	2.1	-3.2	-3.6	-1.7	-2.7
Dimension 3	-3.3	-5.0	-3.2	-4.4	-6.3	-2.5	-4.7	-5.4	-3.2	-6.1	-4.2	-1.5
Dimension 4	-0.9	1.4	-1.8	2.7	-1.1	-0.4	-0.8	-1.3	1.6	3.5	-0.3	-0.6
Dimension 5	1.3	-3.8	-3.1	-3.6	-1.8	-3.4	-4.6	-4.8	-3.5	-4.8	-3.3	-4.6
Dimension 6	-3.9	-2.4	-1.9	-0.1	-1.1	-2.4	-0.4	0.2	-3.4	-2.5	-2.3	-2.2

Table 3: Dimension scores for the SBC subset texts annotated in the present study

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Dimension 1	63.0	45.8	47.9	36.3	59.4	34.1	54.0	45.4	38.5	55.8	47.3	21.5	54.3	9.1
Dimension 2	-4.8	0.9	-4.9	0.2	-2.7	0.9	-2.2	-0.9	-3.5	2.7	5.7	1.2	-2.3	1.0
Dimension 3	-3.9	-1.9	-3.6	-1.0	-0.8	-3.5	-5.2	-3.9	-3.5	-1.7	-3.3	0.4	-3.3	2.1
Dimension 4	-2.7	-5.5	1.6	-0.3	-6.5	-6.3	4.0	-4.6	-1.4	0.6	0.8	1.5	0.3	-0.5
Dimension 5	-4.6	-2.7	-4.6	-3.0	-4.8	-3.7	-4.6	-4.1	-3.5	-4.1	-3.3	-1.3	-4.4	2.3
Dimension 6	-1.8	1.1	-0.6	-1.4	1.3	-2.4	2.2	-2.4	-1.7	6.9	0.4	-0.3	-3.1	-0.8

List of References

Corpora

- ELC other = Electronic Language Corpus other. 1991. Compiled by M. Collot.
- LLC = London-Lund Corpus of Spoken English. 1990. Compiled by J. Svartvik. <<http://clu.uni.no/icame/manuals/LONDLUND/INDEX.HTM>> (2015-10-13).
- LOB = Lancaster-Oslo/Bergen Corpus. 1978. Compiled by G. Leech, S. Johansson and K. Hofland. <<http://clu.uni.no/icame/manuals/LOB/INDEX.HTM>> (2015-10-13).
- LSWE = Longman Spoken and Written English Corpus. S.a. Pearson Longman. <<http://www.pearsonlongman.com/dictionaries/corpus/>> (2015-10-13).
- SBC = Santa Barbara Corpus of Spoken American English, part 1. 2000. Compiled by J. W. Du Bois, W. L. Chafe, L. Meyer and S. A. Thompson. <<http://www.linguistics.ucsb.edu/research/santa-barbara-corpus>> (2015-10-13).
- UCOW = Uppsala Conversational Writing Corpus. 2004. Compiled by E. Jonsson.

Works cited or referred to

- Aijmer, K. (2002): *English discourse particles: Evidence from a corpus*. Amsterdam: John Benjamins.
- Akinnaso, F. N. (1982): On the differences between spoken and written language. *Language and Speech* 25 (2): 97–125.
- Altenberg, B. (1984): Causal linking in spoken and written English. *Studia Linguistica* 38: 20–69.
- Andersen, G. (2006): Grammatical variation in English corpora: The case of modal verbs. In L. E. Breivik, S. Halverson & K. E. Haugland (eds.) *These things write I vnto thee...: Essays in honour of Bjørg Bækken*, 11–27. Oslo: Novus Press.
- Anderson, J. F., F. K. Beard & J. B. Walther (2010): Turn-taking and the local management of conversation in a highly simultaneous computer-mediated communication system. *Language@Internet* 7, article 7.
- Anglemark, L. (2009): Address terms in computer mediated communication: Email, chat and weblogs. Unpublished PhD dissertation. Department of English, Uppsala University, Uppsala.
- Atkinson, D. (2001): Scientific discourse across history: A combined multi-dimensional/rhetorical analysis of the *Philosophical Transactions of the Royal*

- Society of London*. In S. Conrad & D. Biber (eds.) *Variation in English: Multi-dimensional studies*, 45–65. Harlow: Longman.
- Atkinson, D. & D. Biber (1994): Register: A review of empirical research. In D. Biber & E. Finegan (eds.) *Sociolinguistic perspectives on register*, 351–371. New York: Oxford University Press.
- Babineaux, L. (forthcoming; cited in Condon & Čech 2001:8): Simultaneous transmission and discourse planning in computer-mediated communication. Forthcoming Master's thesis, University of Louisiana at Lafayette, Lafayette, LA.
- Bachman-Mann, M. (1944): The quantitative differentiation of samples of written language. *Psychological Monographs* 56 (2): 39–74.
- Baron, N. S. (1981): *Speech, writing and sign*. Bloomington, IN: Indiana University Press.
- Baron, N. S. (1984): Computer-mediated communication as a force in language change. *Visible Language* XVIII (2): 118–141.
- Baron, N. S. (1998): Letters by phone or speech by other means: The linguistics of email. *Language & Communication* 18: 133–170.
- Baron, N. S. (2000): *Alphabet to email: How written English evolved and where it's heading*. London: Routledge.
- Baron, N. S. (2004): See you online: Gender issues in college student use of instant messaging. *Journal of Language and Social Psychology* 23: 397–423.
- Baron, N. S. (2008): *Always on: Language in an online and mobile world*. Oxford: Oxford University Press.
- Baron, N. S. (2010): Discourse structures in instant messaging: The case of utterance breaks. *Language@Internet* 7, article 4.
- Baym, N. (1996): Agreements and disagreements in a computer-mediated discussion. *Research on Language and Social Interaction* 29 (4): 315–346.
- Beaman, K. (1984): Coordination and subordination revisited: Syntactic complexity in spoken and written narrative discourse. In D. Tannen (ed.) *Coherence in spoken and written discourse*, 45–80. Norwood, NJ: Ablex.
- Bennett, D. (2007): Teen slang now following Internet shortcuts. *Canadian Press* 2007-02-08. <http://www.ctv.ca/CTVNews/Entertainment/20070208/teen_slang_070208/> (2015-10-13).
- Bernstein, B. (1964): Elaborated and restricted codes: Their social origins and some consequences. In J. J. Gumperz & D. Hymes (eds.) *The ethnography of communication*, special issue of *American Anthropologist* 66 (6) part 2: 59–69.
- Bernstein, B. (1970): *Class, codes and control. Volume 1: Theoretical studies towards a sociology of language*. London: Routledge & Kegan Paul.

- Besnier, N. (1988): The linguistic relationships of spoken and written Nukulaelae registers. *Language* 64 (4): 707–736.
- Besnier, N. (1991): Literacy and the notion of person on Nukulaelae atoll. *American Anthropologist* 93 (3): 570–587.
- Besnier, N. (1995): *Literacy, emotion, and authority: Reading and writing on a Polynesian atoll*. Cambridge, UK: Cambridge University Press.
- Biber, D. (1984): A model of textual relations within the written and spoken modes. Unpublished PhD dissertation. University of Southern California, Los Angeles, CA.
- Biber, D. (1985): Investigating macroscopic textual variation through multifeature/multidimensional analyses. *Linguistics* 23: 337–360.
- Biber, D. (1986a): Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language* 62: 384–414.
- Biber, D. (1986b): On the investigation of spoken/written differences. *Studia Linguistica* 40: 1–21.
- Biber, D. (1988): *Variation across speech and writing*. Cambridge, UK: Cambridge University Press.
- Biber, D. (1989): A typology of English texts. *Linguistics* 27 (1): 3–43.
- Biber, D. (1990): Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing* 5: 257–269.
- Biber, D. (1991): Oral and literate characteristics of selected primary school reading materials. *Text* 11: 73–96.
- Biber, D. (1993): Representativeness in corpus design. *Literary and Linguistic Computing* 8: 1–15.
- Biber, D. (1995): *Dimensions of register variation: A cross-linguistic comparison*. Cambridge, UK: Cambridge University Press.
- Biber, D. (2003): Variation among university spoken and written registers: A new multi-dimensional analysis. In P. Leistyna & C. Meyer (eds.) *Corpus analysis: Language structure and language use*, 47–70. Amsterdam: Rodopi.
- Biber, D. (2004): Modal use across registers and time. In A. Curzan & K. Emmons (eds.) *Studies in the history of the English language II: Unfolding conversations*, 189–216. Berlin: Mouton de Gruyter.
- Biber, D. (2006): *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, D. (2008): Multi-dimensional approaches. In A. Lüdeling & M. Kytö (eds.) *Corpus linguistics: An international handbook*, vol. 2, 822–855. Berlin: Walter de Gruyter.

- Biber, D. & E. Finegan (1989): Drift and the evolution of English style: A history of three genres. *Language* 65: 487–517.
- Biber, D. & E. Finegan (1991): On the exploitation of computerized corpora in variation studies. In K. Aimer & B. Altenberg (eds.) *English corpus linguistics: Studies in honour of Jan Svartvik*, 204–220. London: Longman.
- Biber, D. & E. Finegan (1992): The linguistic evolution of five written and speech-based English genres from the 17th to the 20th centuries. In M. Rissanen, O. Ihalainen, T. Nevalainen & I. Taavitsainen (eds.) *History of Englishes: New methods and interpretations in historical linguistics*, 688–704. Berlin: Mouton de Gruyter.
- Biber, D., S. Conrad & R. Reppen (1998): *Corpus linguistics: Investigating language structure and use*. Cambridge, UK: Cambridge University Press.
- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999): *Longman grammar of spoken and written English*. Harlow: Pearson Education.
- Biber, D. & E. Finegan (2001): Diachronic relations among speech-based and written registers in English. In S. Conrad & D. Biber (eds.) *Variation in English: Multi-dimensional studies*, 66–83. Harlow: Pearson Education.
- Biber, D., M. Davies, J. K. Jones & N. Tracy-Ventura (2006): Spoken and written register variation in Spanish: A multi-dimensional analysis. *Corpora* 1 (1): 1–37.
- Biber, D. & S. Conrad (2009): *Register, genre, and style*. Cambridge, UK: Cambridge University Press.
- Blankenship, J. (1962): A linguistic analysis of oral and written style. *Quarterly Journal of Speech* 48 (4): 419–422.
- Blankenship, J. (1974): The influence of mode, sub-mode, and speaker predilection on style. *Speech Monographs* 41 (2): 85–118.
- Blashki, K. & S. Nichol (2005): Game geek's goss: Linguistic creativity in young males within an online university forum. *Australian Journal of Emerging Technologies and Society* 3 (2): 77–86.
- Bloomfield, L. (1933): *Language*. New York: Holt, Rinehart and Winston.
- Brown, P. & C. Fraser (1979): Speech as a marker of situation. In K. R. Scherer & H. Giles (eds.) *Social markers in speech*, 33–62. Cambridge, UK: Cambridge University Press.
- Bybee & S. Fleischman (eds.) (1995a): *Modality in grammar and discourse*. Amsterdam: John Benjamins.
- Bybee, J. & S. Fleischman (1995b): Modality in grammar and discourse: An introductory essay. In J. Bybee & S. Fleischman (eds.) *Modality in grammar and discourse*, 1–14. Amsterdam: John Benjamins.

- Cameron, D. (2001): *Working with spoken discourse*. London: Sage Publications.
- Carroll, J. B. (1960): Vectors of prose style. In T. A. Sebeok (ed.) *Style in language*, 283–292. Cambridge, MA: MIT Press.
- Castello, E. (2008): *Text complexity and reading comprehension tests*. Bern: Peter Lang.
- Chafe, W. (1982): Integration and involvement in speaking, writing, and oral literature. In D. Tannen (ed.) *Spoken and written language: Exploring orality and literacy*, 35–54. Norwood, NJ: Ablex.
- Chafe, W. (1985): Linguistic differences produced by differences between speaking and writing. In D. R. Olson, N. Torrance & A. Hildyard (eds.) *Literacy, language, and learning: The nature and consequences of reading and writing*, 105–123. London: Cambridge University Press.
- Chafe, W. (1994): *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Chafe, W. & J. Danielewicz (1987): Properties of spoken and written language. In R. Horowitz & S. J. Samuels (eds.) *Comprehending oral and written language*, 83–113. San Diego: Academic Press.
- Chafe, W. & D. Tannen (1987): The relation between written and spoken language. *Annual Review of Anthropology* 16: 383–407.
- Cherny, L. (1994): Gender differences in text-based virtual reality. In M. Bucholtz (ed.) *Cultural performances: Proceedings of the Berkeley Conference on Women and Language*, 102–115. Berkeley, CA: Berkeley Women and Language Group, University of California.
- Cherny, L. (1999): *Conversation and community: Chat in a virtual world*. Stanford, CA: CSLI Publications.
- Cho, T. (2010): Linguistic features of electronic mail in the workplace: A comparison with memoranda. *Language@Internet* 7, article 3.
- Chomsky, N. (1964): *Current issues in linguistic theory*. The Hague: Mouton & Co.
- Chomsky, N. (1965): *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chotlos, J. W. (1944): A statistical and comparative analysis of individual written language samples. *Psychological Monographs* 56: 75–111.
- Coates, J. (1983): *The semantics of the modal auxiliaries*. London: Croom Helm.
- Coates, J. (1993): *Women, men and language*, 2. ed. Essex, England: Addison Wesley Longman.
- Coates, J. (1995): The expression of root and epistemic possibility in English. In J. Bybee & S. Fleischman (eds.) *Modality in grammar and discourse*, 55–66. Amsterdam: John Benjamins.

- Coleman, J. (1996): *Public reading and the reading public in late medieval England and France*. New York: Cambridge University Press.
- Collot, M. (1991): The situational features and textual dimensions of electronic language. Unpublished Master's thesis. Concordia University, Montreal.
- Collot, M. & N. Belmore (1996): Electronic language: A new variety of English. In S. C. Herring (ed.) *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*, 13–28. Amsterdam: John Benjamins.
- Condon, S. L. & C. G. Čech (2001): Profiling turns in interaction: Discourse structure and function. *Proceedings of the 35th Hawaii International Conference on System Sciences (HICSS-35)*, 1–10. Los Alamitos, CA: IEEE Press.
- Condon, S. L. & C. G. Čech (2010): Discourse management in three modalities. *Language@Internet* 7, article 6.
- Connor, U. & T. Upton (2003): Linguistic dimensions of direct mail letters. In P. Leistyna & C. Meyer (eds.) *Corpus analysis: Language structure and language use*, 71–86. Amsterdam: Rodopi.
- Conrad, S. (2001): Variation among disciplinary texts: A comparison of textbooks and journal articles in biology and history. In S. Conrad & D. Biber (eds.) *Variation in English: Multi-dimensional studies*, 94–107. Harlow: Longman.
- Conrad, S. & D. Biber (eds.) (2001a): *Variation in English: Multi-dimensional studies*. Harlow: Longman.
- Conrad, S. & D. Biber (2001b): Multi-dimensional methodology and the dimensions of register variation in English. In Conrad, S. & D. Biber (eds.) *Variation in English: Multi-dimensional studies*, 13–42. Harlow: Longman.
- Crystal, D. (2001): *Language and the Internet*. Cambridge, UK: Cambridge University Press.
- Crystal, D. (2004a): *A glossary of netspeak and textspeak*. Edinburgh, UK: Edinburgh University Press.
- Crystal, D. (2004b): *The language revolution*. Cambridge, UK: Polity Press.
- Crystal, D. (2008a): *Txtng: The gr8 db8*. Oxford, UK: Oxford University Press.
- Crystal, D. (2008b): *A dictionary of linguistics and phonetics*, 6. ed. Oxford, UK: Blackwell.
- Crystal, D. (2011a): *Internet linguistics: A student guide*. Abingdon, Oxon: Routledge.
- Crystal, D. & D. Davy (1969): *Investigating English style*. New York: Longman.
- Danet, B., L. Ruedenberg & Y. Rosenbaum-Tamari (1998): Hmmm...Where's that smoke coming from?: Writing, play and performance on Internet Relay Chat. In F. Sudweeks, M. McLaughlin & S. Rafaeli (eds.) *Network and netplay: Virtual groups on the Internet*, 41–76. Cambridge, MA: MIT Press.

- Davis, B. & J. Brewer (1997): *Electronic discourse: Linguistic individuals in virtual space*. Albany: State University of New York Press.
- De Argaez, E. (2015): *Internet world stats: Usage and population statistics* <<http://www.internetworldstats.com/stats.htm>> (2015-10-13).
- DeCoursy, L. (2001): ICQ celebrates 100 million registered users. *Time Warner press release* 2001-05-09 <<http://www.timewarner.com/newsroom/press-releases/2001/05/09/icq-celebrates-100-million-registered-users>> (2015-10-13).
- Derks, D., A. E. R. Bos & J. von Grumbkow (2007): Emoticons and social interaction on the Internet: The importance of social context. *Computers in Human Behavior* 23: 842–849.
- Dery, M. (1993): Flame wars. In M. Dery (ed.) *Flame wars: The discourse of cyberculture*, 1–10. Durham, NC: Duke University Press.
- DeVito, J. A. (1964): A quantitative analysis of comprehension factors in samples of oral and written technical discourse of skilled communicators. Unpublished PhD thesis. University of Illinois, Urbana, IL.
- DeVito, J. A. (1965): Comprehension factors in oral and written discourse of skilled communicators. *Speech Monographs* 32: 124–128.
- DeVito, J. A. (1966): Psychogrammatical factors in oral and written discourse by skilled communicators. *Speech Monographs* 33: 73–76.
- DeVito, J. A. (1967a): Levels of abstraction in spoken and written language. *Journal of Communication* 17: 354–361.
- DeVito, J. A. (1967b): A linguistic analysis of spoken and written language. *Central States Speech Journal* 18: 81–85.
- DeVito, J. A. (1970): *The psychology of speech and language: An introduction to psycholinguistics*. New York: Random House.
- Diani, G. (2004): The discourse function of *I don't know* in English conversation. In K. Aijmer & A-B. Stenström (eds.) *Discourse patterns in spoken and written corpora*, 157–171. Amsterdam: John Benjamins.
- Dresner, E. (2005): The topology of auditory and visual perception, linguistic communication, and interactive written discourse. *Language@Internet* 2, article 2.
- Dresner, E. & S. C. Herring (2010): Functions of the non-verbal in CMC: Emoticons and illocutionary force. *Communication Theory* 20: 249–268.
- Drieman, G. H. J. (1962): Differences between written and spoken language: An exploratory study. *Acta Psychologica* 20: 36–57, 78–100.
- Du Bois, J. W., W. L. Chafe, C. Meyer & S. A. Thompson (2000): *Santa Barbara Corpus of Spoken American English, part 1*. Philadelphia: Linguistic Data Consortium. ISBN 1-58563-164-7.

- Elbow, P. (1991): Reflections on academic discourse. *College English* 53 (2): 135–155.
- Elsner, M. & E. Charniak (2008): You talking to me? A corpus and algorithm for conversation disentanglement. In *Proceedings of ACL-08: HLT*, 834–842. <<http://www.cs.brown.edu/~melsner/chat.pdf>> (2015-10-13).
- Elsner, M. & E. Charniak (2010): Disentangling chat. *Computational Linguistics* 36 (3): 389–409.
- Ervin-Tripp, S. (1972): On sociolinguistic rules: Alternation and co-occurrence. In J. J. Gumperz & D. Hymes (eds.) *Directions in Sociolinguistics: The ethnography of communication*, 213–250. New York: Holt, Rinehart and Winston.
- Fairbanks, H. (1944): The quantitative differentiation of samples of spoken language. *Psychological Monographs* 56: 47–74.
- Ferguson, C. A. (1959): Diglossia. *Word* 15: 325–340.
- Ferguson, C. A. (1994): Dialect, register, and genre: Working assumptions about conventionalization. In D. Biber & E. Finegan (eds.) *Sociolinguistic perspectives on register*, 15–30. New York: Oxford University Press.
- Ferrara, K., H. Brunner & G. Whittemore (1991): Interactive written discourse as an emergent register. *Written Communication* 8 (1): 8–34.
- Finegan, E. & D. Biber (1986): Toward a unified model of sociolinguistic prestige. In D. Sankoff (ed.) *Diversity and diachrony*, 391–398. Amsterdam: John Benjamins.
- Flesch, R. (1948): A new readability yardstick. *Journal of Applied Psychology* 32 (3): 221–233.
- Foertsch, J. (1995): The impact of electronic networks on scholarly communication: Avenues to research. *Discourse Processes* 19: 301–328.
- Forsyth, E. N. (2007): Improving automated lexical and discourse analysis of online chat dialog. Unpublished Master's thesis. Naval Postgraduate School, Monterey, CA.
- Forsyth, E. N. & C. H. Martell (2007): Lexical and discourse analysis of online chat dialog. *Proceedings of the International Conference on Semantic Computing 2007*: 19–26.
- Fossum, E. (1944): An analysis of the dynamic vocabulary of junior college students. *Speech Monographs* 11 (1): 88–96.
- Fowler, R. & G. Kress (1979): Critical linguistics. In R. Fowler, B. Hodge, G. Kress & T. Trew (eds.) *Language and control*, 185–213. London: Routledge.
- Freiermuth, M. R. (2003): *Features of electronic synchronous communication: A comparative analysis of online chat, spoken and written texts*. PhD dissertation. Oklahoma State University, Stillwater, OK.

- Gajadhar J. & J. Green (2003): An analysis of nonverbal communication in an online chat group. Open Polytechnic of New Zealand, working paper, March 2003.
- Geertz, C. (1988): *Works and lives: The anthropologist as author*. Cambridge, UK: Polity.
- Geisler, C. (2002): Investigating register variation in nineteenth-century English: A multi-dimensional comparison. In R. Reppen, S. M. Fitzmaurice & D. Biber (eds.) *Using corpora to explore linguistic variation*, 249–271. Amsterdam: John Benjamins.
- Georgakopoulou, A. (2004): To tell or not to tell? Email stories between on- and offline interactions. *Language@Internet* 1, article 1.
- Georgakopoulou, A. (2011a): Computer-mediated communication. In J-O. Östman & J. Verschueren (eds.) *Pragmatics in practice*, 93–110. Amsterdam: John Benjamins.
- Georgakopoulou, A. (2011b): “On for drinkies?”: Email cues of participant alignments. *Language@Internet* 8, article 4.
- Gibson, J. W., C. R. Gruner, R. J. Kibler & F. J. Kelly (1966): A quantitative examination of differences and similarities in written and spoken messages. *Speech Monographs* 33 (4): 444–451.
- Giese, M. (1998): Self without body: Textual self-representation in an electronic community. *First Monday* 3 (4) < <http://firstmonday.org/ojs/index.php/fm/article/view/587/508> > (2015-10-13).
- Goody, J. (1987): *The interface between the written and the oral*. Cambridge UK: Cambridge University Press.
- Greenbaum, S. & J. Svartvik (1990): The London-Lund Corpus of Spoken English. In Svartvik, J. (ed.) *The London-Lund Corpus of Spoken English: Description and research* (Lund Studies in English 82). Lund: Lund University Press.
- Greenfield, P. M. (1972): Oral and written language: The consequence for cognitive development in Africa, the United States and England. *Language and Speech* 15: 169–178.
- Gromov, G. (1995): *Roads and crossroads of the Internet history*. <http://www.netvalley.com/internet/history_of_internet.pdf> (2015-10-13).
- Gruner, C. R., R. J. Kibler & J. W. Gibson (1967): A quantitative analysis of selected characteristics of oral and written vocabularies. *Journal of Communication* 17 (2): 152–158.
- Gumperz, J. J. (1965): Language. *Biennial Review of Anthropology* 4: 84–120.

- Gumperz, J. J., H. Kaltman & M. C. O'Connor (1984): Cohesion in spoken and written discourse. In D. Tannen (ed.) *Coherence in spoken and written discourse*, 3–20. Norwood, NJ: Ablex.
- Guo, J. (1995): The interactional basis of the Mandarin modal néng 'CAN'. In J. Bybee & S. Fleischman (eds.) *Modality in grammar and discourse*, 205–238. Amsterdam: John Benjamins.
- Hafner, K. & M. Lyon (1996): *Where wizards stay up late: The origins of the Internet*. New York: Simon & Schuster.
- Halliday, M. A. K. (1978): *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.
- Halliday, M. A. K. (1979): Differences between spoken and written language: Some implications for literacy teaching. In G. Page, J. Elkins & B. O'Connor (eds.) *Communication through reading. Proceedings of the Fourth Australian Reading Conference*, 37–52. Adelaide, SA: Australian Reading Association.
- Halliday, M. A. K. (1985a): *Spoken and written language*. Oxford: Oxford University Press.
- Halliday, M. A. K. (1985b): *An introduction to functional grammar*, 1. ed. London: Arnold.
- Halliday, M. A. K. (1987): Spoken and written modes of meaning. In R. Horowitz & S. J. Samuels (eds.) *Comprehending oral and written language*, 55–82. San Diego: Academic Press.
- Halliday, M. A. K. (2004): *An introduction to functional grammar*, 3. ed. London: Arnold.
- Halliday, M. A. K. & R. Hasan (1976): *Cohesion in English*. London: Longman.
- Halliday, M. A. K. & R. Hasan (1989): *Language, context, and text: Aspects of language in a social-semiotic perspective*, 2. ed. Oxford: Oxford University Press.
- Hård af Segerstad, Y. (2002): *Use and adaptation of written language to the conditions of computer-mediated communication*. PhD dissertation. Department of English, Göteborg University, Gothenburg.
- Helt, M. (2001): A comparison of British and American spoken English. In S. Conrad & D. Biber (eds.) *Variation in English: Multi-dimensional studies*, 171–183. Harlow: Longman.
- Herring, S. C. (ed.) (1996a): *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*. Amsterdam: John Benjamins.
- Herring, S. C. (1996b): Two variants of an electronic message schema. In S. C. Herring (ed.) *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*, 81–106. Amsterdam: John Benjamins.

- Herring, S. C. (1998): Virtual gender performances. Talk presented at Texas A&M University, September 25, 1998.
- Herring, S. C. (1999): Interactional coherence in CMC. *Journal of Computer-Mediated Communication* 4 (4).
- Herring, S. C. (2001): Computer-mediated discourse. In D. Schiffrin, D. Tannen & H. E. Hamilton (eds.) *The handbook of discourse analysis*, 612–634. Oxford: Blackwell Publishers.
- Herring, S. C. (2002): Computer-mediated communication on the Internet. *Annual Review of Information Science and Technology* 36: 109–168.
- Herring, S. C. (2003): Gender and power in online communication. In J. Holmes & M. Meyerhoff (eds.) *The handbook of language and gender*, 202–228. Oxford: Blackwell.
- Herring, S. C. (2004a): Computer-mediated discourse analysis: An approach to researching online behavior. In S. A. Barab, R. Kling & J. H. Gray (eds.) *Designing for virtual communities in the service of learning*, 338–376. New York: Cambridge University Press.
- Herring, S. C. (2004b): Slouching toward the ordinary: Current trends in computer-mediated communication. *New Media & Society* 6 (1): 26–36.
- Herring, S. C. (2007): A faceted classification scheme for computer-mediated discourse. *Language@Internet* 4, article 1.
- Herring, S. C. (2011a): Commentary. In C. Thurlow & K. Mroczek (eds.) *Digital discourse: Language in the new media*, 340–347. New York: Oxford University Press.
- Herring, S. C. (2011b): Computer-mediated conversation. Part II: Introduction and overview. *Language@Internet* 8, article 2.
- Herring, S. C. (2013a): Discourse in Web 2.0: Familiar, reconfigured, and emergent. In D. Tannen & A. M. Trester (eds.) *Discourse 2.0: Language and new media*, 1–25. Washington, DC: Georgetown University Press.
- Herring, S. C. (2013b): Relevance in computer-mediated conversation. In S. C. Herring, D. Stein & T. Virtanen (eds.) *Handbook of pragmatics of computer-mediated communication*, 245–268. Berlin: Walter de Gruyter.
- Herring, S. C. (2013c): Grammar and electronic communication. In C. A. Chapelle (ed.) *Encyclopedia of applied linguistics*, 2338–2346. Hoboken, NJ: Wiley-Blackwell.
- Herring, S. C., D. O. Kutz, J. C. Paolillo & A. Zelenkauskaitė (2009): Fast talking, fast shooting: Text chat in an online first-person game. *Proceedings of the Forty-Second Hawai'i International Conference on System Sciences (HICSS-42)*, 1–10. Los Alamitos, CA: IEEE Press.

- Herring, S. C., D. Stein & T. Virtanen (eds.) (2013): *Handbook of pragmatics of computer-mediated communication*. Berlin: Walter de Gruyter.
- Hodge, R. & G. Kress (1988): *Social semiotics*. Cambridge, UK: Polity.
- Horn, E. H. (1926): A basic writing vocabulary: 10,000 words most commonly used in writing. *University of Iowa Monographs in Education* 4 (first series), Iowa City.
- Horowitz, M. W. & A. Berkowitz (1964): Structural advantage of the mechanism of spoken expression as a factor in differences in spoken and written expression. *Perceptual and Motor Skills* 19 (2): 619–625.
- Horowitz, M. W. & J. B. Newman (1964): Spoken and written expression: An experimental analysis. *Journal of Abnormal and Social Psychology* 68 (6): 640–647.
- Horowitz, R. & S. J. Samuels (eds.) (1987): Comprehending oral and written language: Critical contrasts for literacy and schooling. In R. Horowitz & S. J. Samuels (eds.) *Comprehending oral and written language*, 1–52. San Diego: Academic Press.
- Hughes, R. (1996): *English in speech and writing: Investigating language and literature*. London: Routledge.
- Hunt, K. W. (1964): *Differences in grammatical structures written at three grade levels, the structures to be analyzed by transformational methods*. Tallahassee: Florida State University.
- Hymes, D. H. (ed.) (1964): *Language in culture and society: A reader in linguistics and anthropology*. New York: Harper & Row.
- Hymes, D. H. (1972): On communicative competence. In J. B. Pride & J. Holmes (eds.) *Sociolinguistics*, 269–293. Baltimore: Penguin Books.
- Hymes, D. H. (1974): *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania.
- Irvine, J. (1979): Formality and informality in communicative events. In J. Baugh & J. Sherzer (eds.) *Language in use: Readings in sociolinguistics*, 211–228. Englewood Cliffs, NJ: Prentice-Hall.
- Johansson, S., G. Leech & H. Goodluck (1978): *Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English for use with digital computers*. Department of English, University of Oslo, Oslo <<http://clu.uni.no/icame/manuals/LOB/INDEX.HTM>> (2015-10-15).
- Johns, A. M. (1997): *Text, role and context*. Cambridge, UK: Cambridge University Press.

- Johnson, W. (1939): *Language and speech hygiene: An application of general semantics* (General Semantics Monographs I). Institute of General Semantics, Chicago.
- Johnson, W. (1944): Studies in language behavior: I. A program of research. In *Psychological Monographs* 56 (2): 1–15.
- Jonsson, E. (1998): Electronic discourse: On speech and writing on the Internet. Unpublished Master's thesis. Luleå University of Technology, Luleå.
- Kennedy, G. (2002): Variation in the distribution of modal verbs. In R. Reppen, S. M. Fitzmaurice & D. Biber (eds.) *Using corpora to explore linguistic variation*, 73–90. Amsterdam: John Benjamins.
- Kerbrat-Orrechioni, C. (2004): Introducing polylogue. *Journal of Pragmatics* 36 (1): 1–24.
- Ko, K. (1994): Structural characteristics of electronic discourse: A comparative analysis of InterChange text. Unpublished PhD dissertation. Ball State University, Muncie, Indiana.
- Ko, K. (1996): Structural characteristics of computer-mediated language: A comparative analysis of InterChange discourse. *Electronic Journal of Communication/La Revue Electronique de Communication* 6 (3) <<http://www.cios.org/ejpublic/006/3/006315.html>> (2015-10-13).
- Korsgaard Sorensen, E. (1993): Dialogues in networks. In P. Bøgh Andersen, B. Holmqvist & J. F. Jensen (eds.) *The computer as medium*, 389–421. Cambridge, UK: Cambridge University Press.
- Kroll, B. (1977): Combining ideas in written and spoken English: A look at subordination and coordination. In E. Ochs Keenan & T. L. Bennett (eds.) *Discourse across time and space*, 69–108. Los Angeles: University of Southern California.
- Kytö, M. (2000): Robert Keayne's *Notebooks*: A verbatim record of spoken English in early Boston? In S. C. Herring, P. van Reenen & L. Schøsler (eds.) *Textual parameters in older languages*, 273–308. Amsterdam: John Benjamins.
- Labov, W. (1969): Contraction, deletion, and inherent variability of the English copula. *Language* 45: 715–762.
- Labov, W. (1972a): *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1972b): *Language in the inner city: Studies in the black English vernacular*. Philadelphia: University of Pennsylvania Press.
- Lakoff, R. T. (1982): Some of my favorite writers are literate: The mingling of oral and literate strategies in written communication. In D. Tannen (ed.) *Spoken and written language: Exploring orality and literacy*, 239–260. Norwood, NJ: Ablex.

- LeBlanc, T. R. (2005): "Is there a translator in teh house?": *Cultural and discourse analysis of a virtual speech community on an Internet message board*. Master's thesis. Louisiana State University, Baton Rouge, LA <http://etd.lsu.edu/docs/available/etd-04072005-145922/unrestricted/LeBlanc_thesis.pdf> (2015-10-13).
- Lee, C. K. M. (2011): Micro-blogging and status updates on Facebook: Texts and practices. In C. Thurlow & K. Mroczek (eds.) *Digital discourse: Language in the new media*, 110–128. New York: Oxford University Press.
- Lee, D. Y. W. (2001): Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language Learning and Technology* 5 (3): 37–72.
- Lee, D. Y. W. (forthcoming): Modelling variation in spoken and written English: The multi-dimensional approach revisited.
- Lin, J. (2007): *Automatic author profiling of online chat logs*. Master's thesis. Naval Postgraduate School, Monterey, CA <http://faculty.nps.edu/cmartell/07Mar_Lin.pdf> (2015-10-13).
- Ling, R. (2005): The sociolinguistics of SMS: An analysis of SMS use by a random sample of Norwegians. In R. Ling & P. Pedersen (eds.) *Mobile communications: Renegotiation of the social sphere*, 335–349. London: Springer.
- Ling, R. & N. S. Baron (2007): Text messaging and IM: A linguistic comparison of American college data. *Journal of Language and Social Psychology* 26 (3): 291–298.
- Ling, R. & N. S. Baron (2013): Mobile phone communication. In S. C. Herring, D. Stein & T. Virtanen (eds.) *Handbook of pragmatics of computer-mediated communication*, 191–215. Berlin: Walter de Gruyter
- Liu, G. Z. (1999): Virtual community presence in Internet relay chatting. *Journal of Computer-Mediated Communication*, 5 (1).
- Love, A. M. (2002): Introductory concepts and "cutting edge" theories: Can the genre of the textbook accomodate both? In J. Flowerdew (ed.) *Academic discourse*, 76–91. Harlow: Longman.
- Mann, C. & F. Stewart (2000): *Internet communication and qualitative research: A handbook for researching online*. London: Sage Publications.
- Mar, J. (2000): Online on time: The language of Internet relay chat. In D. Gibbs & K-L. Krause (eds.) *Cyberlines: Languages and cultures of the Internet*, 151–174. Albert Park, Australia: James Nicholas Publishers.
- Marckworth, M. L. & W. J. Baker (1974): A discriminant function analysis of co-variation of a number of syntactic devices in five prose genres. *American Journal of Computational Linguistics*, Microfiche 11.

- Marcus, M., B. Santorini & M. Marcinkiewicz (1993): Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19 (2): 313–330.
- Martin, J. R. (1985): *Factual writing: Exploring and challenging social reality*. Geelong, VIC: Deakin University Press.
- Martin, J. R. (1992): *English text: System and structure*. Amsterdam: John Benjamins.
- Martin, J. R. (2001a): Cohesion and texture. In D. Schiffrin, D. Tannen & H. E. Hamilton (eds.) *The handbook of discourse analysis*, 35–53. Oxford: Blackwell.
- Martin, J. R. (2001b): Language, register and genre. In A. Burns & C. Coffin (eds.) *Analysing English in a global context: A reader*, 149–166. London: Routledge.
- Marvin, L. E. (1995): Spoof, spam, lurk and lag: The aesthetics of text-based virtual realities. *Journal of Computer-Mediated Communication* 1 (2).
- Maynor, N. (1994): The language of electronic mail: Written speech? In G. D. Little & M. Montgomery (eds.) *Centennial usage studies*, 48–54. Tuscaloosa: University of Alabama Press.
- McGrath, J. E. (1990): Time matters in groups. In J. Galegher, R. E. Kraut & C. Egido (eds.) *Intellectual teamwork: Social and technical foundations of cooperative work*, 23–61. Hillsdale, NJ: Lawrence Erlbaum.
- Morgan, J. (2011): Why did LOL infiltrate the language? *BBC News* 2011-04-08. <<http://www.bbc.co.uk/news/magazine-12893416>> (2015-10-13).
- Morrisett, L. (1996): Habits of mind and a new technology of freedom. *First Monday* 1 (3) <<http://firstmonday.org/ojs/index.php/fm/article/view/483/404>> (2015-10-13).
- Murray, D. E. (1985): Composition as conversation: The computer terminal as a medium for communication. In L. Odell & D. Goswami (eds.) *Writing in nonacademic settings*, 203–227. New York: Guilford.
- Murray, D. E. (1988): The context of oral and written language: A framework for mode and medium switching. *Language in Society* 17 (3): 351–373.
- Newson, L. (2011): Multimodal creativity and identities of expertise in the digital ecology of a World of Warcraft guild. In C. Thurlow & K. Mroczek (eds.) *Digital discourse: Language in the new media*, 131–153. New York: Oxford University Press.
- Nichol, S. & K. Blashki (2006): The realm of the game geek: Supporting creativity in an online community. *IADIS Conference on Web Based Communities 2006* <http://www.academia.edu/1426133/The_realm_of_the_game_geek_supporting_creativity_in_an_online_community> (2015-10-13).

- Nilsson, S. (2003): A brief overview of the linguistic attributes of the blogosphere. Unpublished Master's thesis. Umeå University, Umeå.
- Nuckolls, K. (2005): IM communicating: A conversational analysis of instant message conversations. Unpublished Master's thesis. Portland State University, Portland, Oregon.
- Ochs, E. (1979): Planned and unplanned discourse. In T. Givón (ed.) *Discourse and syntax*, 219–244. Norwood, NJ: Ablex.
- O'Donnell, R. (1974): Syntactic differences between speech and writing. *American Speech* 49 (1/2): 102–110.
- O'Donnell, R. C., W. J. Griffin & R. C. Norris (1967): A transformational analysis of oral and written grammatical structures in the language of children in grades three, five, and seven. *The Journal of Educational Research* 61 (1): 35–39.
- Olson, D. R. (1977): From utterance to text: The bias of language in speech and writing. *Harvard Educational Review* 47 (3): 257–281.
- Ooi, V. (2002): Aspects of computer-mediated communication for research in corpus linguistics. In P. Peters, P. Collins & A. Smith (eds.) *New frontiers of corpus research: Papers from the twenty first international conference on English language research on computerized corpora, Sydney 2000*, 91–104. New York: Rodopi.
- Örnberg, T. (2003): Linguistic presence on the Internet: Communication, world-view and presence in online virtual environments. Unpublished D-extended essay, Department of Modern Languages, Umeå University, Umeå.
- Oxford English dictionary* (2000-): [Online version <http://www.oed.com>] Oxford: Oxford University Press (2015-10-13).
- Pak, A. & P. Paroubek (2010): Twitter as a corpus for sentiment analysis and opinion mining. *Proceedings of the Seventh Conference on International Language Resources and Evaluation LREC'10, Valletta, Malta, European Language Resources Association ELRA*, 1320–1326.
- Paolillo, J. C. (2011): “Conversational” codeswitching on Usenet and Internet relay chat. *Language@Internet* 8, article 3.
- Peterson, E. E. (2011): How conversational are weblogs? *Language@Internet* 8, article 8.
- Petrović, S., M. Osborne & V. Lavrenko (2010): The Edinburgh Twitter Corpus. *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media*, 25–26. <<http://www.aclweb.org/anthology/W10/W10-05.pdf#page=37>> (2015-10-13).
- Pichlmair, M. (2010): *Pwned – 10 tales of appropriation in video games*. <http://publik.tuwien.ac.at/files/pub-inf_4395.pdf> (2015-10-13).

- Pioch, N. (1997): *A short IRC primer*. <<http://www.irchelp.org/irchelp/ircprimer.html>> (2015-10-13).
- Poole, M. E. & T. W. Field (1976): A comparison of oral and written code elaboration. *Language and Speech* 19 (4): 305–312.
- Portnoy, S. (1973): A comparison of oral and written verbal behavior: An empirical approach. In K. Salzinger & R. S. Feldman (eds.) *Studies in verbal behavior*, 99–151. New York: Pergamon.
- Protalinski, E. (2015) Facebook passes 1.44B monthly active users and 1.25B mobile users; 65% are now daily users. *VentureBeat* 2015-04-22. <<http://venturebeat.com/2015/04/22/facebook-passes-1-44b-monthly-active-users-1-25b-mobile-users-and-936-million-daily-users/>> (2015-10-13).
- Quirk, R., S. Greenbaum, G. Leech & J. Svartvik (1985): *A comprehensive grammar of the English language*, 2. impr. Harlow: Longman.
- Rafaeli, S., F. Sudweeks, J. Konstan & E. Mabry (1998): ProjectH: A collaborative quantitative study of computer-mediated communication. In F. Sudweeks, M. McLaughlin & S. Rafaeli (eds.) *Network and netplay: Virtual groups on the Internet*, 265–281. Cambridge, MA: MIT Press.
- Reid, E. (1991): *Electropolis: Communications and community on Internet relay chat*. Honours thesis. University of Melbourne, Melbourne <<http://cyber.eserver.org/reid.txt>> (2015-10-13).
- Reid, E. (1994): *Cultural formations in text-based virtual realities*. Master's thesis. University of Melbourne, Melbourne <<http://www.aluluei.com/cult-form.htm>> (2015-10-13).
- Renouf, A. (1986): The elicitation of spoken English. In G. Tottie & I. Bäcklund (eds.) *English in speech and writing: A symposium*, 177–197. Stockholm: Almqvist and Wiksell.
- Reppen, R. (2001): Register variation in student and adult speech and writing. In S. Conrad & D. Biber (eds.) *Variation in English: Multi-dimensional studies*, 187–199. Harlow: Longman.
- Rowe, C. (2011): *Whatchanade?* Rapid language change in a private email sibling code. *Language @Internet* 8, article 6.
- Sacks, H., E. Schegloff & G. Jefferson (1974): A simplest systematics for the organization of turn-taking for conversation. *Language* 50 (4): 696–734.
- Santorini, B. (1990): *Part-of-speech tagging guidelines for the Penn Treebank project*, 3rd revision, 2nd printing. Philadelphia: University of Pennsylvania.
- Sapir, E. (1921): *Language: An introduction to the study of speech*. New York: Harcourt, Brace.

- Saussure, F. de (1966): *Course in general linguistics*, 1. ed., edited by C. Bally & A. Sechehaye in collaboration with A. Riedlinger and translated with an introduction and notes by W. Baskin. New York: McGraw-Hill.
- Schiffrin, D. (1982): *Discourse markers: Semantic resource for the construction of conversation*. PhD dissertation. University of Pennsylvania, Philadelphia, Pennsylvania.
- Schiffrin, D. (1985): Conversational coherence: The role of *well*. *Language* 61: 640–647.
- Schiffrin, D. (1994): *Approaches to discourse*. Oxford: Blackwell.
- Schiffrin, D., D. Tannen & H. E. Hamilton (eds.) (2001): *The handbook of discourse analysis*. Oxford: Blackwell.
- Schulze, M. (1999): Substitution of paraverbal and nonverbal cues in the written medium of IRC. In B. Naumann (ed.) *Dialogue analysis and the mass media*, 65–82. Tübingen: Niemeyer.
- Scoble, R. & S. Israel (2006): *Naked conversations: How blogs are changing the way businesses talk with customers*. Hoboken, NJ: John Wiley & Sons.
- Severinson Eklundh, K. (1986): *Dialogue processes in computer-mediated communication: A study of letters in the COM system* (Linköping Studies in Arts and Sciences 6). University of Linköping.
- Severinson Eklundh, K. (2010): To quote or not to quote: Setting the context for computer-mediated dialogues. *Language@Internet* 7, article 5.
- Spitzer, M. (1986): Writing style in computer conferences. *IEEE Transactions on Professional Communication* PC29 (1): 19–22.
- Squires, L. (2007): Whats the use of apostrophes? Gender difference and linguistic variation in instant messaging. *American University TESOL Working Papers*, no. 4.
- Stubbs, M. (1996): *Text and corpus analysis*. Oxford: Blackwell.
- Surman Paley, K. (2001): *I-writing: The politics and practice of teaching first-person writing*. Carbondale: Southern Illinois University Press.
- Svartvik, J. (ed.) (1990): *The London-Lund Corpus of Spoken English: Description and research* (Lund Studies in English 82). Lund: Lund University Press.
- Sveningsson, M. (2001): *Creating a sense of community: Experiences from a Swedish web chat*. PhD dissertation. Department of Communication Studies, Linköping University, Linköping.
- Swales, J. (1990): *Genre analysis: English for academic and research settings*. Cambridge, UK: Cambridge University Press.

- Tagliamonte, S. & D. Denis (2006): LOL for real! Instant messaging in Toronto teens. Paper presented at the Linguistic Association of Canada and the United States (LACUS), July 31 – August 4. Toronto, Canada.
- Tagliamonte, S. & D. Denis (2008): Linguistic ruin? LOL! Instant messaging and teen language. *American Speech* 83 (1): 3–34.
- Tannen, D. (ed.) (1982a): *Spoken and written language: Exploring orality and literacy*. Norwood, NJ: Ablex.
- Tannen, D. (1982b): Oral and literate strategies in spoken and written narratives. *Language* 58 (1): 1–21.
- Tannen, D. (1985): Relative focus on involvement in oral and written discourse. In D. R. Olson, N. Torrance & A. Hildyard (eds.) *Literacy, language, and learning: The nature and consequences of reading and writing*, 124–147. Cambridge, UK: Cambridge University Press.
- Tannen, D. (1990): *You just don't understand*. New York: Dell Books.
- Tannen, D. (1994): *Gender and discourse*. New York: Oxford University Press.
- Thompson, P. A. & D. A. Foulger (1996): Effects of pictographs and quoting on flaming in electronic mail. *Computers in Human Behavior* 12: 225–243.
- Thurlow, C. & K. Mroczek (eds.) (2011): *Digital discourse: Language in the new media*. New York: Oxford University Press.
- Tottie, G. (1981): Negation and discourse strategy in spoken and written English. In D. Sankoff & H. Cedergren (eds.) *Variation omnibus*, 271–284. Edmonton, Alberta: Linguistic Research.
- Tottie, G. (1982): Where do negative sentences come from? *Studia Linguistica* 36: 88–105.
- Tottie, G. (1983a) *Much about 'not' and 'nothing': A study of the variation between analytic and synthetic negation in contemporary American English*. Lund: Gleerup.
- Tottie, G. (1983b): The missing link? Or, why is there twice as much negation in spoken English as in written English? In S. Jakobson (ed.) *Proceedings from the Second Scandinavian Symposium on Syntactic Variation* (Stockholm Studies in English 62), 67–74. Stockholm: Almqvist & Wiksell.
- Tottie, G. (1986): The importance of being adverbial: Focusing and contingency adverbials in spoken and written English. In G. Tottie & I. Bäcklund (eds.) *English in speech and writing: A symposium*, 93–118. Stockholm: Almqvist and Wiksell.
- Tottie, G. (1991): *Negation in English speech and writing: A study in variation*. London: Academic Press.

- Tottie, G., B. Altenberg & L. Hermerén (1983): *English in speech and writing: A manual for students* (ETOS report 1). Lund: The Departments of English at the Universities of Lund and Uppsala.
- Tsui, A. B. M. (1991): The pragmatic functions of *I don't know*. *Text – Interdisciplinary Journal for the Study of Discourse* 11 (4): 607–622.
- Turkle, S. (1995): *Life on the screen: Identity in the age of the Internet*. New York: Simon & Schuster.
- Ure, J. (1971): Lexical density and register differentiation. In G. E. Perren & J. L. M. Trim (eds.) *Applications of linguistics: Selected papers of the Second International Congress of Applied Linguistics*, 443–452. London: Cambridge University Press.
- Van de Velde, K. & J. Meuleman (2004): *Lexical tensions in 'internet english': 1337 as language?* <<http://www.verbumvanum.org/kristof/index.html>> (2015-10-13).
- Van der Wege, M. & H. H. Clark (1997): Turn-taking systems for computer-mediated communication. In M. G. Shafto & P. Langley (eds.) *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, 1077. Mahwah, NJ: Erlbaum.
- Voelker, C. H. (1942): The one-thousand most frequent spoken words. *The Quarterly Journal of Speech* 28 (2): 189–197.
- Waldner, P. (2009): *Typing the untypeable: Paralinguistic features in IRC (Internet Relay Chat)* Diplomarbeit. Universität Wien, Vienna <http://othes.univie.ac.at/7148/1/2009-10-14_0216767.pdf> (2015-10-13).
- Wales, K. (1996): *Personal pronouns in Present-day English*. Cambridge, UK: Cambridge University Press.
- Walther, J. B. & K. P. D'Addario (2001): The impacts of emoticons on message interpretation in computer-mediated communication. *Social Science Computer Review* 19 (3): 324–347.
- Wells, R. (1960): Nominal and verbal style. In T. A. Sebeok (ed.) *Style in language*, 213–220. Cambridge, MA: MIT Press.
- Werry, C. C. (1996): Linguistic and interactional features of Internet relay chat. In S. C. Herring (ed.) *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*, 47–63. Amsterdam: John Benjamins.
- Wilkins, H. (1991): Computer talk: Long-distance conversations by computer. *Written Communication* 8 (1): 56–78.
- Wolf, A. (2000): Emotional expression online: Gender differences in emoticon use. *CyberPsychology and Behavior* 3 (5): 827–833.

- Woodburn, R., R. Proctor, J. L. Amott & A. F. Newell (1991): A study of conversational turn-taking in a communication aid for the disabled. In D. Diaper & N. Hammond (eds.) *Proceedings of HCI 91: People and Computers VI*, 359–371. Cambridge, UK: Cambridge University Press.
- Woolbert, C. H. (1922): Speaking and writing: A study of differences. *Quarterly Journal of Speech Education* 8 (3): 271–285.
- W3Techs (2015): *Usage of content languages for websites*. <http://w3techs.com/technologies/overview/content_language/all> (2015-10-13).
- Yates, J. A. & W. Orlikowski (1993): *Knee-jerk anti-LOOPism and other e-mail phenomena: Oral, written and electronic patterns in computer-mediated communication*. Working paper, Cambridge, MA: MIT. <<http://ccs.mit.edu/papers/CCSWP150.html>> (2015-10-13).
- Yates, S. (1993): The textuality of computer-mediated communication: Speech, writing and genre in CMC discourse. Unpublished PhD thesis. Open University, Milton Keynes.
- Yates, S. (1996): Oral and written linguistic aspects of computer conferencing. In S. C. Herring (ed.) *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*, 29–46. Amsterdam: John Benjamins.
- Zipf, G. K. (1949): *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge, MA: Addison-Wesley Press.
- Zitzen, M. (2004): *Establishing topical coherence in electronic communication: Topic shift markers in asynchronous and synchronous computer mediated communication (CMC)*. Stuttgart: Ibidem.

