

Use and Adaptation of Written Language to the Conditions of Computer-Mediated Communication

Ylva Hård af Segerstad



Department of Linguistics
University of Gothenburg, Sweden
2002

Use and Adaptation of Written Language
to the Conditions of
Computer-Mediated Communication

Ylva Hård af Segerstad

Doctoral Dissertation

Publicly defended in Stora Hörsalen,
Humanisten, University of Gothenburg,
on December 21, 2002, at 10.00,
for the degree of Doctor of Philosophy

Department of Linguistics
Göteborg University, Sweden
2002

This volume is a revised version of the dissertation.

ISBN 91-973895-3-6

Use and Adaptation of Written Language to the Conditions of Computer-Mediated Communication

Abstract

The purpose of the present study is to investigate how written language is used and adapted to suit the conditions of four modes of computer-mediated communication (CMC). Texts from email, web chat, instant messaging and mobile text messaging (SMS) have been analyzed. The general human ability to adapt is deemed to underlie linguistic adaptation. A linguistic adaptivity theory is proposed here. It is proposed that three interdependent variables influence language use: synchronicity, means of expression and situation. Two modes of CMC are synchronous (web chat and instant messaging), and two are asynchronous (email and SMS). These are all tertiary means of expression, written and transmitted by electronic means. Production and perception conditions, such as text input technique, limited message size, as well as situational parameters such as relationship between communicators, goal of interaction are found to influence message composition.

The dissertation challenges popular assumptions that language is deteriorating because of increased use in CMC. It is argued that language use in different modes of CMC are variants, or repertoires, like any other variants. Contrary to popular assumptions, results show that language use is adapted creatively and is well suited the particular modes of CMC. A number of linguistic features are shown to be characteristic of the modes of CMC investigated in the present study. Strategies such as syntactical and lexical reductions are employed to reduce *time*, *effort* and *space*. These techniques often appear to serve multifunctional purposes, by expressing interpersonal intimacy by the choice of words and phrases, while reducing keystrokes. This clearly indicates linguistic awareness.

Texts in email, web chat, instant messaging and SMS are found to contain unconventional and not yet established abbreviations based on Swedish as well as words from other languages, unconventional or spoken-like spelling, unconventional use of punctuation and use of non-alphabetical graphical means (emoticons, asterisks). Thus, written language is found to have been developed and enhanced to suit the conditions of computer-mediated communication.

KEY WORDS:

computer-mediated communication; human adaptability; email, web chat; instant messaging; SMS.

The thesis is written in English.

Skicka Chatta Bifoga Adress Typsnitt Färger Spara som utkast

Till: all.readers@dissertation.nu

Kopia:

Ämne: acknowledgement

Signatur: Ingen

ahhh, finally. i know its been a long time. ;D dissertationcreation is a lonely business. it couldn't have been done completely in isolation, mindyou, even though the actual typing of text was done by me alone.

jonas, husband, without you everything is impossible. without you there would be no reason to strive. it's true. especially dissertation-writing. my *beloved* family: brothers, mum, dad and you'll have me back now. hmm, at least i think so!!!!

annelie ädel, truest of true friends: you're invaluable! *michael lindström*, I don't think you know how much I appreciate your steady friendship! heavily texting *hanna fjällman*!

all my competent and inspiring colleagues at my home department: *leif grönqvist* (i honestly believe our department would collapse without you! + you're a good friend too), *sally boyd*, *åsa nordqvist*, *åsa wengelin*, *daniel ridings*, *torbjörn lager*, *staffan larsson*, *ulla veres*. the list would be too long!! oh, not to forget: my supervisors: *Jens allwood* and *elisabeth ahlsén*.

Being a research student has taken me to many places in the world where i'd never been to before. one of the best parts of going to conferences is when you get to know new people who sometimes turn out to be personal friends. *malin sveningsson* (may all our bathrooms be tiled from now on :-D), *t.l. taylor*, *nalini kotamraju*, *alexandra weilenman*, *marianne sandberg*: CU on the other side, gals! thanks to *charles ess* for sharing your wisdom and encouraging a disillusioned doctoral student! the friendly enthusiasm of *tuija virtanen* warms the coldest of hearts.

The studies of CMC naturally all required messages to analyze. I'm so grateful to all you guys who contributed with your email, chat, instant messaging, and sms messages. special thanks to my *sms informants*!

genie perdin, you've excelled in scrutiny and balance when checking my Swenglish in this text!!! just imagine what it was like before her professional check! *shudder* thanks also to *urban lindblom*, who believed me when i needed it. thanks so much to *Benjamin lyngfelt* for advice on layout, and for always being so friendly!

one wouldn't embark on a study of this extent if one didn't have a genuine interest in language and human behavior. (I mean, how many times haven't we all asked ourselves whether one has to be maniac to start writing a dissertation, or whether you go mad while doing it????) luckily, I believe there's still some more to be researched out there...

Ylva Hård af Segerstad.
November 2002



Table of Contents

PART I	1
1 INTRODUCTION	1
1.1 INTRODUCTION	1
1.2 PURPOSE OF THE DISSERTATION	5
1.2.1 Hypotheses	6
1.3 OUTLINE OF DISSERTATION	7
1.4 A NOTE ON THE COMPOSITION OF THIS DISSERTATION	8
2 BACKGROUND	10
2.1 INTRODUCTION	10
2.2 HUMAN COMMUNICATION.....	11
2.2.1 Multilevel organization of communicative interaction.....	13
2.2.2 Communication as rational and cooperative action and interaction	15
2.2.3 Communication and Social/Situational Variables.....	16
2.2.4 Participants, Situation and Context	18
2.2.5 Registers, Genres, Activities	19
2.2.6 Communication management.....	22
2.2.7 Establishing shared information	26
2.2.8 Constraints on Grounding, or Sharing Information.....	28
2.2.9 Frequency effects and the principle of least effort.....	31
2.3 VARIABLES CONDITIONING HUMAN COMMUNICATION.....	32
2.4 SPOKEN AND WRITTEN COMMUNICATION.....	36
2.4.1 Speech and writing: manifestations of the same system?.....	36
2.4.2 Speech and writing, spoken and written language.....	39
2.4.3 Variables Conditioning Speech and Writing	40
2.4.4 Are differences between speech and writing due to means of expression, or genre?.....	46
2.5 COMPUTER-MEDIATED COMMUNICATION.....	50
2.5.1 Does CMC have a negative influence on language?	51
2.5.2 A third medium?.....	53
2.5.3 Modes of text-based CMC.....	55
2.5.4 Variables conditioning CMC	57
2.5.5 Properties of the modes of CMC investigated in the present study.....	63
2.6 CHAPTER SUMMARY	76
3 MATERIAL & METHODS	78
3.1 INTRODUCTION	78
3.2 ETHICAL CONSIDERATIONS OF COLLECTING AND HANDLING DATA	79
3.3 GENERAL METHOD OF ANALYSIS	81

3.4	A NOTE ON SELECTION OF MATERIAL	82
3.5	REPRESENTATIVITY OF THE DATA.....	83
3.6	METHODS AND MATERIAL IN FOUR MODES OF CMC.....	84
3.6.1	Analyses of Language Use in Email	84
3.6.2	Analyses of Language Use in Web Chat.....	86
3.6.3	Analyses of Language Use in Instant Messaging.....	89
3.6.4	Analyses of Language Use in SMS.....	92
3.7	OVERVIEW OF DATA FROM FOUR MODES OF CMC	94
3.8	CHAPTER SUMMARY	94
PART II		95
4	ELECTRONIC LETTERS TO UNKNOWN AUTHORITIES	96
4.1	INTRODUCTION	96
4.1.1	Aim of the present study	98
4.2	PREVIOUS STUDIES OF EMAIL	99
4.2.1	Analysis of messages in an electronic Swedish conference system	99
4.2.2	Analysis of messages in electronic mailing lists	99
4.2.3	Analysis of private email.....	101
4.2.4	Closure on actions	102
4.3	THE PRESENT STUDY: ELECTRONIC LETTERS TO UNKNOWN AUTHORITIES	102
4.3.1	The Swedish business letter template and norms for informal letters .	104
4.4	MATERIAL AND METHOD	104
4.4.1	Material.....	104
4.4.2	Method.....	107
4.5	RESULTS	107
4.5.1	Overview of the corpora	107
4.5.2	Analysis of paper letters.....	112
4.5.3	Salutations and signatures	113
4.6	CONCLUSIONS FROM THE CATEGORIES	115
4.6.1	Salutation conventions.....	116
4.6.2	Closing conventions.....	117
4.6.3	Word frequency	117
4.6.4	Mean length of messages.....	118
4.7	CONCLUSIONS.....	119
4.8	CHAPTER SUMMARY	121
5	WRITTEN CONVERSATION IN A WEB CHAT.....	123
5.1	INTRODUCTION	123
5.2	REAL-TIME CHAT	125
5.2.1	Real-time Chat.....	125
5.2.2	Linguistic Characteristics of Chat	127
5.2.3	Constraints on Written Communication	128

5.2.4	Advantages of written communication	129
5.2.5	Strategies to adapt to text-only limitations	131
5.3	MATERIAL & METHOD	131
5.3.1	Material.....	131
5.3.2	Method.....	134
5.4	RESULTS	134
5.4.1	Strategies	140
5.5	CONCLUSIONS.....	150
6	INSTANT MESSAGING WITH WEBWHO	153
6.1	INTRODUCTION	153
6.2	AWARENESS TOOLS.....	154
6.2.1	Instant Messaging and Education.....	154
6.2.2	Awareness Support and Active Maps.....	155
6.3	WEBWHO IN USE	158
6.4	COMPUTER-MEDIATED COMMUNICATION.....	160
6.4.1	Features of CMC and factors that influence language use.....	160
6.4.2	Modes of Communication	162
6.4.3	Awareness of presence in WebWho.....	163
6.4.4	Data Collection and Analyses.....	164
6.4.5	Quantitative Analyses, Automatic Measures.....	165
6.4.6	Qualitative, Manual Analyses.....	165
6.5	RESULTS	166
6.5.1	Comparison with Speech and Writing, and with CMC.....	166
6.5.2	Physical and Virtual Awareness of Presence.....	169
6.5.3	Message Categories.....	169
6.6	DISCUSSION.....	181
6.7	CHAPTER SUMMARY	184
7	SMS – TEXT MESSAGING VIA MOBILE PHONES.....	186
7.1	INTRODUCTION	186
7.2	WHAT IS SMS?.....	186
7.3	PRODUCTION AND PERCEPTION CONDITIONS OF SMS	188
7.3.1	Text input	188
7.3.2	Screen size.....	189
7.3.3	Text input using predictive software	191
7.3.4	Texting via web based SMS services	192
7.4	PREVIOUS RESEARCH ON SMS.....	194
7.4.1	Accounts and popularity.....	194
7.4.2	Use and function of SMS	194
7.5	LINGUISTIC CHARACTERISTICS OF SMS.....	197
7.5.1	Reduced forms in SMS communication	197
7.5.2	Writing and graphic means	199

7.5.3	Means to reduce text in SMS	201
7.6	THE PRESENT STUDY	206
7.6.1	Research questions	206
7.7	METHODS FOR DATA COLLECTION AND ANALYSES.....	207
7.7.1	A note on the choice of material and its implications on the results ...	207
7.7.2	Data collection	208
7.7.3	Methods for analyzing the data	211
7.8	RESULTS	212
7.8.1	A comparison of word frequency between modes of communication	213
7.8.2	Linguistic features in the Swedish SMS corpus	214
7.8.3	Punctuation.....	215
7.8.4	Spelling.....	217
7.8.5	Grammar.....	224
7.8.6	Graphical (non-alphabetic) means	226
7.9	CONCLUSIONS	228
7.10	231
7.10	FUTURE RESEARCH ON MOBILE MESSAGING	231
7.11	CHAPTER SUMMARY.....	231
8	CONCLUDING DISCUSSION.....	233
8.1	INTRODUCTION	233
8.2	LINGUISTIC FEATURES IN CMC	234
8.3	SUMMARY OF THE DISSERTATION.....	236
8.3.1	Chapter 1: Introduction.....	236
8.3.2	Chapter 2: Background.....	236
8.3.3	Chapter 3: Methods and Material.....	237
8.3.4	Chapter 4: Email.....	237
8.3.5	Chapter 5: Web chat.....	241
8.3.6	Chapter 6: Instant Messaging	245
8.3.7	Chapter 7: Text Messaging via Mobile Phones – SMS.....	248
8.4	COMPARISONS AMONG THE SUB STUDIES.....	252
8.4.1	Word frequency, mean length of utterance, vocabulary richness	252
8.4.2	Generally compared findings.....	256
8.5	LINGUISTIC ADAPTIVITY THEORY	259
8.5.1	Instrumental rationality.....	259
8.5.2	A genuinely new medium?	260
8.5.3	Is language deteriorating?	262
8.6	THE MAIN CONTRIBUTIONS OF THE DISSERTATION	263
8.7	IMPLICATIONS OF RESULTS	264
8.8	FUTURE RESEARCH	265

REFERENCES

Part I

1 Introduction

1.1 Introduction

This section comprises a general introduction to investigation in the field of computer-mediated communication, or CMC for short, and to what kind of research the reader might expect to find. Motivations are provided as to why this field is interesting and worth studying. Basic characteristics of written and spoken language are given, compared with CMC as an intermediate or hybrid mode of communication, with references to a more in-depth treatment of data in the background chapter. It is claimed that the ability to adapt is central to human behavior, and that written language adaptations observed in CMC are an example of this. The purpose of this dissertation is stated, together with the questions that will be dealt with in this work. The introduction section is concluded with an outline of this dissertation.

This doctoral dissertation is an investigation of how written language is used, adapted and developed to suit the needs and provisions in different kinds of text-based computer-mediated communication which are used today. We use computer-mediated communication (henceforth CMC) more and more in our everyday lives, both for private and professional purposes. Communication that is mediated by various modes of technology seems to affect the written language that we use in this context. Communication via the Internet and mobile (cellular) phones are still rather new modes of interaction, and these are becoming increasingly more popular for social and recreational purposes and are being demanded for work related tasks in a great variety of professional fields. Four different modes of CMC will be investigated in this work: email, web chat, instant

messaging, and SMS (Short Message Service: text messaging via mobile phones).

There are modes of CMC that are not text-based, but it is at present much more common to send an email than to communicate via video conferencing systems or web cameras. This thesis will address only the above-mentioned, text-based modes of CMC. Much modern communication is text-based, providing us with more tangible evidence of the language that we use and observe, and draws our attention to how this kind of language use changes, as it were, continually. There is considerable popular attention directed to the increasing use of text-based modes of communication and on current language issues. On occasion, one notes short paragraphs in newspapers where writers voice their irritation concerning language use that they observe in email messages or in chat rooms. Many observers fear that due to the growing influence of the Internet and computer-mediated communication, people are going to lose knowledge of how language ought to be used. They are alarmed at finding language changing, thus disregarding or being unaware of the fact that language is normally in a constant state of flux. The work in this dissertation is a descriptive linguistic study of language used in a number of text-based modes of CMC widely used today. The reader should not expect an investigation made in order to make prescriptive statements as to how language ought best to be used. This is generally of little interest to the linguist. The main interest of this dissertation is to analyze how written language is used in text-based CMC, as well as how and why this usage differs from the norms of traditional written language. Text-based CMC also shows evidence that the system and norms of written language are in the process of being developed and adapted to meet the constraints and provisions of means of expression and situation, as well as to suit the communicative wishes of the users (cf. Crystal 2001). CMC has often been treated as a sort of hybrid between spoken and written communication (cf. Du Bartell 1995, among others), exhibiting features that are characteristic of both spoken and written language. It is important to point out here that CMC can not to be treated as a single mode of communication resulting in the same kind of language use everywhere (cf. Ferrara, Brunner et al. 1991). There are great differences between the settings and of how means of expression and situation are involved, for example, when sending an email or when communicating in a chat room, so we must not fail to distinguish between different modes and uses of CMC.

An overview of research on written and spoken language is needed in order to find out in what ways different modes of CMC exhibit written or

spoken language features, why different modes manifest these differently, and what specific “e-style”¹ adaptations may occur. The background below will provide a more thorough account of relevant research on the conditions and settings of written and spoken language, as well as of research on different modes of CMC.

Prototypical spoken language is found in face-to-face interaction, which employs simultaneous multimodal channels for sending information to guide interpretation via visual, auditory, and in some cases even tactile cues. Written language is monomodal and linear, relying on visual representations of speech.² Face-to-face spoken communication is an interaction between two or more interlocutors present at the same time and place, and is often characterized as spontaneous and unedited, composed on the spur of the moment, as there often is no time to plan and correct the flow of conversation, which results in self-corrections, pauses, false starts, etc. (cf. Linell 2001/in preparation, and others). Prototypical traditional written language is found in formal pen-and-paper (or type-written) letters newspaper articles and books. The traditional prototypical writing process is a solitary activity in which the writer has time to plan what to communicate and to edit, so that the reader is provided with a final version of a text without any trace of revision. It is often supposed to be fact that spoken language is characterized by a syntax that is less complex than that of written language, and that it employs less varied vocabulary (Chafe and Danielewicz 1987). This may not be the full story, which we will get back to below (cf. Linell 2001/in preparation). Some genres and uses of written language take on what is typically associated with what is regarded as representative for spoken language, and the other way around. A formal talk or lecture may resemble written repertoire, and memo notes or informal letters between friends are in most cases written in a casual style often disregarding the normative rules of traditional written language. People adapt their language to what the situation requires. There are no absolute boundaries between what “counts as” written language or what is representative of spoken interaction. Differences between spoken and written language are not always clear-cut. One might regard the phenomena as a continuum between the two modes, as argued in the works of Biber (1988), Linell (2001/in preparation) and

¹ (Maynor 1994) refers to the most apparent “e-style” features as “the lack of capital letters, simplified spellings, clipping, and icons [emoticons]”, which will be accounted for in detail below (see Chapter 4).

² The written language that is analyzed in this study employs the roman alphabetic writing system, and will henceforth be referred to as “writing”, in brief.

others. This issue will be addressed in more detail below (see the Background chapter).

In CMC new uses for written language are found, and the situations are different from those in traditional writing. As mentioned above, different modes of CMC have different settings; it is important to make the distinction between asynchronous and synchronous CMC. Asynchronous CMC does not require the communicators to be online and available at the same time or place in order for communication to take place successfully. In this respect, asynchronous CMC resembles the settings of traditional writing. Examples of asynchronous CMC are email and SMS. Messages are composed off-line providing the sender the benefit of time for planning and editing the message. Thus, asynchronous CMC takes on some features afforded to written language. Like spoken interaction, synchronous CMC requires its interlocutors to be online simultaneously for successful communication. Synchronous CMC allows for interactive written communication, for example, in different forms of chat (web chat, IRC, etc.). The conditions of synchronous and asynchronous CMC, as well as an account of research that is relevant for this study will be examined further in Section 2.3 below.

Regardless of which mode of CMC is being analyzed, all text-based communication still has the characteristic in common of being language represented in visual form. The writing system is perhaps not as well suited or developed for its goals as it might be, not even in traditional writing settings. Compared with spoken interaction, written communication might be considered a “lean” medium constrained by factors such as linearity, monomodality, and effort of production (cf. Daft and Lengel 1984). What might be observed in the written language of text-based CMC are not only spoken language characteristics due to medium and situation, but also developments made to overcome the difficulties of a communication system that has the disadvantage of being constrained by linearity, etc. It might be argued that writing was never intended to convey all the information that spoken interaction conveys in the first place, and that it is used for other purposes than spoken interaction. As such, the observable adaptations made to suit the situation and meet the communicative needs and wishes of the users might be regarded as examples of human linguistic adaptability, which in turn might tell us something about the human ability for adaptation in general (cf. Brenner 1975).

During work on the studies that comprise this dissertation, this author has had occasion to meet people who appreciate work on a topic that is of the essence in human life: language and communication. Moreover, the

particular sort of communication focused on in this dissertation, CMC, is something that is gaining more and more significance in the lives of many people today, and something that people seem to have a lot of opinions and concerns about. Most people employ a somewhat different perspective to what they perceive to be happening to written language, than a linguist does. Laymen often fear that language, especially written language, is in the process of being degraded into something that is less suited to carry all the nuances and distinctions that it used to communicate before. Far from being new, the fear that language is being corrupted has been recorded since the days of Plato. People believed that the use of language by the uneducated, and perhaps worst of all, by young people, caused language to degenerate into something less 'pure' or 'original' than previously, and thus less expressive. It was believed that if this process was allowed to continue, one day humans would no longer be able to communicate or to understand each other. These misconceptions are at the heart of this thesis: the ability to adapt - be it the adaptation of our behavior to suit the climate of the environment around us or the adaptation of the use of language - is central to human behavior. Moreover, it will be argued that young people's use of language is no less degenerate than any other use, but may be seen as one of the driving forces for the development of language to suit the needs and opportunities of communication.

1.2 Purpose of the dissertation

The purpose of this dissertation is to investigate how written language is used and adapted to a number of modes of computer-mediated communication. More specifically, this thesis will address the following issues.

- How is written Swedish used and adapted in various modes of CMC?
- In what ways do these modes of CMC differ from the norms of traditional written language and why?
 - Which are the written and spoken language features in each mode respectively, and why are these used?
 - Which specific "e-style" characteristics are found in different CMC modes?
- Which variables influence these adaptations?

Both medium and situational variables are assumed to influence the adaptation of written language to CMC, not all of those variables are relevant to research in this work. Medium variables (cf. Herring Forthcoming) to be investigated below are:

- Production conditions
 - Synchronicity
 - Possibility to signal feedback (1-way vs. 2-way message transmission)
 - Effort of production (physical and cognitive)
 - Cost of transmission (economic)
 - Persistence of transcript
 - Levels of persistence
- Size of message buffer
- Anonymous messaging
- Private messaging

Situational variables (cf. Herring Forthcoming) to be investigated are:

- Participation structure
 - Relation between communicators
- Participant characteristics
- Purpose
- Activity
- Norms

Chapter 3 (Material & methods) will provide a detailed account for all of the above-mentioned interdependent variables and motivate the choice of the same as relevant to research in the present study.

1.2.1 Hypotheses

The following hypotheses will be tested in the course of this dissertation.

Hypothesis 1

Synchronous CMC will exhibit more features of spoken language than asynchronous CMC. The study comprises two synchronous modes (web chat and instant messaging) and two asynchronous modes (email and SMS).

Hypothesis 2

Unlimited buffer size will result in more edited, written language features than limited buffer size. The study comprises two modes with unlimited buffer size (email and instant messaging), and two modes with limited buffer size (web chat and SMS).

Table 1. Hypothesized effects of synchronicity and buffer-size in four modes of computer-mediated communication.

	<i>Synchronous</i>	<i>Asynchronous</i>	
<i>Limited buffer</i>	WEB CHAT	SMS	<i>Spoken-like</i>
<i>Unlimited buffer</i>	INSTANT MESSAGING	EMAIL	<i>Written-like</i>
	<i>Spoken-like</i>	<i>Written-like</i>	

1.3 Outline of dissertation

This dissertation is composed of analyses of how written language is used and adapted to four modes of CMC: email, web chat, instant messaging and SMS. The case studies will be treated in separate chapters.

A theoretical background to this study follows in the Background chapter. The first section deals with theories of general human communication. The second section investigates previous studies of written and spoken language, stating the conditions and commonly held stereotypical features of these modes. The third section deals with various previous studies of CMC. This chapter suggests a taxonomy of a number of variables that place conditions on communication in CMC.

The theoretical background is followed by an overview of the methods used for collecting and analyzing data from the four different modes of CMC that form the material for this dissertation, as well as comparisons among the four studies that comprise this dissertation.

Following the account of methods are separate reports on the studies comprising email, web chat, instant messaging, and SMS, with an account of the methods specific for each case study together with a report on results followed by discussion.

The study on email is an analysis of electronic letters to the city council in Göteborg city, Sweden, compared to traditional pen and paper letters to the same institution. This particular setting for messages was chosen partly

because it is a corpus of email messages that is publicly accessible³, and comparable to traditional letters of the same type. This is an asynchronous mode of communication, much like traditional letter writing in that respect.

The web chat study investigates language use in a public web chat. Data were collected by logging all contributions to a specific chat room during one week. Web chat is a synchronous mode of communication that requires all participants to be logged in at a chat room at the same point in time.

The study of instant messaging analyzes messages sent through the messaging service of the WebWho software that is used in a large university computer lab. WebWho was primarily intended to provide a visual overview of the presence of students logged in at the workstations in the lab. The instant messaging service provides interesting data due to the sender's awareness of the receiver's presence when sending the messages.

The investigation of SMS messages comprises analyses of mobile text messages. The messages were created both on the keypad of mobile phones as well as on computer keyboards utilizing web based SMS services. The messages are limited to a maximum of 160 characters each using either of the two techniques of production. Mobile text messaging does not require synchronous communication.

The dissertation concludes with a discussion of the results and a general discussion of the implications of these results. Using the findings and experience gained on completing this study, future research in CMC and linguistics is suggested. Human adaptability to given situations is a clear theme for this dissertation.

1.4 A Note on The Composition of this Dissertation

This dissertation is composed of studies of four different modes of CMC. Three of these four studies have been published as articles in refereed scientific journals: the study of email (Hård af Segerstad 2000b), the study

³ All official public documents in Sweden may be accessed by any citizen on request. The principle of public accessibility concerns official public documents (in Swedish, "allmänna handlingar"). By this is meant documents that have been 1) received by the public authority, or 2) produced by the public authority (Wirén 1998). More on this in the chapter on the email study (Chapter 4).

of web chat (Hård af Segerstad Forthcoming), and the study of instant messaging (Hård af Segerstad and Ljungstrand 2002).

This particular dissertation is a kind of hybrid between a monograph, which is the traditional format for a dissertation in the humanities at Swedish universities, and a compound dissertation, which is more common for dissertations in the natural sciences. A compound dissertation is comprised of a collection of refereed articles, an introduction and a concluding and discussing chapter. The format of this dissertation is as follows: it opens with an introduction, which is followed by a theoretical background, and a chapter on the methods of analysis and data collection. These three first chapters make up Part I of the dissertation. Part II consist of the four studies on modes of CMC, as well as a concluding discussion. Each of the case study chapters will have a more detailed background of the specific field of investigation, as well as a specific account of methods for analysis and data collection. The case studies are followed by a concluding discussion that compares the results of the case studies and the implications they have on the research questions in the dissertation. The final chapter is concluded with a reference to future research.

2 Background

2.1 Introduction

This chapter provides a background to research that is relevant for the questions raised in the studies that constitute this dissertation. It opens with an outline of relevant research on human communication in general, as a foundation for explaining what is involved when people communicate. This is followed by an account of relevant research on written and spoken language, providing a background to the assumptions and hypotheses that will be investigated in this work. This is followed by a discussion of previous interdisciplinary research on computer-mediated communication (CMC), primarily dealing with aspects of written and spoken language, upon which the methods of analysis in the present study are based. Results and methods of previous CMC research are evaluated and discussed where applicable to the purposes and material in this dissertation. A description of the modes of CMC that were investigated in this study is also given. The chapter concludes by summarizing the main points of the thesis in order to set a background for the subsequent chapter, which defines methods for data collection and analyses and outlines the material that was used in the studies.

This dissertation deals with how written language is used and adapted to suit the conditions for communicating effectively in a number of different modes of CMC⁴. There are both constraints and enablements⁵ (Allwood

⁴ Computer-mediated communication, or the alternate term computer-mediated discourse (CMD) (Herring Forthcoming). Further explanation is given in the section on CMC (Section 2.4).

2000; Herring 2001) in these modes that make communication and language use different from that of “traditional” written and spoken interaction (Soukup 2000). The purpose here is to analyze how people tailor their use of written language to the conditions of medium and situation, as well as to their communicative wishes. The aim is to answer the questions of how and why different modes of text-based CMC contrast with traditional writing and speech.

2.2 Human communication

Humans are not the only species that communicates, but complex language is probably unique to the human race. Pinker (1994) is a strong advocate of the view that the ability to speak and the brain structures necessary for language are unique to humans:

Language is a human instinct, wired into our brains by evolution like web-spinning in spiders or sonar in bats.

A complex set of interdependent variables is involved in all forms of communication (Allwood 1976; 1995). Linguistic, sociological, and psychological research has often chosen to focus on one or several of the variables involved in communication that fit the aims of the particular research questions at hand. Chafe (1986), and many others, point out that speaking has been with us from the very beginning of human history, whereas writing as we know it has existed for a comparatively brief period. Communication through speech is a resource available for all normally equipped human beings across different social groups and cultures (Linell 2001/in preparation). Writing and reading are usually acquired through deliberate instruction, which depends on previously acquired speaking abilities.

The first section of this chapter will focus on previous research addressing human communication from a general point of view, focusing on spoken interaction. Allwood (1995) defines communication in general as the sharing of information or understanding.

⁵The concept of “enablements” occurs in Allwood (2000) to supplement concepts such as possibilities, potentialities, affordances that do not always cover the intended meaning of “resources”. See further Section 2.2.1 below.

If communication is intentional it is claimed to have at least one joint purpose, i.e., the purpose of sharing information, or perhaps better, sharing understanding [...]

Allwood points out that this is indeed exactly what the etymology of communication indicates: the Latin word *communicare*, which means “to share, impart” (Lewis and Short 1963). Senders or coders of linguistic signals, i.e. speakers or writers, activate and share information with receivers or decoders of the same signal, i.e. hearers or readers. Regarding communication as activating and sharing information, which is Allwood’s view, gives a different perspective of communication than that which is intended by the term “conduit metaphor”. The term “conduit metaphor”, introduced by Reddy (1979), is based on the belief that ideas are like physical objects, and that the purpose of language is to provide a package for ideas for transfer between minds. Eubanks (2001) points out that according to the conduit metaphor, language contains meaning; speakers and writers use linguistic containers to send meaning to audiences; and at the end of the line, audiences remove the unaltered meaning from its container. The difference between the conduit metaphor and Allwood’s notion of communication as sharing of information lies in the activating and interactive aspect of information.

The medium of sending, or sharing, differs from situation to situation, as do the conditions for what may be communicated and how. The effects that various conditions of different communicative settings have on language use and message content will be dealt with further in the section on written and spoken language (Section 2.4). Even though complex language is unique to *Homo sapiens*, the ability to adapt to the conditions of context and situation necessary for survival certainly is not. Other animals also have this ability. *The ability to use and adapt language to the conditions of context and situation* is central to human behavior and most likely central to the evolutionary success of the human species (Diamond 1992). Wallace (1999) argues that *Homo sapiens* are both set in their ways and amazingly adaptable. In this dissertation, use and adaptations of linguistic behavior in computer-mediated communication are regarded as an example of this ability. On account of this ability to adapt, *Homo sapiens* may perhaps be even better named *Homo flexibilis*.

2.2.1 Multilevel organization of communicative interaction

All communication involves a complex set of finely interwoven and interdependent variables. Allwood's activity based communication analysis (Allwood 2000) takes into account the complexity of the relations that are established between the participants in an event of communication. Allwood argues that communication serves as an instrument in activities. People communicate for various reasons. Allwood builds his theory on Wittgenstein's insights concerning meaning. Meaning is determined by use in three types of context: perceptual context, social activity and activated background information.

Like Austin (1962) and Searle (1969), Allwood claims that speaking (by implication also writing and communication in general) should be regarded as a species of social action. Allwood also draws on pragmatic principles of rational communication and cooperation – similar to, but not identical with – those proposed by Grice (1957; 1975), see Section 2.2.2 below. As Grice notes, "meaning is a kind of intending," and the hearer's or reader's recognition that the speaker or writer means something by x is part of the meaning of x . Allwood does not share this view completely, and argues that meaning is and can be more than intending (cf. Allwood 1997).

In contrast to the assumptions of structuralism⁶ (a theory that privileges *langue*, the system, over *parole*, the speech act), speech act theory holds that the investigation of structure always presupposes something about meaning, language use, and extralinguistic functions. Speech act theory is a theory of language based on Austin's *How to Do Things with Words* (Austin 1962). The major premise of Speech act theory is that language is as much, if not more, of a mode of action as it is a means of conveying information. Allwood argues that communication in many ways seems to build on the human ability for rational coordinated, and thus cooperative, interaction. Following Allwood's Activity Based Communication Analysis, conveying information is in itself a mode of action. As Searle (1969) argued:

All linguistic communication involves linguistic acts. The unit of linguistic communication is not, as has generally been supposed, the symbol, word, or sentence, or even the token of the symbol, word, or sentence, but rather the production or issuance of the symbol or word or sentence in the performance of a speech act.

⁶ See Saussure (1916).

Meaning, then, should at least partly be regarded as a species within the genus intending-to-communicate, since language itself is highly complex, rule-governed intentional behavior. A theory of language is part of a theory of action. The basic emphasis of Speech act theory is on what an utterer (U) means by his utterance (x) rather than what x means in a language (L). This view leaves little space to either convention or receiver's interpretation (cf. Allwood 1976; 2000), which can be claimed to be part of meaning as well.

Allwood argues that the physical, biological, psychological, and social levels of organization are involved in any human activity. These levels constrain and enable communication whether it occurs in spoken or written form.

- (i) Physical
- (ii) Biological
- (iii) Psychological
 - (a) Perception, understanding and emotion
 - (b) Motivation, rationality and agency
- (iv) Social
 - (a) Culture, social institution
 - (b) Language
 - (c) Activity
 - (d) Communication

Communicators are physical entities; their communicative contributions are physical (acoustic/optical) processes or entities, which physical level indicates.

On the *biological level*, it is claimed that communicators are biological organisms, whose communicative contributions can be seen as biological activation and directed behavior. The *psychological level* deals with perception, understanding and emotion, as well as motivation, rationality and agency (see Section 2.2.2). Communicators are perceiving, understanding and emotional beings. Utterances, or communicative contributions, are perceptually comprehensible, emotionally charged phenomena. Communicators are also motivated (including ethical, cooperative motives), rational agents, whose communicative contributions are motivated, rational acts.

On the *social level*, communicators are members of a culture and one or more social institution. Therefore their communicative contributions can be characterized as cultural and social institutional acts. In addition, are

members of one or more linguistic communities, and their contributions are normally linguistic acts.

The notion of *activity* deals with the fact that the communicators normally play a role in a social activity. Their communicative contributions are contributions to that social activity through their role, e.g. as a sales clerk telling the customer the price of some goods, or a teacher lecturing. Communicators normally, at a given point in time, focus primarily either on sending or receiving information. That is, they are either in the sender (speaker, writer, etc.) role or in the receiver (addressee, listener, reader, etc.) role.

2.2.2 Communication as rational and cooperative action and interaction

Grice formulated a rough general principle that he labeled the Cooperative Principle. Participants in conversation are expected to observe this super ordinate principle, *ceteris paribus* (Grice 1975):

Make your conversational contribution such as required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

It can be argued that the conversational maxims which follow from the general cooperative principle⁷ (maxims of quantity, quality, relation and manner) are too vague, overlapping and that they are not exhaustive. (cf. Allwood 1995; 2000) . Researchers have added, rearranged or reduced the maxims, like Wilson and Sperber (1987) who reduced the maxims to a single one, that of Relevance. Even so, the general principle holds the message that participants in an interaction adapt their contributions according to the situation in which they are involved, the motives that the speakers might have, and so on. In other words, participants tailor their conversational contributions in accordance with the specific requirements of a particular phase of interaction.

Allwood, along with Grice, argues that communication in many ways seems to build on the *human ability for rational coordinated interaction*. Communicators make certain rational assumptions that guide the choice of how to express ideas as well as how to interpret what other people say or write. As we have seen above, Allwood regards communicators as beings who are perceiving, understanding, and rational. Furthermore, they can be

⁷For further reading, see for example Grice (1975), or Levinson (1983).

seen as rational agents pursuing various motives and goals, some of which are ethical and cooperative, and that they occupy various activity roles. Allwood proposes six principles of communication seen as a species of *rational motivated action and interaction*.

- | | |
|-----------------|---|
| (A) Agency | (i) Intentionality: intent to do things, intentions tied to behavior |
| | (ii) Volition: will to do things |
| (B) Motivation | (i) General: motivation underlies action, ethical, cooperative motives etc. |
| | (ii) Pleasure/pain: the wish to seek pleasure and avoid pain |
| (C) Rationality | (i) Adequacy: efficiency |
| | (ii) Competence: making sure of the preconditions for an action |

The difference between Grice's theory of rational and cooperative communication and Allwood's theory becomes apparent upon examination of the role of ethics. Allwood argues that if you consider other persons by "doing unto others what you would have them do unto you" you make it possible for others to be rational, motivated agents, and are thus cooperating. At the very least, communication involves cognitive consideration, i.e. an attempt to perceive and understand another person's actions. Here, action is defined as behavior that involves intention and volition. Motivation underlies action and often involves the wish to seek pleasure and avoid pain. Rationality is analyzed in terms of adequate (efficient) and competent (making sure of preconditions) action. These principles are involved both when producing (sending information) and perceiving (understanding) messages. The motivation to seek pleasure and avoid pain coupled with adequacy and efficiency may be related to the reduced language use found in many modes of CMC, and will be investigated further below (see, for example, Section 7.5).

2.2.3 Communication and Social/Situational Variables

Allwood argues that there is an interaction between the inherent "meaning potential" of an expression and the use it is put to in linguistic constructions (grammatical structure), communicative functions (expressive and evocative), as well as in joint social activities (occurrence in specific type of social activity). Allwood characterizes social activity by its

type, purpose or function (the reason for its existence). Each activity is of a certain type, and has one or more purposes (obvious/acknowledged goals) or functions (less obvious/acknowledged goals). These purposes/functions give rise to procedures that help define what the activity is all about.

Activities may be pursued for many reasons, and are associated with certain standard activity roles, or standard tasks in the activity, which usually are performed by one person. This role is associated with certain competence requirements, obligations and rights. If instruments and machines are used in an activity, they will create their own patterns of communication. Instruments are necessary in some activities and more ancillary in others. A blackboard may be necessary in some kinds of teaching, whereas a computer and Internet connection are among the necessary instruments for real-time chat. Other physical circumstances influence communication; an example would be when communicating in a noisy environment people have to raise their voices in order to make themselves heard. Hymes (1974) also claims that for some activities, instruments and machines play an important role, and communication is affected by the instrument which is used as a vehicle to mediate it.

Allwood points out that in most human activities, communication plays an important instrumental role, and that a certain degree of cooperation is essential for both activities and communication. Hymes argues that the rules for participants and settings vary according to what the purpose of the interaction is. Hymes argues that the purposes and outcomes of a speech event are the conventionally recognized and expected outcomes: his term "Purposes and Goals" designates the purpose of an event from the standpoint of the community, and need not be the purposes of those engaged in that event (cf. Allwood's view described in Section 2.2.2, above). Hymes points out that the conventionally expected or ascribed outcome or goal must be distinguished from the purely situational or personal. The two aspects have to do with what the speaker intends to accomplish and to what degree he or she does, in fact, accomplish the intended purpose (Wolfson 1989).

Hymes (1974) developed a model to promote the analysis of discourse as a series of speech events and speech acts within a cultural context. Hymes set forth a framework for analyzing and describing the patterns of speaking in a given speech community. His term speech event is intended to be technical and refers to specific activities involving speech. Hymes' taxonomy serves as a starting point for investigation of the way speech is

used in a given society or community⁸. He proposed a set of components of speech that can be subsumed under the acronym SPEAKING. Hymes' SPEAKING taxonomy comprises the categories Setting/Scene, Participants, Ends, Act sequence, Key, Instrumentalities, Norms, and Genres⁹. Hymes intended the components to provide an overall picture of the possible variables that may operate to form the basis of analysis. The components in the list overlap, as was his intention, since some components of speech may be more relevant to some analyses than others (Wolfson 1989).

2.2.4 Participants, Situation and Context

Hymes' term "Setting" refers to the time and place of a speech act. In general, it refers to the physical circumstances in which speech takes place (Hymes 1974). "Scene" is distinct from setting, and designates the psychological setting of speech, or what may be seen as "the cultural definition of an occasion". Hymes points out that the same persons in the same setting may redefine their interaction as a changed type of scene. The scene may change from formal to informal, from serious to festive, etc. Note that cultural aspects do not necessarily affect psychological issues.

Hymes' view of *participants* concerns who is talking and who is listening, and includes four components: speaker or sender of a message; addressor (since in some situations, narration in a book, etc, the speaker is not the same person who actually gives the message, this component is included for use where necessary); hearer or receiver or audience; and addressee. People adapt their verbal language use according to who is talking and who is listening, and it seems plausible to assume that they do likewise in written communication. This seems to be the case in real-time chat, for instance, when participants contribute provocative messages directed to a particular participant, in the full knowledge that all others logged on to the chat may read it. Another example might be teenage girls discussing boyfriends or the past weekend's activities in loud voices in the back of a bus, where the intended audience seems to be the passengers on the bus in general as much as the other participants in the group – if not more so!

Hymes introduces "Key" to provide for the tone, manner or spirit in which an act is done (Hymes 1974). Expressive, or stylistic cues that establish these features may be signaled nonverbally, as with a wink, gesture, posture, vocal emphasis, pauses, etc. Some of the keys that

⁸ Hymes' framework was intended to include both oral and written forms.

⁹ For a good overview of Hymes' model, see Wolfson (1989).

provide expressive cues in traditional written language are punctuation and capitalization. People engaged in computer-mediated written communication seem to have a need for compensating for the lack of vocal and visual cues, and utilize a multitude of strategies to express cues employed in CMC (Herring 1999).

Hymes provides for the specific behaviors that are considered appropriate for different kinds of speaking in different societies. “Norms” are social rules governing the event and the participants' actions and reaction. Norms of interaction concern rules for turn taking, that one must not interrupt, for instance. Naturally, these are associated with certain social structures, or social relationships. Norms of interpretation involve the way different kinds of speech are regarded and understood by members of a given group. Norms are relative to the group and the situation, and are often examples of rational behavior attuned to the situation (Allwood 2000). As mentioned above, Allwood argues that meaning is determined by use in three types of context: perceptual context, social activity and activated background information, as are frequently norms. In the playful setting of recreational chat, the norms might allow participants to interrupt and try to get people's attention, use non-standard spelling and slang, etc. In a serious, formal written communication such as a traditional normative letter, these would not be considered appropriate.

2.2.5 Registers, Genres, Activities

Allwood (1976; 1995; 2000) argues that it is the activity in which people are involved that guides the form of linguistic expression and use. A shift between varieties in the repertoire may signal social intimacy or distance, formal or informal setting, etc. Hymes (1974) points out that humans are skilful in adapting their repertoire to the social context and situation of interaction. This idea is taken to be a comprehensive factor in the use and adaptation of written language to computer-mediated communication.

Hymes points out that “no normal person, and no normal community are limited to a single style of speech, to an unchanging monotony that would preclude indication of respect, insolence, mock seriousness, humor, role distance, and intimacy by switching from one mode of speech to another” (Hymes 1974). Hymes further stresses that a general theory of the interaction of language and social life must encompass the multiple relations between linguistic means and social meaning.

Clark and Brennan (1991) argue that language use is more than the sum of a speaker speaking and a listener listening. It is the joint action that

emerges when speakers and listeners, writers and readers perform their individual actions in coordination, as ensembles.

Biber (1988) identifies fundamental variations between speech and writing, and based his research on texts taken from the LOB (Lancaster-Oslo-Bergen) corpus of written English and the London-Lund Corpus of spoken English. Categories, or “genres” based on external criteria. Biber's fundamental claim is that the frequent co-occurrence of a group of linguistic features in texts is an indication of an underlying function shared by those features. Biber differentiates between “genre” which he uses to describe categorization according to external criteria, and “text type” referring to groupings of texts that are similar with respect to their linguistic form, irrespective of genre categories (Sinclair and Ball 1996). As mentioned above, Allwood approaches communication from the point of view of how current activity and mutual cooperation (Allwood 1976) influence the specific interaction. The question of whether means of expression (i.e. speaking vs. writing) or activity/genre exerts more influence on language use is still open. A discussion of this issue will conclude this dissertation.

As mentioned above, the components in Hymes' list are not mutually exclusive, as was his intention, since some components of speech may be more relevant to some analyses than others (Wolfson 1989) making use of his taxonomy problematic. Hymes' keys, for instance, does not seem to be clearly distinct from act-sequence; instrumentalities overlaps with both act-sequence and keys, genres overlaps with act-sequence, norms and keys. Hymes argues that how something is said is part of what is said. Furthermore, he points out that the means of expression condition or control content (Hymes 1974). Hymes intends content to be a question of topic, or change of topic. Message form and message content are central to Hymes' term “speech act”, and are the focus of its “syntactic structure”. Message form and message content are also closely interdependent. By the term “Genres”, Hymes refers to categories of communication, the kind of speech act or event (poems, prayers, curses, proverbs, form letters, myths, etc.). The notion of genre implies being able to identify formal characteristics traditionally recognized. Hymes points out that often enough, a genre recurs in several events, and often coincides with speech events. The relation between “genre” and “language in an activity” seems to be that Biber uses the term much more to mean “genre in which written language is used”, whereas Allwood's “language in an activity” is used for any kind of language use, spoken, written or signed.

Hymes' framework is intended to include both oral and written forms of communication; his term *channels*¹⁰ referring to whether the medium of communication is spoken or written (Wolfson 1989). Talk may take place face-to-face or by means of the telephone, for example. Writing may be by hand, or typewritten. The forms and styles of speech or writing employ different language codes, varieties and registers.

Written communication may take the form of a formal letter, a personal note, etc. Hymes argues that "speech styles", or forms, applies to any and all organization of linguistic features of verbal means, in relation to a social context. "Speech styles" suggests an aspect of persons, situations and genres, which is highly relevant to the research in this dissertation.

People seem to adapt their speech styles according to channels, relation to other participants and purposes etc, both in spoken interaction and in writing. In written communication, people might use a casual register with many spoken language and dialect features or they might use a more formal register and careful grammatical "standard" forms, depending on circumstances, norms of the situation, activity or community.

Clark and Brennan claim that in language use and in conversation the participants try to minimize the collaborative effort, i.e. the work that both participants do from the initiation of each contribution to its mutual acceptance (Clark and Brennan 1991). It seems that, in general, people do not like to work any harder than they have to. An example to the contrary is pointed out in Allwood's treatise on one-word utterances (Allwood 2002). If people always strove for least collaborative effort, they would not get involved in small talk at all, which indicates that there are other factors involved in interaction than just minimizing effort. Utterances, and indeed interaction, are often multifunctional; humans pursue conversations that demand a fair amount of effort just to pass the time of day. An example from CMC where more than the necessary exertion is used is when people put in a lot of extra effort in composing elaborate graphic messages. There are several possible reasons for doing more than necessary, perhaps to convey something special, to show that you "talk the talk and walk the walk", so to speak, or for the sheer joy of language play.

¹⁰ Channels and forms of speech are joined as means or agencies of communication, and are therefore labeled "Instrumentalities", or *I*, in Hymes' mnemonic.

2.2.6 Communication management

When people communicate they do not just utter their contributions randomly. Even spontaneous, everyday informal conversation follows conventionalized norms and guidelines – speech management – for structuring interaction. Communication needs to be managed in order to reach the goal of “mutual understanding” or “shared information” (in the service of a particular social activity). This management must be attuned to basic physical and biological constraints as well as to the general rational and ethical requirements (outlined in Sections 2.2.2 and 2.2.3). Allwood mentions two types of communication management (*Interactive Communication Management* and *Own Communication Management*, see below) for which there are, in different languages, systematic but partly distinct acoustic and bodily means of giving and eliciting relevant information.

Communication management (CM) functions are linguistic processes and mechanisms whereby a speaker manages his or her linguistic contributions to a communicative interaction, as well as for managing the flow of interaction. Communication management is primarily a feature of spoken, real-time interaction. Allwood assumes two types of functions (Allwood 2000):

Interactive communication management (ICM), which is characterized by linguistic processes and mechanisms whereby speakers manage the flow of interaction with regard to feedback, turn management, and sequences.

Own communication management (OCM)¹¹, which enables the communicator to choose or change his/her message.

First, we examine the three ICM functions.

1. *Basic communication feedback* (Allwood 1976; Allwood, Nivre et al. 1993), which concerns whether sender and receiver are willing and able to:
 - Continue (start or end) communicating
 - Perceive the message which is communicated
 - Understand the message which is communicated
 - Attitudinally and behaviorally react to the message.

¹¹This notion accounts for what has previously been dealt with as hesitations and self-repairs (Allwood, Nivre et al. 1993).

This information communicates their ability and wish to continue, whether or not they perceive and understand, and their reactions are normally evocative of the intention of the previous utterance, for example morphemes like *yes*, *no*, *ok*, *oh* or mechanisms like repetition and pronominal reformulation (see Table 2 below).

2. *Turn management* (Sachs, Schegloff et al. 1974) concerns the mechanisms for distribution of the rights and obligations of communication (more properly, for distributing the sender and receiver roles). Turn management functions include: having-not having a turn, taking (interrupting)-refusing a turn, accepting-giving up a turn, keeping-losing a turn and assigning a turn. Turn management is, thus, conceptually related to the ability and willingness to continue communication, which is also an important part of what is signaled by the linguistic feedback system.
3. *Sequential structuring* (Sacks and Schegloff 1973) (of communicative acts, subactivities, topics, etc.); since this last type of management function concerns, for example, entering or leaving subactivities of an activity, it is often globally determined by the activity as a whole.

ICM mechanisms are features of spoken interaction, depending much on the synchronicity of spoken interaction (face-to-face or over the telephone) in which there is minimal delay between exchanges of contributions. Which kinds of interactive communication management mechanisms might be used depends on the communicative situation and the means for sending information (whether the interaction takes place face-to-face, over the phone, or via letters, for instance). The main ways of giving feedback linguistically are the following (Allwood 2000):

1. Bodily - Mainly head movements
2. Spoken
 - (i) Feedback words like *yes*, *no*, *m* with various phonological and morphological operations allowing expansion of these words (for example *yes* in Swedish: *ja* ->*jaa*, *ja* -> *a* or *ja* -> *ingressive a*)
 - (ii) Repetition of words in a previous utterance to show agreement or to elicit confirmation or more information.
 - (iii) Pronominal or other types of reformulation, e.g. B can agree to A's utterance of *it's raining* by saying *it is*.

Other ways of giving bodily feedback might be through gaze, facial expressions, or shrugging one's shoulders. Traditional written communication such as via letters or newspaper text, seldom displays ICM features, as these are asynchronous modes of interaction. Data from the Swedish spoken language corpus (Allwood (ed.), 2000) and the written language corpus show that the most common feedback and OCM morphemes are those presented in Table 2 below.

Secondly, if we examine OCM, we find that linguistic processes and mechanisms constitute it, whereby speakers manage what to say and, when necessary, change what has just been said. OCM mechanisms are fundamental to enabling individual contributions to collective thinking to be as flexible as possible. Hesitation morphemes, self-corrections, false starts, etc. are examples of how a speaker manages his or her own communication. The Swedish spoken and written language corpus, statistics of which is represented in Table 2 below, is comprised of texts from newspapers and novels. Sometimes, there are sequences of dialogue in newspaper text or in novels; this is where most of the feedback and OCM morphemes are found. The corpus tagger might have misinterpreted some morphemes, which might explain the occurrence in written language of features that are primarily characteristic of spoken interaction. Among other things, this study will analyze the occurrence of feedback and OCM morphemes in four modes of CMC, in an attempt to investigate its written- and spoken-like qualities and attempt to give explanations as to why this is so.

Table 2. The most frequent ICM and OCM morphemes in spoken and written Swedish.

	<i>Spoken language</i>	<i>Written language</i>
<i>Feedback morphemes</i>	1. ja (yes) 2. va (what) 3. (j)a (yes, yeah) 4. nä (no) 5. m 6. m: 7. mm 8. jo (yes contrary) 9. nej (no) 10. jaa (yees)	1. bara (just, only) 2. ja (yes) 3. kanske (maybe) 4. nej (no) 5. just (just, exactly) 6. å 7. bra (good, fine) 8. jo (yes contrary) 9. visst (sure) 10. ha
<i>OCM morphemes</i>	1. e(h) 2. ä(h) 3. e(h) 4. ö(h) 5. ä:(h) 6. ö:(h) 7. (h)m 8. s+ 9. e1 10. (h)m:	1. ö 2. äh 3. öh 4. vetu (youknow) 5. eh

Feedback in spoken and written interaction

Feedback in communication is a vital means for facilitating the flow of the interaction. The speaker can monitor whether his or her utterances are received by the listener, and how the listener reacts to what is being communicated. The listener signals whether he or she can hear, understand and is willing and able to partake in the interaction (Allwood 2000). Spoken, face-to-face interaction has the advantage of simultaneous multimodality (e.g. it is possible to give and receive feedback using auditory and visual – perhaps even tactile - cues simultaneously). Feedback in telephone conversation is monomodal, as speaker and listener cannot see or touch each other, thus all visual and tactile cues have to be left out. Most probably, the interlocutors draw on their previous knowledge and experience of what normally occurs in face-to-face interaction providing them with cues to interpret what cannot be seen but might sometimes be

heard.¹² Feedback in this sense is interactive, requiring the interlocutors to conduct the interaction in real-time, preferably face-to-face.

As written communication normally is an asynchronous and solitary activity (Chafe and Danielewicz 1987) it is not possible to give and receive the same kind of feedback signals. Written communication is normally not conducted "on-line" in real time, but is asynchronous. The communicators are normally separated in time, as well as in space. Written communication was designed and developed to suit other circumstances and conditions than those of spoken interaction. Written language relies on an agreed upon system for representing speech in visual form, mostly representing graphic renderings of spoken words in normative spelling and grammar. Writing cannot be described as "representing speech" altogether for many reasons (Linell 2001/in preparation): spoken interaction involves much more than just words and utterances. The simultaneous nature of spoken interaction allows the interlocutors to be less explicit since much of the information is given nonverbally. Writing, on the other hand, can be utilized for purposes beyond those of verbal interaction. Feedback in written communication is even more restrained than in telephone conversation, relying only on graphic characters which in various combinations represent spoken words and to some extent also give cues to tone of voice and pauses (capitalization and punctuation). So, feedback in writing mostly represents what might be represented by verbal means in a linear, monomodal fashion.

The reasons for how feedback signals are given and received, and indeed why they are used, rely heavily on whether the interlocutors are present face-to-face online in time and space – whether they can see and hear each other in real time - or if the communication is off-line and there is a time delay between producing/sending and receiving/answering (the level of synchronicity and number of modalities that might be used). Cost and effort in production and sending of messages also play an important part.

2.2.7 Establishing shared information

Allwood (2000) argues that there are two main kinds of communicative intentionality involved in communicative acts: the communicative

¹² Results from a study by Nivre and Richthoff (1988) suggest that turn taking in telephone conversations is regulated much more by verbal means, than in face-to-face interaction, just because the non-verbal channels are not available. There is no point in nodding your head in agreement when speaking on the phone, because the listener cannot see you. Even so, many of us have probably experienced doing so many times, just to find oneself having to supply the nod with an explanatory verbal agreement!

intentionality: expressive and evocative. For successful cooperation and interactive communication, it is required that individuals employ individual context and dependent communicative acts of sending and receiving (understanding) in such a way that a kind of cooperation results. There is a basic human social tendency to be contactable for coordination of information in interaction. Communicators have obligations of consideration as well as obligations of responsiveness, and must signal results of an evaluation of whether they are able and willing to continue the interaction, as well as an evaluation of their ability and willingness to perceive and understand what is being communicated. Moreover, communicators are obliged to evaluate the main evocative intention of the previous utterance. A receiver's obligations are frequently combined also with the obligations and conventions that are connected with a particular activity of a particular role in an activity, as mentioned above (Section 2.2.5). The sender and receiver have a number of communicative obligations based on a basic social tendency to be available and contactable for co-ordination of information.

According to Allwood (2000), obligations of consideration entail that a speaker should have the attitude that is being expressed, and a statement, for example, ideally implies that the sender believes in what he or she states (Sincerity). Further, the speaker should have motivation and competence for the communicative act he or she is performing. Making a claim requires some form of evidence for the claim (Grounding). The speaker should also consider whether the interlocutor can and wants to comply with the main evocative intentions of the utterance. For example, in making a claim the sender should consider whether the listener has enough background information or might be in possession of counter-evidence to the claim.

Obligations of responsiveness entail that the receiver should continuously evaluate whether she or he is willing and able to continue, perceive, understand, and respond to the main evocative intention of the contribution. The receiver also has an obligation to report on the result of the evaluation. This report includes providing feedback signals and attitudinal information ("The obligation of responsiveness"). Allwood points out that the receiver's obligations are frequently combined with the obligations and conventions that are connected with a particular activity or particular role in an activity. A person is under different type of pressure when answering a question in a formal job interview than in a casual chat in a bar, for example. Hymes' norms of interaction in combination with setting and scene correspond to Allwood's sender commitments and receiver obligations.

2.2.8 Constraints on Grounding, or Sharing Information

Clark and Brennan (1991) state a criterion for grounding in a different sense than Allwood's term above ("mutual, or shared understanding"): "that our addressees and we mutually believe that they have understood what we meant well enough for the current purposes in our interaction". According to Clark and Brennan, grounding techniques depend both on purpose and medium, and these sometimes interact. Before ever beginning to coordinate content, the interlocutors must assume a vast amount of shared information or common ground, i.e. mutual knowledge, mutual beliefs, and mutual assumptions. To coordinate this process, they need to update their common ground moment by moment (cf. Allwood's obligations, Section 2.2.2). Clark and Brennan point out that grounding is basic to communication, and indeed to all collective actions. He claims that purpose (what people are trying to accomplish in their communication) and medium (the techniques available in the medium for accomplishing this purpose) are factors that shape how grounding works. These factors are not exhaustive, as we shall see.

Allwood argues that by giving and receiving various sorts of feedback signals, communicators monitor continued contact, that what is being expressed is perceived, understood, as well as the reaction to the main evocative intention (Allwood 2000). Clark and Brennan argue that people ultimately seek positive evidence of understanding; acknowledgements (back-channel responses), continuers, assessments, gestures. In conversation, people monitor what their partners are doing moment-by-moment – in particular, what they are paying attention to (Clark and Brennan 1991). This is what Allwood treats as ICM (cf. Section 2.2.6 above).

Further, Clark and Brennan claim a number of constraints on grounding: co presence; visibility; audibility; co temporality; simultaneity; sequentiality; reviewability; revisability. Of course, these constraints differ according to the medium of communication. Face-to-face spoken interaction and traditional written communication, when compared, show almost opposite features, as visualized in Table 3 below.

Table 3. Constraints on grounding (adapted from Clark and Brennan 1991).

	<i>Face-to-face speech</i>	<i>Trad. Writing</i>
<i>Copresence</i>	+	-
<i>Visibility</i>	+	-
<i>Audibility</i>	+	-
<i>Cotemporality</i>	+	-
<i>Simultaneity</i>	+	-
<i>Sequentiality</i>	+	+
<i>Reviewability</i>	-	+
<i>Revisability</i>	-	+

Cotemporality, in the terminology of Clark and Brennan, means that the receiver receives the messages at roughly the same time as it is produced. He points out that in most conversations, an utterance is produced almost simultaneously with when it is received and understood. Clark and Brennan argue, furthermore, that in media such as traditional letters and electronic mail, this is not the case (Clark and Brennan 1991). By simultaneity, Clark and Brennan mean that the interlocutors can send and receive at once and simultaneously. Messages can be conveyed and received by both parties at once, as when a hearer smiles during a speaker's utterance. Clark and Brennan point out that simultaneous utterances also are allowed in the keyboard teleconferencing program called *Talk* (see Section 2.3.3). Split-screen chat is simultaneous as well as cotemporal, both parties' typing appears letter by letter in two distinct halves of the screen (a 2-way protocol in Herring's terms (Forthcoming)). A web chat in which messages are transmitted message-by-message, is cotemporal but not simultaneous (a 1-way protocol in Herring's terms (Forthcoming)). Other researchers have termed this "quasi-synchronous" or "near-synchronous"¹³ (Ferrara, Brunner et al. 1991; Wilkins 1991; Garcia and Jacobs 1999, and others).

The ability to maintain the topic at hand, and change of topic, is part of human communicative competence, and is of particular importance to study of coherence of discourse. Allwood (1976) accounts for cohesion in dialogue by deriving it from different types of communicative functions and obligations. Feedback morphemes and mechanisms are probably the most important cohesive devices in spoken language, whether they occur as single utterances or part of a larger utterance (often in initial position).

¹³ See further in Section 2.4.4.

Clark and Brennan discuss medium and purpose in relation to the costs of grounding (Clark and Brennan 1991). When a medium lacks one of these characteristics, it generally forces people to use alternative grounding techniques. The cost of the various techniques of grounding may change. Clark lists the following costs on grounding: formulation, production, reception, understanding, start-up, delay, asynchrony, speaker change, display, fault, and repair. The speaker pays the cost of formulation and production while the addressee pays reception and understanding costs. Note that these costs are not independent of each other.

It costs time and effort to formulate and reformulate utterances. It costs more to plan complicated than simple utterances, more to retrieve uncommon words than common ones, etc. The act of producing an utterance itself has a cost that varies from medium to medium. It takes little effort (for most of us) to speak or gesture, more effort to type on a computer keyboard or typewriter, and the most effort (for many of us, anyway) to write by hand. People are willing to use more words talking than in typewriting to accomplish a goal, and the faster a person is at typing, the more words he or she is willing to use (Horowitz and Berkowitz 1964).

Listening is generally easy while reading is harder. It also costs to have to wait while a speaker produces a turn. It is more costly to understand certain words, constructions, and concepts, than others, regardless of the medium. Costs can be compounded when contextual clues are missing.

When examining the cost of starting a new discourse one finds that start-up costs are minimal in face-to-face interaction, a bit higher in telephone conversations, and higher still in email communication. In the case of written computer-mediated communication, the sender may or may not be notified that the message is delivered. Once a message is delivered, there is no guarantee that the addressee will read it right away or respond to it. The cost for delaying utterances is high (planning, revision, etc.) in face-to-face interaction, because delays may be interpreted as being the end of a turn, utterances must be formulated quickly. In media without cotemporality – such as email and personal, traditional letters – delays that would be crippling in conversation are imperceptible.

In face-to-face interaction, people time their utterances with great precision, e.g. time acknowledgements to mark what it is they are acknowledging. In media without copresence, visibility, audibility, or simultaneity, timing is much less precise; without cotemporality it is altogether impossible.

The cost of changing speakers is higher in media with fewer cues for turn taking. Costs are quite high, for example, in keyboard

teleconferencing, where the points of speaker change are not as easily marked or readily recognized. Conventions for marking out these points may be developed. One effect of speaker change costs is that people try to do more within one turn. In face-to-face interaction people can point, nod at, or present an object for the other interlocutors. It is also easy to monitor listeners' facial expressions for cues regarding attention, understanding, attitude, etc, as well as with the help of gaze to show who is the intended addressee. In media without copresence, gestures cost a lot, are severely limited, or are out of the question.

There are costs associated with misspeaking or mistakes as well. In conversation, a hearer may expect faults from a speaker because the production of speech is so spontaneous. In written communication, faults are not as easily justified, because the sender has already had a chance to revise them. Some repairs take little time or effort, others take a lot. In media that are not cotemporal, repairs initiated or made by others become very costly, so speakers/senders will try hard to avoid relying on others to repair misunderstandings.

Crystal (2001) argues that if someone leaves out the punctuation in an email, people do not immediately draw the conclusion that the sender does not know his (or her) grammar. Rather, it is more plausible to draw the conclusion that the sender is in a hurry. Crystal argues that people attend to or deviate from traditional norms of letter writing very much depending on the recipient.

2.2.9 Frequency effects and the principle of least effort

As mentioned above, it seems that people do not like to work any harder than they have to. The amount of work one is willing to do depends on the amount of pleasure received or avoidance of pain. Clark and Brennan argue that in language use and in conversation, participants try to minimize the collaborative effort – the work that both do from the initiation of each contribution to its mutual acceptance (Clark and Brennan 1991). This law of minimizing effort cannot be an absolute principle (cf. Section 2.3). Considering how often we are engaged in small talk, for instance, it becomes obvious that there are other principles at work in communicative interaction. One of these principles is the effect that frequent use of certain words and expressions have, a point that will be investigated in text-based CMC in the studies that follow. Dahl (2001) claims that frequent items become reduced because of habituation - less energy is spent on well-learned tasks. He considers the relationship between informational value and effort

spent to be equally important. As the frequency of an item rises, its informational value as an element decreases; hence, less energy will be spent on it. He gives an example of the frequency effect, when a title which is used routinely, for politeness reasons, may be reduced phonetically.

Lexical shortening of frequent words has been noted by Zipf (1932), among others. Words with high usage frequency are shortened compared to lower frequency words. Gregory et al. (1999) suggest that speakers use their knowledge of the probability of a word or combinations of words in sentence production. They claim that highly probable (less informative) words are shorter or more reduced in conversational speech. This is true whether the high probability of the word is based on frequency, collocation with neighboring words, repetition of the word in the conversation, or the semantic association of the word with its conversational context. Gregory et al. combine two explanations that traditionally have been modeled separately. *Frequency models* deal with the fact that words with high use frequency are shortened compared to low frequency words, whether in the lexicon (Zipf 1932) or during phonetic production (Fidelholtz 1975; Bybee and Scheibman 1999). *Predictability models* focus on the fact that words that are highly predictable from the context are shortened during production (Jespersen 1922; Bolinger 1981; Fowler and Housum 1987).

The effect of frequent use, the principle of least effort and time pressure, and the interlocutors' pragmatic knowledge about the world and how communication works are at work when words and phrases are reduced. The principle of least effort (Zipf 1949) may be regarded as normal tendency, but may sometimes be overruled because of other factors, such as the situation demanding clear articulation. Text-based CMC messages show evidence that words are shortened, reduced and even left out. Frequency and probability effects as well as the principle of least effort are among the variables that influence this adaptation; a fact which will be argued in the analyses of the data in the studies that comprise this dissertation.

2.3 Variables Conditioning Human Communication

This section will introduce a number of variables which are taken to condition human communication in general: “synchronicity”, “means of expression”, and “situation”. These variables are grounded in the theories

of human communication in general, and the theories of social and situational factors described above.

Table 4. Variables conditioning human communication in general.

Variables	Conditions
<i>Synchronicity</i>	Time aspect: levels of synchronicity, one-way or two-way communication
<i>Means of expression</i>	Dimensions of expression: primary, secondary, tertiary
<i>Situation</i>	Ability to perceive and understand Physical environment and context Ability for rational, coordinated interaction Type of activity Motives for communication Relation between communicators Pragmatic knowledge Addressee(s) Topic or subject

Synchronicity

Prototypical human communication is face-to-face spoken interaction. Prior to writing, the telegraph, the telephone, and audio recording, human communication took place when people could simultaneously both see and hear one another. The level of synchronicity conditions how language is used and what can be communicated. Both expression and content are dependent on whether communication takes place in real-time or is separated in time. What people say, and how they say it are conditioned by whether other parties are there to interact or not. The variable of Synchronicity focuses on the time aspect, rather than physical location, even though both variables clearly are interdependent. To what extent, and whether, communication can be one-way or two-way has to do with both level of synchronicity and the variable means of expression.

Means of expression

The variable *Means of expression* includes production and perception conditions connected with various means of expression. Allwood acknowledges that the dimensions of expression - the means by which

humans communicate - can be subdivided in many ways. He suggests one possible way:

“Primary means of expression” are means of communication that can be controlled directly without extra aids, e.g. bodily movements, voice, speech, gestures, touch, song etc.

“Secondary means of expression” simply consist of the instruments that are used to augment and support the primary means of communication. They are used to overcome spatial distance and to preserve information over time. Some secondary means directly reproduce primary means, others require several steps to be recoded. Examples of means that directly reproduce primary means are; pen, chisel, megaphone, telephone, while for example the telegraph and email require several steps to be recoded.

When coming to the term *“Tertiary means of expression”*, Allwood admits that one might object that the label “means of expression” is not entirely adequate. He argues that tertiary means are simply human artifacts that are not secondary means of expression. Such artifacts express technical, functional and aesthetic ideas and intuitions. Examples like tables, chairs, houses, roads, household appliances, cars, etc are given. These are rather symbolic than directly communicating. Allwood further argues that tertiary means of expression can often be regarded as collective while primary and secondary means, even if they are also bound by convention, give greater room for the expression of single individuals.

I would suggest that what Allwood labels secondary means of expression should be divided into either two subsets, or that the secondary means which require several steps for being decoded be labeled tertiary means. “Tertiary means of expression” may then be labeled “Quaternary means of expression” in this taxonomy.

The secondary means that directly reproduce primary means reproduce it in different ways: a megaphone strengthens the acoustic signal, while a pen or chisel makes optical, written signs (more or less) permanent which are then able to be carried over spatial distances.

The secondary means of expression require several steps; email – as well as traditional letters - require both mediation from primary means and transmission in order to be recoded. If we focus on the means of *transmission*, we come one step further from the means that reproduce primary means (mediated expressions). I will use the term Tertiary means of expression to indicate CMC.

Electronic means of transfer enables expression on the tertiary level to carry out other kinds of interaction than would be possible on the

secondary level. One example is synchronous interactive written discourse with multiple participants, as in a web chat, for instance.

Situation

I will argue that humans have the ability to adapt to the constraints and enablements of a given communicative situation. As we have seen, there is a complex set of interdependent, finely interwoven variables that condition human communication; at a minimum, language use is adapted according to these.

The variable *Situation* includes the physical, psychological and sociological aspects of communication. Humans have psychological and physical (the physical brain is a necessary requirement for psychological activity) equipment to be able to perceive and understand. Humans also have the ability for rational, coordinated interaction (cf. Section 2.2.2 above). Communication is adapted to suit the conditions of the physical environment and context in which the communication takes place (in a building site, at the kitchen table, in the pub, at a lecture, etc.). This is closely connected to activity and roles in an activity: type of activity fosters the requirements of formality of repertoire, together with roles in the activity, motives for communication (purposes and goals), which individuals are involved in the interaction and whether the communication is one-to-one, one-to-many, or many-to-many. The type of relation that holds between communicators (colleague, friend, lover, stranger, someone you want to impress etc.) Topic or subject of communication is closely connected with type of activity, relation between participants, purposes and goals of communication. Pragmatic knowledge, such as shared background information, conventions of communication, ethical conduct, etc., also influences how messages are formulated.

To sum up, the section above has dealt with general human communication and socio-cultural factors. Theories of human communication need to account for the various levels of organization involved in communication. Hymes (1974) claims that a theory must consider the context, or setting, participants involved, purposes, instrumentalities, genres, etc. Allwood discusses what he terms multi-layered constraints and enablements such as physical, biological, psychological, and social levels of organization. Allwood also argues that communication in many ways builds on the human ability for coordinated (cooperative) interaction (Allwood 2000). This ability was analyzed using

six principles regarding communicators as rational, motivated agents. Type, purpose and function, together with activity roles of the communicators and instruments characterize a social activity. Interlocutors need to establish a degree of contact, perception, understanding and reaction for interaction to flow.

The theories reported above will be used as a background to support the analyses of the material and evaluation of the results in the present work. The claims concerning factors that affect communication (among the most important are: context; participants involved; purposes and activity; instrumentalities; genres) will be revisited in connection with the studies that comprise this dissertation. Frequency and probability effects as well as the principle of least effort dealt with by Clark and Brennan (1991), Zipf (1932) and Dahl (2001) are among the factors that influence the adaptation of language use, which will be argued below.

2.4 Spoken and Written Communication

From considerations of human ability to language and human communication in general, we now turn to more specific aspects of human communication, namely written and spoken communication. As pointed out in the introductory chapter, CMC has often been treated as a sort of hybrid between spoken and written communication, employing features that are characteristic of both modes. In order to investigate spoken-like and written-like features of the CMC modes included in this study – and to look into possible features that are specific to CMC - it is necessary to get an overview of the findings of relevant and important previous studies of written and spoken language. Recall that CMC cannot be treated as a single form of communication as there are many modes that have different constraints and enablements, as well as being used for many different purposes.

2.4.1 Speech and writing: manifestations of the same system?

Several approaches to the relationship between speech and writing have been taken in linguistic study over the centuries. Wengelin (2002) summarizes three different views of the relation between spoken and written language that have been taken throughout history. In the west, written language has been seen, since antiquity, as primary while spoken

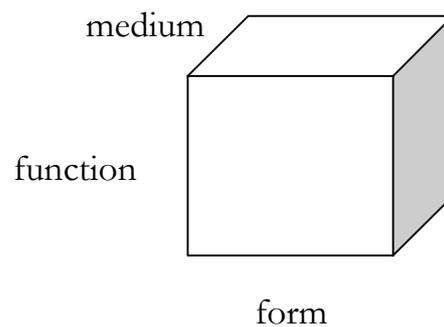


Figure 1. Three interrelated aspects of language: medium, function, and form (adapted from Halliday, 1985: 78).

language has often been seen as dependent and even derived from the former. Not all of the ancient philosophers held that written language is primary, though, Plato argued that spoken language is primary. This idea has been put forward by scholars of the 20th century as well, among them Saussure (1916), Bloomfield (1933) and Jespersen (1922). Bloomfield in fact did not regard written language as language at all but rather held that:

Writing is not language but merely a way of recording language by means of visible marks.

Bloomfield argued that writing is just an external device that linguists use for observing features of speech of past times. However, it does not seem that Bloomfield believed in this sincerely, since the idea stands in contrast with what he actually practiced.

A third possible way to view the relationship between speech and writing is to regard them as alternative forms of language in their own right (a view taken by for example by Chafe (1994) and Halliday (1985)). This view regards spoken and written communication as each having its own validity, without putting one manifestation of language before another. Both speech and writing may be seen as effective adaptations to suit the situation in and goals for which they are used, an idea we will come back to many times in this study. Halliday (1985) argues that one has to think about both written and spoken language in terms of three interrelated aspects: the nature of the means of expression, the functions served, and the formal properties displayed, as pictured in Figure 1 below. This is the view taken in this dissertation.

As we shall see, there are no straightforward differences between speech and writing that hold for each and every instance of language use (see Section 2.3.). Cook (1997) claims that the overall differences relate to the actual physical medium through which speech and writing are transmitted, to the memory systems by which they are processed, and to the functions for which they are used. The different media of sound waves and optical signs affect how they may be used to carry expression, but the differences are not always clear-cut. Genres, purposes or communicative activity seem to be more important than means of expression. Halliday also points out that the distinction between speech and writing is becoming blurred as a consequence of modern technology (Halliday 1985).

Spoken language has been characterized as being less structured and organized than written language, as it is being created on the spur of the moment and does not allow for extended planning in order to pack complex nuances neatly. Halliday notes that there is a tradition of regarding spoken language as formless and featureless (Halliday 1985). He points out that, in fact, spoken language is no less structured and is as highly organized as written. Furthermore, he claims, it could not be otherwise since:

*Both are manifestations of the same system. The two are both **language**; and **language is more important than either**. It is a mistake to become too much obsessed with the medium.¹⁴*

This thesis concurs with Halliday's idea that both speech and writing are manifestations of the same system, but it will be argued that means of expression, situation and synchronicity are such important factors that they cannot be ignored. Speech, certainly, is not unstructured or superficial. Halliday (1985) argues that spoken language tends to be disvalued in cultures that are strongly dependent on written language because it is not the primary means of access to power. Linell (2001/in preparation) points out that

The conditions under which written language is generally taught have promoted the quite common belief that (some variants of) written language represent(s) the "grammatical" correct language, whereas many variants of spoken language are incorrect, defective, incoherent, ugly or rude.

Cook (1997) reminds us that it is easy to fall into the trap of treating spoken language as if it were a poor relation of written language. Speech

¹⁴ Sic.

was not meant to be written down, and most written text does not sound like spontaneously produced spoken language if read aloud (Halliday 1985).

Having touched above upon theories that recognize register, genre and stylistic variations in previous sections in this chapter (see Section 2.2.5) it is obvious that there are variants of written language as well as variants of spoken language. Comparisons between spoken and written language show that the two means of expression show some similarities and some differences in patterns of features. In order to investigate the spoken-like and written-like features in a number of modes of CMC, the point of departure for the analyses in the present study has to be established with respect to previous research on spoken and written language.

2.4.2 Speech and writing, spoken and written language

Speech and writing, according to Linell (2001/in preparation), are two rather different ways of representing our experience. Spoken language represents phenomena as processes, written language represents phenomena as products, which seems to be a rather simple way to put it. As mentioned above, the different media of sound and writing affect how they may be used. The permanency of writing leads to advantages for storage and for careful use of language, which, in turn, have led to its status within society as a mark of education. The swiftness and immediacy of speech are advantageous for uses that emphasize social relationships and flexible interaction (Cook 1997). We will now turn to features, or properties that have been found to be characteristic of typical speech and writing.

Biber (1988) uses the notions of ‘stereotypical’ speech and writing to refer to the most frequent types of speech and writing, which are face-to-face conversation and informational written exposition. Chafe and Danielewicz (1987) choose informal conversation and formal academic writing to exemplify the extreme ends of the scale of typical speech to typical writing. Linell (2001/in preparation) argues that speech is linguistically structured speech behavior which operates together with non-verbal signals, various kinds of background knowledge that the speaker and the listener have, the listener’s responses as well as other characteristics of the physical and social context in which the communicative activities (cf. Allwood 2000) are embedded. In addition, both production and comprehension of speech are transient events that partially overlap and occur at very high rates of speed.

2.4.3 Variables Conditioning Speech and Writing

Table 5. Variables influencing the conditions of prototypical speech and writing.

<i>Variables</i>	<i>Conditions</i>	
	<i>Speech</i>	<i>Writing</i>
<i>Synchronicity</i>	Synchronous Dialogue, two-way interaction Process: continuous movements Simultaneous channels Time pressure Linear time	Asynchronous Monological, one-way (non-interactive) Single channel No time pressure Non-linear time
<i>Means of expression</i>	Multimodal channels (“rich mode”) Acoustic (speech) + visual (nonverbal) Ephemeral, fleeting Light physical and cognitive effort	Monomodal channel (“lean mode”) Optical visual (written signs) Persistent, permanent Physical and cognitive effort
<i>Situation</i>	Background knowledge Context and situation dependent	Autonomous, independent Lacking immediate context

Synchronicity

Speech

Synchronicity is the relationship between different things in time. Typical spoken interaction is spontaneous, informal everyday conversation. Interlocutors (speaker(s) and listener(s)) in this type of communicative situation are present in time (synchronous) and space (face-to-face). Contributions are perceived as they are produced (two-way interaction). As a consequence of the synchronous variable, communication is dialogical and often allows for interaction¹⁵. The communicators send information through several channels (acoustical and optical) simultaneously: speech

¹⁵ Broadcasted talks on the radio are also to be regarded as communication, but not interactive communication. In this sense listening/interpretation/understanding is not regarded as an active part of the communication.

and nonverbal signals. Interaction, or dialog, requires at least two participants.

Spoken interaction is constrained by the pressure of time: the speaker must produce his or her utterances quickly and readily and the listener must respond just as rapidly, under the pressure of the emotive and social atmosphere of face-to-face communication. Time pressure gives speakers less time to plan; speech is characterized by "draft version" formulation, in which features of the interlocutors' communication management (pauses, hesitations, self corrections, etc.), or editings are frequent. Lexical density is the proportion of content words in a text. It tends to be lower for spoken than for written language. Lexical density is likely to be approximately twice as high in writing as for speech (Halliday 1985). Halliday also suggests that spoken and written language also differ markedly in formality; lexical density might be a measure of formality.

Similarities between speech and writing were found when comparing word frequency and frequency of parts of speech in spoken and written Swedish (Allwood 1998). Thus, for example, word types were found to be mostly the same in both speech and writing. "Syncategorematic" words, or closed class words, were most frequent in both modalities. Differences were that the number of word types was smaller in speech than in writing, that high-frequency expressions had a relatively higher frequency in speech than in writing. The most common words were different between the two modalities, and speech contained a number of features such as feedback, OCM and interjections, which are more associated with dialogic forms of interaction than with writing. The most frequent part of speech in spoken Swedish was found to be pronouns, followed by verbs and adverbs, while the most frequent counterparts in written Swedish were nouns followed by verbs and pronouns (cf. Wengelin 2002).

One can *do* things simultaneously, but senders must necessarily send and perceive words *in temporal order*, which means that both speakers and writers have to "solve the linearization problem" (Levelt 1981), i.e. put ideas into linear form. You cannot say, or write, everything at exactly the same time. Levelt argues that this is a problem that both speaker/listener and writer/reader have to overcome. Information conveyed in speech or writing has to be structured according to two powerful constraints on linearization that derive from *mutual knowledge* and *working memory limitations*. The speaker utters strings of sounds, which he has to order in such a way that the listener has a chance of grasping the meaning. The fact that words follow one another in real time to form messages seems inescapable. It also seems that the "working memory" of a listener/reader cannot cope with too much for too long, or she or he will lose track of the message.

Therefore the speaker/writer has to use strategies like little ‘pegs’ on which to hook information. E.g. anaphoric references can then be used as “return addresses”. It is the same with writing.

Interaction in face-to-face spoken communication means that the communicative acts and contributions are integrated in the whole complex network of interdependent variables, and are not isolated in time, space or context. The variables *synchronicity* and *means of expression* are naturally interdependent: communicators monitor interaction continuously using information sent simultaneously through several channels. Further, contributions are adapted according to factors such as relation to other communicators, goal of interaction, surrounding context and environment, etc.

Writing

Traditional written communication is asynchronous; rapid responses are not necessary. Reading is faster than listening: considerable sections may be scanned (almost) simultaneously or at least repetitively, while a listener has to hear a speaker out. Speech is much more linear in time, it is not possible to “fast-forward” real-time speech-signals. Typical written communication is the formal letter or academic text. The communicators (writer and reader(s)) in this type of communicative situation are separated in time (asynchronous) and also in space (distributed), which makes written communication monologic, or at least, interactive with a substantial time delay. Linell (2001/in preparation) and Cook (1997) both point out that written texts are typically not perceived and interpreted at the same time and place as they are produced. Writing is thus in this sense one-way (Cook 1997), and non-interactive. Biber (1988) points out that there are severe real-time constraints in speech, but none in writing as the writer has time to choose words and expressions, and to revise. This normally holds for traditional writing, but there might be situations during which there definitely are severe time constraints on even traditional writing. Consider the time pressure when writing for a deadline, for instance. The freedom of time to revise and choose leisurely between words and expressions no longer holds for interactive written communication.

Means of expression

Speech

Spoken interaction is a means of communication that is multimodal, speakers use several channels to send information through. Both acoustical and optical signals can be utilized. Usually, the listener and speaker are physically present in the same speech situation; they can see each other and are aware of what is going on. Speech can be produced very quickly. Spoken interaction requires relatively little effort i.e. low physical cost (Allwood 2000), and cognitive effort.

Both production and comprehension of speech are transient events that partially overlap and occur at very high rates of speed (Linell 2001/in preparation). Both spoken and nonverbal signals are ephemeral due to the temporary nature of acoustic waves and visual information from bodily movements. Speech is dynamic, ephemeral behavior distributed in time; it proceeds continuously and its inherent dynamics, its changes at various levels, must be subject to on-line monitoring and analysis by both communicating parties. As time proceeds participants can no longer observe that which was produced earlier.

Writing

Writing has been described as a “lean” medium. Writing is constrained by its lack of channels and cannot be used for sending simultaneous multimodal information. Writing is, by nature a monomodal mode of communication, communicators send information through a single channel, or medium (optical). Writing is also constrained by the time and effort that production of messages costs. Written signs are persistent and may be reread and stored for the future. The absence of immediate context must be compensated for, i.e. referents must be fully described, and arguments must be represented more extensively. A written text is also relatively autonomous: it is not dependent on the surrounding situation, especially not the physical environment in which it is read.

Speech behavior has many features of continuous movement (rather than a chain of successive states). A written text is made up of discrete symbols, i.e. letters (at least in print) and (graphic) words which are organized in certain regular spatial patterns (according to syntactic rules as well as various conventions of punctuation and paragraph division). These symbols are the approximate counterparts of only some of the structural

(i.e. segmental: phonological, grammatical, lexical) features of spoken language. In an attempt at marking out in writing prosodic features and the non-verbal signals in speech situations, graphic means such as capitalization, underlinings, italics, etc., are used.

Biber (1988) argues that the relation of communicative participants to the text has to do with the degree of permanence of the text; whether they have opportunity for interaction with the text during production (planning or revising) and in comprehension (one-way or two-way interaction). The speed of production is connected with the type of text input technique (Horowitz and Berkowitz 1964).

As facility increases in the writing mode, cognitive and linguistics indices do approach those previously found in spoken expression, but still differ significantly from them.

According to Horowitz & Berkowitz' study, some of the many differences found between spoken and written expression are due to the greater ease of speaking one's mind than writing it. Any mode of writing¹⁶ that increases the ease of production of written content should result in the production of cognitive and linguistic material more nearly like that produced in spoken expression.

Situation

Speech

The whole interaction between speaker and listener is dependent on the situation (context). Speakers' speech behavior is continuously accompanied and supplemented (occasionally contradicted) by various non-verbal signals, which means that the verbal message as such is often much less explicit than in writing. An utterance in a normal situation involving face-to-face interaction is not an isolated speech act: it is part of an integrated communicative act which comprises the use of a range of semiotic resources, including spoken language and non-verbal communication (gaze, facial expressions, gestures and body movements etc.) as well as aspects of the surrounding material. The message is conveyed in several ways simultaneously.

Prosody – rises and falls in pitch, accents, pauses, rhythms, and variations in voice quality – is a salient feature of spoken language, but is

¹⁶Horowitz & Berkowitz compared hand writing, type writing and steno typing in 1964.

not fully represented in written language. Chafe (1988) explores the relationship between what he calls the covert prosody of writing and the principal device that writers use in order to make it at least partially overt, punctuation. Chafe and Danielewicz' (1987) results from analyses of the scale of informal conversation to formal academic writing show that typical speech employs a relatively limited vocabulary, as mentioned above, and is inclined to hedge lexical choices, as well as to be referentially implicit. Colloquial words and phrases are frequent during conversation.

Writing

A written text and its component parts (letters, words, sentences, paragraphs, etc.) have the character of objects; they are constant and static, as well as atemporal (Linell 2001/in preparation). Face-to-face speech may rely on shared contextual information and often concerns first and second person in the here and now. In spoken interaction people can be rather implicit, and do not have to explicitly refer to things that the interlocutors have in fresh memory or to things in the immediate context. Writing cannot rely on information from a shared context, multimodal signals and constant feedback. It has to be more explicit. Writers have time to plan and edit the product, of which the final product shows no traces.

Written texts typically lack an immediate context. Written text is, as a rule and in comparison with spoken utterances, relatively explicit; the writer must compensate for the absence of immediate context, i.e. referents must be fully described and arguments must be represented more extensively. A written text is also relatively autonomous; it is not dependent on the surrounding environment, especially not the physical one. The medium of writing is adapted for a monological function.

Paolillo (1999) approaches communication from a sociolinguistic point of view, and argues that people in regular contact with one another tend to share more linguistic features, and tend to borrow more features of each other's language varieties, even in situations where those varieties are different languages. Likewise, people who have less contact with one another tend to share fewer linguistic features with one another. The variable *situation* incorporates variety and frequency of contacts among people. The relationship between communication participants is another important aspect of the communicative situation. Biber (1988) lists major situational parameters, among them are the relation of communicative participants to each other, the relation of communicative participants to the external context, and the primary purpose of communication (Biber's variables *Physical channel* and *Relation of communicative participants to the text* are

referred to under the variable *Means of expression* above). *Situation* also incorporates *activity* as an ingredient in how the situation is shaped and the reasons for communication being pursued (cf. Section 2.2.5 above).

2.4.4 Are differences between speech and writing due to means of expression, or genre?

Differences between language use in spoken and written communication can in part be explained by examining the communicative situation with its different production conditions, the aims and purposes of communication, and relationship between communicators, as outlined in the variables above (Section 2.3)

Table 6. Linguistic consequences of influencing variables on typical spoken and written language.

<i>Speech</i>	<i>Writing</i>
First draft status	Final draft/product (no traces of editing)
Structurally simple	Structurally complex, elaborate
Fragmented	Organized and planned
Concrete	Abstract
Implicit, exophoric (or situation dependent) reference	Explicit, endophoric (or situation independent) reference
OCM and ICM features	Decontextualized, autonomous
Intonation and prosody	Subordinations, passives
Lexical uniformity	Lexical diversity
	Density in content
	Punctuation (conventions and rules)
	Neutrality of social roles

There is no linguistic or situational characterization of speech and writing that is true of all spoken and written genres. Biber uses the notion of ‘typical’ speech and writing to refer to the most frequent or common types of speech and writing (Biber 1988): face-to-face conversation and informational exposition as typical writing. Biber draws the conclusion, after investigating differences between speech and writing, that there is no single absolute difference between the two. Rather:

There are several dimensions of variations, and particular types of speech and writing are more or less similar with respect to each dimension.

It is likely that Biber's study is biased towards written language, due to its focus on the words uttered in spoken interaction or written in writing. Biber's analysis of spoken and written language showed no differences concerning the linguistic features he chose to investigate. He disregarded that, through speech, communicators can communicate via other means than merely words. Spoken interaction also involves sounds and gestures, definite examples of differences between speech and writing. Biber's study is a bottom-up analysis of empirical data in which variables that suit written language are applied to spoken language data, ignoring features of spoken language that have no resemblance in writing: intonation, sound, overlap, hesitation, pauses. Overlaps and hesitation are denoted in writing by punctuation, a fact that is ignored in Biber's study.

Biber views the difference between speech and writing to be an effect of genre, and not medium. These ideas are then justified through his choice of variables and corpora. Biber argues that among the situational differences between typical speaking and writing, only two of them approach absolute distinctions between speaking and writing: the channel difference (vision and talk) and the opportunity for interaction with the text. As mentioned above, spoken interaction is multimodal.

Even during the time that Biber conducted his study, people used graphical means to indicate information other than that which can be conveyed through the lexical-syntactical sub channel in writing. People use several means to convey this, for example through different font sizes, bold and italic face, pictures, logotypes, writing words in all capitals or all small-case, elaborated use of punctuation marks, etc. This is taken even further in CMC.

Chafe and Danielewicz (1987) discussed important linguistic features that characterize different types of spoken and written language from dinner conversations to academic papers (shown in Table 7 below). Taking into account the cognitive and social demands made on speakers, listeners, writers, and readers in their interactions, they analyze the reasons for these language differences. Chafe and Danielewicz argue that:

*Language users, whether they are speakers or writers, have certain variable resources available to them. Every speaker or writer possesses a repertoire of devices which are combined in varying mixtures depending on context, the purpose, and the subject matter of language use. In other words, **language adapts to its varying environments**.*¹⁷

Chafe and Danielewicz analyzed material from two genres of speech: conversationalists (i.e. those adept at conversation) and academic lecturers and two of writing (letter writers and academic writers) for variety of vocabulary, level of vocabulary, clause construction, sentence construction, involvement and detachment.

Conversationalists, at the spoken and informal end of the scale, exhibit properties that are appropriate to language that is produced rapidly in an environment where the immediate presence of the audience plays an important role. The language of academic lecturers, which represents less informal speech, is still controlled by the constraints of rapid production. They were found to strive after some of the elegance and detachment of formal writing. Letter writers take some advantage of what the deliberateness of writing allows. Interestingly enough, being on the casual end of the writing side of the scale, letter writers also maintain, and sometimes even surpass, the casualness and involvement of speech. Academic writers at the other far end of the scale, represent language use that shows a maximum adaptation to the deliberateness and detachment of the writing environment.

The fact that letter writers sometimes even surpass vocabulary casualness and involvement of speech is interesting in the light of CMC research. It resembles findings from analyses of email messages (Maynor 1994; Cho Forthcoming), suggesting that users attempt to imitate an informal 'oral style'.

¹⁷ Bold face mine.

Table 7. Chafe & Danielewicz report their findings in comparing spoken and written interaction. (Adapted from Chafe and Danielewicz 1987).

	<i>Speakers (conversationalists)</i>	<i>Academic lecturers</i>	<i>Letter writers</i>	<i>Academic writers</i>
<i>Vocabulary</i>	Employ a relatively limited vocabulary	Limited vocabulary (equal to conversationalists)	More varied	Maximally varied
<i>Hedges</i>	Hedge their lexical choices	Hedges	Rarely hedges	Avoid hedges
<i>Reference</i>	Are referentially implicit	Referentially implicit	Sometimes implicit	Avoid implicit references
<i>Style</i>	Considerable use of colloquial words and phrases	Literate vocabulary (somewhat greater than conversationalists)	Moderate number of colloquial words or contractions, a greater number of literary items	Maximally literate; no colloquial items or contractions
<i>Intonation units</i>	Create relatively brief intonation units	Extended length and syntactically elaborate sentences	Intermediate between conversationalists and academic writers	Maximally long and coherent
<i>Interaction</i>	Interact with their audiences	Interact slightly with audience	No interaction	No interaction
<i>Involvement/detachment</i>	Show ego involvement	Some use of first pronoun	Greater degree of involvement with ego than all other language users (in this study)	Little ego involvement
<i>Concrete/abstract</i>	Talk about specific times and places	Some concrete spatio-temporal references	Greater degree of involvement with reality than all other language users (in this study)	Little involvement with concrete reality
	Language that is produced rapidly in an environment where the immediate presence of the audience plays an important role.	Mixed kind of language: controlled by the constraints of rapid production, but striving after some of the elegance and detachment of formal writing	Takes some advantages of what the deliberateness of speech allows, but maintains the casualness and involvement of speech	Maximum adaptation to the deliberateness and detachment of the writing environment

2.5 Computer-Mediated Communication

December (1996) gives a concise definition of CMC by saying that it is the asynchronous and synchronous creation and transmission of messages using digital techniques. Herring (2001) argues that the study of computer-mediated discourse (CMD) is to be regarded as a specialization within the broader interdisciplinary study of computer-mediated communication. CMD is to be distinguished from the broader field of CMC by its focus on language and language use in computer-networked environments, and by its use of methods of discourse analysis to address that focus. Crystal (2001) makes the same distinction, but uses the term ‘electronic discourse’ to emphasize interaction and dialogue elements, claiming that ‘CMC’ focuses on the medium itself. Herring defines CMD as

Computer-mediated discourse is the communication produced when human beings interact with one another by transmitting messages via networked computers.

Messages are typed on a computer keyboard and read as text in a computer screen, typically by a person or persons at a different location from the message sender. Crystal (2001) identifies five broad text-based Internet-using situations. He claims that these are sufficiently different to cause the language they contain to be significantly distinctive. The five situations are:

- electronic mail (e-mail)
- asynchronous discussion groups (bulletin boards)
- synchronous real-time chat groups
- virtual worlds
- world wide web (www)

The linguistic properties vary depending on the kind of messaging system used and the social and cultural context embedding particular instances of use. What these modes¹⁸ have in common is that activity that takes place through them is constituted primarily, - in many cases exclusively – by visually-presented language (Herring 2002). These characteristics of CMC

¹⁸The five situations are not mutually exclusive. Emails often contain web attachments, and some real-time chat rooms allow participants to contact each other via email as well, for example.

have important consequences for understanding the nature of computer-mediated language. Crystal (2001), like Ferrara, et al. (1991), argues that many of the expectations and practices associated with spoken and written language no longer apply. He points out that

[...] there is no way of predicting whether this new language-using situation will make use of old conversational norms or invent fresh stylistic techniques to facilitate interaction, or what particular combination of new and old will prove the most effective.

This claim will be investigated below.

As mentioned in the introductory chapter, the development of CMC technologies is exceedingly swift. The rapid progress of technologies will make CMC change continually more quickly than scholarship can publish (Soukup 2000). Soukup argues that users, as well as software and hardware manufacturers, seem to be shifting from the text-based use of the Internet toward the new, three-dimensional, multimedia based world wide web (Soukup 2000). Multimedia CMC involves communication and information with audio and video in chat rooms, web pages, email and mobile phones. Most CMC used today, however, is still text-based. The present study is investigating text-based CMC; the different conditions of multimedia CMC and its effects on language use will be left to future analyses. For the sake of simplicity, in the present work, the term CMC will henceforth refer to text-based computer-mediated communication.

2.5.1 Does CMC have a negative influence on language?

According to popular opinion, the Internet and CMC will be bad for the future of language. People are anxious that a sort of "Technospeak" will take over - in analogy with "Newspeak"¹⁹ as envisaged by Orwell (1949) in the novel "Nineteen Eighty-Four" - and that standards and norms of traditional written language will be lost, and creativity and expressiveness will be diminished as globalization imposes uniformity (Baron 1984; Crystal 2001).

It has often been claimed that the language of CMC is becoming less expressive and less sophisticated than other forms of writing (cf. Hale

¹⁹ The purpose of Newspeak was to drastically reduce the number of words in the English language in order to eliminate ideas that were deemed dangerous and, most importantly, seditious to the totalitarian dictator, Big Brother and the Party. For further understanding, see (Orwell 1949).

1996, among others). The language in e-mail, for instance, is said to be structurally simpler than traditional forms of writing. It is supposedly made up of shorter, grammatically less complex sentences, and contains more sentence fragments and typographical errors (Hale 1996). However, Herring reports (1998) that she found that the language in e-mail messages posted to professional discussion lists tended to be linguistically sophisticated, making use of complex grammar and containing few errors. Demographic factors and purpose of communication are important factors in conditioning linguistic expression. It must also be argued that the situation in which these messages were posted is very special, and one could well expect that type of activity (i.e. asynchronous communication among academics) to influence the messages (cf. Section 2.2.5 above). Ko (1996) found synchronous CMC to be simpler even than spontaneous speech in terms of range of vocabulary used and measures of word and sentence length. Addressing the claim that CMC uses less expressive and sophisticated language, Herring (2002) claims that:

Social meanings appear to be conveyed effectively through CMC. Users achieve this in part through creative uses of language, such as novel spellings, repeated punctuation, and ASCII graphics designed to convey attitude, non-speech sounds and facial expressions.

This is especially common in synchronous chat, despite the fact that expressive language often requires extra keystrokes, and thereby goes against the principle of economy of effort that otherwise conditions chat language.

Du Bartell (1995) argues that we expect written language to be edited, planned, articulated without recourse to non-standard constructions, slang and vulgar expressions. From speech, we expect more or less the opposite: we expect slang, non-standard grammatical constructions, sudden topic shifts and spontaneity. “CMC gives us these in writing. CMC discourse exhibits the type of grammatical constructions that appear in non-edited non-standard spoken language of face-to-face interaction” (Du Bartell 1995).

Crystal (2001) argues that the Internet enables a dramatic expansion of the range and variety of language and provides opportunities for personal creativity. Crystal argues that CMC is a hybrid of speech and writing, and is not concerned that English will be ruined by its often quite casual treatment. On the contrary, he argues that children who spend their days sending instant messages are in no danger of becoming illiterates. Crystal points out that people are aware of stylistic differences in language use.

Language use differs according to purpose and activity, which I believe is a plausible argument. People looking for a job will construct their e-mail quite differently than they would if emailing with friends.

2.5.2 A third medium?

Ferrara et al. (1991) examined electronic discourse and described it as an emergent register called Interactive Written Discourse (IWD) with unique features. Ferrara et al. view IWD as a “hybrid” language variety, which displays characteristics of both oral and written language, resembling speech and writing, but which is neither. Like most language varieties, it arises out of a new social context in which a new appropriate form of expression is needed. Therefore, language users draw on existing language varieties to form hybrids. One of the most important features which illustrates how IWD is both like written and spoken language is that it is simultaneously “interactive” and “edited”. Ferrara et al. view CMC as changing not only the way people think about the possibilities of communication, but also the ways they use written language, compose in it, and interact through it. Part of their conclusions are based upon the syntactical reductions that often characterize CMC, with initial pronouns and articles sometimes being omitted (cf. Zipf 1932; Dahl 2001).

Wilkins (1991) also notes that synchronous CMC bears a resemblance to spoken discourse in that markers of personal involvement (e.g., use of second person pronouns and names) along with innovative language use (e.g., plays on words) frequently characterize the corpus she studied.

Herring (2002) notes that computer networks are often considered a medium of communication distinct from writing and speaking. Thus CMC researchers speak of electronic ‘medium effects’ on CMC, rather than treating CMC as a form of ‘writing’ (typing) that happens to be distributed by electronic means. The means of production in CMD is similar to other forms of typing, including allowing for editing and formatting of text in asynchronous modes. Other aspects of computer-mediated communication preclude easy classification with either writing or speaking. CMD exchanges are typically faster than written exchanges (e.g., of letters, or published essays which respond to one another), yet still significantly slower than spoken exchanges, since even so-called “real-time” typing is slower than speaking.

Crystal (2001) argues that CMC is developing into a new medium that shows language users at their most inventive, adapting a variety of styles for a variety of purposes, of which some are formal and some are highly informal. However, Sveningsson (2001) reminds us that:

Computer-mediated communication should not be seen as one homogenous type of communication. Just like other communication media, it manifests itself in various styles and genres, differing both between types of Internet media and within the same type of medium. Some of their properties depend on what technology is used, while others are related to human aspects such as the purpose of communication, groupings and subcultures.

According to Crystal, CMC is fundamentally different from speaking and writing; it shares in their properties, but does something neither could possibly do. Crystal claims that while CMC shares some of the spoken word's transience, it offers other traits, including simultaneity not possible in spoken conversation. There is a hypothetical possibility in Crystal's claim that one, "can have a conversation among 20 people in a computer chat room, something not even the most adroit person could accomplish at a cocktail party". Empirical evidence shows that even skilled chat participants have trouble keeping up more than a few conversational threads. It is easier, though, to keep up conversation with several parties in interactive written discourse where the contributions are persistent rather than ephemeral, as in spoken interaction. Crystal argues that the electronic medium presents us with a channel which facilitates and constrains our ability to communicate in ways that are fundamentally different from those found in other semiotic situations (Crystal 2001). Like Ferrara et al (1991), he views CMC as a hybrid between speech and writing. Crystal suggests further that online language is best viewed as neither written nor spoken language, but rather as a new species of interaction, a "third medium", which is in the process of evolving its own systematic rules to suit new circumstances.

CMC is better seen as written language which has been pulled some way in the direction of speech than as spoken language which has been written down (Crystal 2001). Crystal claims that the answer to why one should regard CMC like this lies in the immediacy of computer-mediated communication. The new medium is different from writing in its immediacy and changeability, Crystal argues, and different from speech in its inability to provide pitch, rhythm, loudness and other voice cues. The evolving discourse of the Internet is quite different from writing, in part because writing's prime characteristic is its stability. Traditional written texts stay in place and remain the same, so that it is possible to refer to a page read earlier. One would be very surprised if the page had changed its character. That is not true for computer-mediated communication, which reflects a characteristic fluidity. A web page might be changed or the text of an email might be reused and cut and pasted to create a new message.

Crystal argues that the analogy of online communications using speech rather than formal writing is too simple: chat rooms, IRC and the like are too constrained by their response times and the slow speed of typing to be considered as a good analogy of speech; Web pages, e-mail and other mechanisms are too transient or easily modified to be equivalent to the printed word. People adapt spelling, grammar and semantics to meet the needs of Internet-based situations. Crystal contends that people attend to or deviate from traditional norms of letter writing depending on the recipient (Crystal 2001). He agrees that much of the language of CMC is non-standard, playful, highly deviant from the normative rules of language, tolerant of typographic and spelling errors, and full of new words. This is viewed as being due to a desire for variety and innovation, and as being the result of a set of language tactics developed for a new medium that have evolved as users have adapted their language creatively to meet changing circumstances.

December (1996) argues that there is no need for CMC researchers to get too wrapped up in splitting hairs about what is CMC and what is not. CMC in the broadest sense involves a wide range of telecommunications activities as well as non-networked transfer of information such as via computer diskettes. Ultimately, the definition of an activity as computer-mediated relies on its value for shedding meaning on the communication act for its validity.

As the aim of this dissertation is to investigate how written language is used, adapted and developed to suit the constraints and affordances of a number of CMC modes, it is essential to discuss and make clear some possible variables that may reveal observed adaptations in the modes under scrutiny in this study. It is also interesting to come to grips with what, exactly, are the features of the so-called “third medium”? What makes it a third medium distinct from speaking and writing? What can CMC do that other communications media cannot? How does this affect language?

2.5.3 Modes of text-based CMC

Different forms, or modes, of CMC have different conditions for communication. They allow for different types of interaction, are used for different purposes and have different impact on language use and message content. A mode is a genre of CMC that combines messaging protocols and the social and cultural practices that have evolved around the use of those protocols (Murray 1987; Herring 2002). Although, as Herring points

out, "cultures of use" of newer CMC may be emergent or latent due to the rapid progress of technologies.

As we saw in the previous section, Crystal contends that the electronic medium presents us with a channel which facilitates and constrains our ability to communicate in ways that are fundamentally different from those found in other semiotic situations (Crystal 2001). Furthermore, he argues that many of the expectations and practices associated with spoken and written language no longer apply. This claim will be investigated below. Much of the differences between how language is used in spoken and text-based interaction seem to lie in the nature of the medium. December argues that an important aspect of CMC is synchronicity (December 1996). As mentioned above, synchronicity is the relationship between different things in time. CMC can either be asynchronous (communication in which messages are exchanged during different time intervals) or synchronous (communication in which messages are exchanged during the same time interval).

CMC varies according to the technology on which it is based, and according to its contexts of use. Thus synchronous CMC (e.g., real time chat) differs systematically from asynchronous CMC in message length, complexity, formality, and interactivity – due, in part, to temporal constraints on message production and processing (Ko 1996; Herring 2002). Possible variables that influence form and content of messages will be dealt with in more detail in Section 2.5.4 below.

As we have seen, spoken language has often been claimed to consist of a narrower range of lexical items, simpler and shorter sentence structures compared to traditional written language (cf. Halliday 1985; Allwood 2000, among others).

CMC is human communication mediated by computers. A complex set of variables is involved when humans communicate. Different theories of human communication give different accounts of which variables are important and how these combine. The following section will give an overview of variables that are considered relevant to the study of text-based CMC. Before entering into a summary of relevant research on different modes of CMC, definitions and examples of asynchronous and synchronous modes of CMC will be given. A more detailed description of the particular CMC modes that were investigated for this dissertation will be given in the chapters dealing with each of them respectively.

2.5.4 Variables conditioning CMC

However we decide to view CMC, it still requires writing for textual communication. Paolillo (1999) argued that

If we are to understand truly how the Internet might shape our language, then it is essential that we seek to understand how different varieties of language are used on the Internet.

The variables conditioning language in CMC were introduced in the sections on general human communication and written and spoken communication. As discussed above, they are interdependent and interwoven in complex ways. This section will deal with them in connection to the particular modes of CMC that are analyzed in this dissertation: asynchronous (email and mobile text messaging (SMS)) and synchronous (web chat and instant messaging (IM)).

Table 8. Variables conditioning CMC.

<i>Variables</i>	<i>Conditions</i>			
	<i>Asynchronous CMC</i>		<i>Synchronous CMC</i>	
	<i>Email</i>	<i>SMS</i>	<i>Web chat</i>	<i>IM</i>
<i>Synchronicity</i>	Production and perception distributed in time and space No time pressure	Production and perception distributed in time and space Time pressure	Production and perception distributed in space Time pressure	Production and perception distributed in space No time pressure
<i>Means of expression</i>	Monological, 1-way (non- interactive) Monomodal Persistent	Dialogical, 2-way (interactive) Monomodal Persistent	Dialogical, 2-way (interactive) Monomodal Less persistent	Dialogical, 2-way (interactive) Monomodal Ephemeral
<i>Situation</i>	Between unknown Autonomous Independent Lacking immediate context	Between friends Dependent on shared background	Between unknown Immediate context	Between friends Dependent on shared background

Synchronicity

Asynchronous CMC

Text-based asynchronous CMC is communication that does not require participants to be online and available at the same time or place in order for communication to take place successfully. Messages are composed off-line, giving the sender the benefit of time for planning and editing the message. Email is a typical example of asynchronous CMC. Production and consumption of messages can occur at the respective paces of sender and receiver (Yates 1996), thus sharing properties of traditional written interaction.

It could be argued that there are different levels of asynchronicity. There are several factors involved in the levels of asynchronicity between email and SMS²⁰, for example. A person has to be by a computer and have an email client access a network to check for new messages, whereas participants in SMS communication may communicate from anywhere, because of the mobile nature of the means of expression. Furthermore, there is no action required for an SMS to be received by a mobile phone, other than that the phone must be switched on at some point.

Asynchronous interaction does not occur in real time, but the time for delivering messages is considerably decreased, compared to traditional "snail mail"²¹. Even though it is not required that sender and receiver are online simultaneously in asynchronous interaction, it is possible that they are, in which case messages may be sent and received within seconds. A message might not be read or responded to until much later, though. Under such circumstances, there are clear differences between email and spoken interaction that go beyond the use of text as a medium. In some text-based CMC one can get feedback that the message is received (some email clients and some mobile phones allow this), but not that it is actually read. Getting feedback that the message is received is possible in face-to-face interaction by the reactions in the listener. Something similar is also possible in spoken interaction over the phone, but perhaps to a lesser extent due to limited modalities involved. Getting to know whether the message is actually shared and comprehended is another matter, though.

²⁰ Short Message Service, i.e. text messaging via mobile phones, as was defined in Chapter 1.

²¹ "Snail mail" is a term for traditional mail sent through the postal service, which takes considerable time to be delivered compared to email. Hence the metaphorical reference to the slow moving animal.

Synchronous CMC

Like spoken interaction, synchronous CMC requires its interlocutors to be online simultaneously. Unlike spoken face-to-face interaction and similar to telephone conversation, the interlocutors do not have to be present in the same physical room. Synchronous CMC allows written communication to become interactive written discourse (Ferrara, Brunner et al. 1991). Most synchronous CMC is text-only²², and thus monomodal. Communication relies solely on what can be communicated through text and other graphic means. It should be pointed out, though, that clients for audio- and video chat are available as well, perhaps not used as extensively as text-only CMC at present.

Examples of synchronous CMC are various forms of real-time chat (web chat, IRC²³, etc.), in which a large number of people may participate in written conversation. The communicative situation could be compared to that of a cocktail party, albeit a virtual one. Users type their written contributions to the conversation, which are displayed in the chat window to everybody that is logged in. Similar to a cocktail party, one may overhear, or rather “oversee”, other conversations going on in the chat room. As in a real-world situation it is difficult to take part actively in more than one or two conversations simultaneously, the main criticism of Crystal’s rather optimistic and hypothetical view.

No synchronous CMC is fully synchronous in the way spoken face-to-face interaction is: there is always the lag and delay of typing and sending the message (cf. Kiesler, Siegel et al. 1984; Ferrara, Brunner et al. 1991). As mentioned above, Crystal (2001) points out that chat rooms, IRC and the like are too constrained by their response times and the slow speed of typing to be considered as a good analogy of speech. “Synchronous” CMC is in this respect merely “quasi-synchronous” (Garcia and Jacobs 1999; Rintel 2001). For the sake of simplicity, the term “synchronous CMC” will nevertheless be used henceforth to refer to the quasi-synchronous chat mode.

Gradations in the levels of simultaneity are closely linked to the notions of 1-way or 2-way communication (Cherny 1999; Herring Forthcoming), i.e. whether sender and receiver are able to see the message as it is being produced or not²⁴. A web chat or IRC requires the sender to type a

²² As Soukup (2000) points out, there seems to be a shift towards chat rooms using graphic representations of the users, rather than text-only representations.

²³ Internet Relay Chat, a type of synchronous chat (cf. Werry 1996)

²⁴ More on the notion of 1-way or 2-way communication in CMC in Section 2.4.6 below, in which medium and situational variables are discussed.

message and hit the enter key or click a button before it is sent off and appears in the communication window of the chat. In split-screen protocols, such as featured in the chat modes *Unix Talk* and *ICQ*²⁵, the synchronous chat displays the communication keystroke-by-keystroke in real time and gives the receiver awareness that the sender is producing his or her message (Herring Forthcoming). MUDs (Multi-User Domains) or MOOs (Multi-User Domains, Object-Oriented) are versions of chat rooms where pairs or larger groups of individuals can get together to "converse" through text messages (Turkle 1995).

December contends that important aspects of CMC are synchronicity, information richness, and social presence (December 1996). CMC varies according to the technologies on which it is based, and according to its contexts of use. Thus synchronous CMC (e.g., real time chat) differs systematically from asynchronous CMC in message length, complexity, formality, and interactivity – due, in part, to temporal constraints on message production and processing (Ko 1996; Herring 2002).

Social presence theorists concentrate more upon the specific characteristics of the cues being exchanged in mediated interactions. One of the most important cue characteristics exchanged in mediated interactions is speed. Put simply, the more speed associated with the exchange of cues in a mediated interaction, the higher "social presence" associated with that medium. For example, e-mail is a medium with low social presence because the exchanges are asynchronous or have long time intervals. In contrast, a phone conversation would be considered to have higher levels of presence since the interlocutors have a synchronous relationship.

Herring (Forthcoming) presents two sets of variable dimensions. The lists were arrived at on the basis of empirical evidence in investigating factors that condition variations in computer-mediated discourse. Her medium variables consist of technological features, some of which would, for the purpose of this thesis, might better be placed under the variable of *Synchronicity*, even though the variables clearly are interconnected. She argues that the level of synchronicity impacts on what types of messages are sent and how these are formulated. Further, whether the receiver is aware that there is a message being composed and that the receiver may read it before it is completed will impact on turn taking and message

²⁵ ICQ is an abbreviation of "I seek you", a software that allows users to have awareness of their friends' on-line presence. ICQ provides a range of communication facilities, including e-mail, instant messaging, split-screen synchronous chat, and SMS. For more information, see the ICQ web page <<http://web.icq.com/>>.

content as well as composition (1-way vs. 2-way message transmission). Herring (1999) points out that researchers in computer-supported cooperative work have identified various limitations imposed on group interaction by the properties of the medium, ranging from high production, reception, and speaker change costs (Clark and Brennan 1991), to "chaos" resulting from the greater openness of computer-mediated systems. She notes that two properties of this medium are often cited specifically as obstacles to interaction management (cf. Section 2.2.3: Allwood 1986; Allwood, Nivre et al. 1993; Allwood 2000):

- Lack of simultaneous feedback, caused by reduced audio-visual cues and the fact that messages cannot overlap;
- Disrupted turn adjacency, caused by the fact that messages are posted in the order received by the system, without regard for what they are responding to.

Herring (Herring 1999) argue that the first property, lack of simultaneous feedback, is a consequence of two separate features of CMC systems. First, text-only CMC is a "lean" medium which relies on fewer channels than face-to-face interaction for transmission of the message. Users do not see or hear their interlocutors and do not have access to non-verbal information about how others are responding. Second, most multi-participant CMC systems make use of "one-way" rather than "two-way" transmission—messages are sent in their entirety when the message originator presses 'send' or 'return', rather than one keystroke at a time (Cherny 1999). One-way transmission protocols include interactive "chat" systems such as IRC (Internet Relay Chat) and MUDs (Multi-User Dimensions), along with e-mail and all forms of asynchronous computer conferencing. (Two-way protocols include split-screen "talk" systems such as UNIX "talk" and VAX "phone".) In one-way systems, it is technically impossible for the addressee to respond while the message is being written; indeed, the addressee may not even be aware she or he is being addressed until a complete message appears on the computer screen.

Research on spoken conversational interaction shows that simultaneous feedback plays an important role in signaling listenership, timing turn-taking effectively, and maintaining continuous interaction (Allwood 1986; cf. Allwood 2000). Conversely, the absence of simultaneous feedback may result in discontinuity and/or overlap within turn sequences, as well as generally making it more difficult for message producers to tailor their messages to respond to recipients' interests and needs (Herring 1999).

Means of expression

CMC is less "rich" because there is most often only one information channel (text), compared to spoken interaction in which the interlocutors can rely on information from both audio- and visual channels. How many channels (text, video, voice) are used in communicating the message conditions what is communicated.

Herring (Herring 2002) suggests a number of variables connected with the medium of communication, or "means of expression", the term used in the present study. The length of time the text persists to be read and reread, and whether it is possible to print or save the text ("Persistence of transcript") is one variable. Another has to do with whether there is a limit to how long a message may be, or how many characters may be used ("Size of message buffer"). Döring (2002a) argues that the technical restriction of 160 characters on SMS makes users economize the number of characters per message. Other factors that influence SMS language use may be that there is a charge for each message sent, and that text input is cumbersome.

Herring goes on to point out that some modes of CMC allow messages to be sent anonymously ("Anonymous messaging"), which also impacts on language use. Whether or not the communication setting allows for dyadic, private messaging without any other participants being able to take part ("Private messaging") might also affect language use. Some modes allow the possibility to filter out unwanted senders or message topics ("Filtering"), and some allow for copying snippets of previous communication to include in a new message, around which a new message may be constructed ("Quoting").

These variables are not just connected with the written language aspect, but since the present study deals exclusively with text-based CMC it will disregard or modify some of Herring's variables.

Situation

December points out that just as context is sort of an inescapable ether in which actions take place online, so too are purposes part of actions online. Regardless of the model for the mechanisms, people do not go online for no reason. The reason may be "to play," "to waste time," or even "to be cool," but there is a reason. Moreover, the purposes people give for using the Net and Web relate to their context, with factors such as peer pressure and social context influencing online behavior (Lea, O'Shea et al. 1992).

Furthermore, CMC is situated in a human context. All acts, all communication is situated in a social context. CMC just does not well up

out of nowhere. December points out that he does not believe that a particular act of CMC can be understood without considering the social systems in which it takes place. It is impossible to isolate language use from language use in context.

Döring (2002a) points out that the linguistic characteristics of text messaging are influenced by the characteristics of the communicators: their interpersonal relationships, relying on pragmatic- and shared background knowledge, individual preferences and collective norms as well as on the communicative context or situation and the reasons for communicating

Herring claims that a number of social and situational factors have also been observed to condition variation in computer-mediated discourse, and especially refers to Baym (1995; 1996). Furthermore, Herring points out that this set of features is based loosely on that which Hymes (1974) proposed in the ethnography of speaking.

Participation structure:	one-to-one, one-to-many, many-to-many
Participant characteristic:	demographics, proficiency, experience, role/status in real life and virtual reality; pre-existing socio-cultural knowledge and norms of interaction
Purpose:	of group; of interaction (e.g. get information, recreational, etc.)
Topic or Theme:	of group as a whole (e.g. topic in discussion lists); of exchanges
Tone:	serious/playful, formal/casual, etc.
Activity:	e.g. debate, job announcement, problem solving, flirtation, game
Norms:	of organization, of social appropriateness, of language
Code:	language, language variety; font/writing system

Some of these variables seem to overlap, and some of them are less relevant to the particular modes of CMC that will be investigated in the present study.

2.5.5 Properties of the modes of CMC investigated in the present study

Email, web chat, instant messaging and SMS are the modes that are investigated in the present study. Email and SMS are both asynchronous

modes, while web chat and instant messaging are synchronous (the granularity of synchronicity varies between the modes in each group, to be discussed in the section on Properties of Synchronous Modes below). We begin with asynchronous CMC by describing the basic features of the modes, as well as the main types of research and findings, then go on to describe synchronous CMC in the same manner. Detailed information will be given below in each chapter that deals with the analyses of the modes respectively.

Properties of Asynchronous Modes

Email

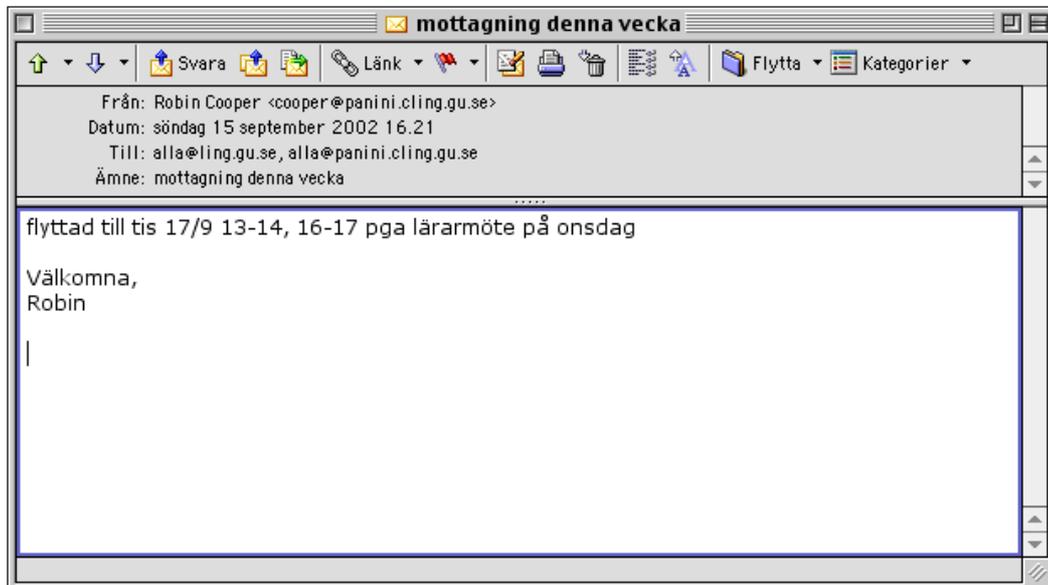


Figure 2. Example of a window showing an email as displayed in an ordinary type of email client (in this case *Microsoft Outlook Express for Mac*).

Email, or electronic mail, is the electronic counterpart of traditional letters. It is an asynchronous form of communication, using writing as a means of expression. Messages are created using an email client, and are typed on a computer keyboard and are read as text on a computer screen. Compared to ordinary letters, the transmission time for email messages is reduced to seconds.

The email software automatically adds a header to each message, giving information about sender address, time and date of posting, etc. Files consisting of, for example, digital images, audio or text files can be attached to email messages. Email has provided a new way for people to

communicate. Email is often perceived as being ephemeral and used instead of telephone or face-to-face conversations, even though messages are normally stored on the local hard disk, and can be printed, redirected, copied. The ephemeral feel, and its apparent intimacy, may also be the reason it fosters more self-disclosure (cf. Sproull and Kiesler 1991).

Messages are used for both private and business purposes. Messages can be sent from one person to another, or from one person to multiple receivers. The latter is the case in electronic mailing lists and discussion groups, where an email goes out to groups rather than individuals. There are various options for this, including web-based bulletin boards and threaded discussions, where new topics and subtopics are kept separate.

Linguistic features of email

Due to ethical considerations, messages from public mailing lists have been the main focus for analysis so far. Extensive work on electronic mail as a medium for dialogue has been made. Severinson Eklundh (1987) argues that the computer, as a medium, may serve communicative purposes previously reserved for direct, spoken conversation. Her analysis of a body of messages from the Swedish conference system COM, showed that messages were structured into coherent dialogues in a way distinct from other forms of written communication. Dialogue sequences appear in email simply as a result of linking between messages and their replies. The exchange of email messages may contain a range of speech-like or “conversational” features. This applies both to the character of individual messages and to the structural features of the entire dialogue. Severinson Eklundh states that the email messages in her study typically were short compared to regular letters. Furthermore, almost all messages used a direct, informal style of address.

Linguists have made contributions to the analysis of email style. Email has been compared to spoken and written data bases (Collot and Belmore 1996; Yates 1996). Collot and Belmore concluded, "The genres which (electronic language) most closely resembles are public interviews and letters, personal as well as professional". Digital writing is playful, dynamic and speech-like (Bolter 1991). Email messages have "oral," and "written," as well as new and uniquely “digital” features (Ferrara, Brunner et al. 1991; Maynor 1994). This empirical profile of the language of email is useful but should not be taken as identifying firm, accepted norms, in any statistical or sociological sense; on the contrary, the language of email is very much in flux. Danet points out that when composing email messages, we draw

on templates of many genres of both oral and written communication, including not only the business letter and the personal letter, but also face-to-face conversation, telephone conversation, intra-organizational memo, telegrams, postcards, and greeting cards. Danet's research shows evidence for blurring of genres of personal and business letters, and for a new acceptability of speech-like features in digital letters (Danet 2001).

Du Bartell (1995) studied the features of the messages in a mailing list, and reasoned that the spoken and written-like characteristics in a written medium result from the constraints imposed by the computer medium – the machine architecture. According to her, the computer medium permits texts that seem both written-like and spoken-like. This thesis maintains that this explanation is too simplified; it disregards the complexity of interdependent variables that influence communication and language use in interaction.

Du Bartell's claim that messages in a mailing list display both written-like and spoken-like features is supported by Baym (1996). Although Baym treats CMC as a single genre, she argues that even though CMC is written, it is marked by many features associated with face-to-face interaction. Her study of Usenet messages²⁶ showed that Usenet interaction is a hybrid between oral, written, interpersonal, and mass communication. Baym concludes that the message features of her study stem from five interrelated factors – the Usenet medium, institutional context of work, topic, participant gender, and the social context which the participants strive to create. Yates (1996), who emphasized the heterogeneous character of electronic messages, also took this wider perspective. He noted that CMC is affected by “the numerous social structural and social situational factors that surround and define the communication taking place”.

Herring (1996) analyzed the schematic organization of electronic messages posted to two academic mailing lists, one predominantly male and the other predominantly female, in order to evaluate the popularly held view that men and women use email for different purposes (information exchange vs. social interaction). Her results did not support the stereotype, but showed that women and men's messages are structured differently, with female users exhibiting alignment with, and male users opposition towards, their addressees.

The basic electronic message scheme was analyzed according to the functional moves salutation, introduction, body, and close. She concludes from analyzing 136 messages that “Surprisingly few messages are preceded

²⁶ Usenet is an extensive collection of topically organized discussion groups distributed through the Internet.

by a salutation (only 13 % on average), and fewer yet are followed by a complimentary close or a postscript” (Herring 1996).

Both Herring and Du Bartell explain the relative lack of epistolary conventions as being due in part to the fact that a header is added automatically to each message by the electronic mailer, including a separate line for whom the message is “from”, whom it is addressed “to”, and the date and time of posting. Partly as a result of having a subject displayed, email messages frequently omit even the typical salutations and farewells associated with other media, regardless whether the speakers²⁷ know each other. Email messages display rather informal register characteristics, even between unknown persons (cf. Du Bartell 1995), which is something Danet also found when analyzing portions of her own email correspondence (Danet 2001). Most previous studies have been based on public mailing lists, and not on private messages, in part due to privacy issues. The identity of the intended receiver(s) and the relationship between correspondents potentially influence linguistic structure and content of email messages. Danet’s email study focused on letters sent to her by people who did not know her but knew her name and status. She remarks that:

A writer of a first letter is likely to take special care in its formulation. Traditional norms for letter form are likely to be salient, and writers are likely to be especially conscious of the impression their message may make on the recipient.

Danet analyzed her email messages holistically using the criteria of the business letter template. In the case of openings and closings, letters were coded “yes” only if they contained both an appropriate opening and an appropriate closing. Abbreviations, spelling, typography, punctuation, and use of exclamation points conforming to the norm were coded “yes”. She found that none of the letters of her corpus conformed to all given criteria, and that variability was extreme. Most letters conformed to expectations regarding syntax and vocabulary as well as those for spelling, typography and layout, but almost none followed paper letter practices regarding openings and closings.

Danet suggests that the new medium invites informality even in business or official contexts. This informality is not due to the technology per se, rather on convergence of new technology with a general tendency toward informality already in place according to Danet. She remarks that the

²⁷ Du Bartell employs the terms ‘speaker’ and ‘listener’ regardless of the medium of linguistic communication.

novelty of the medium can facilitate changes of style and substance in much shorter time than paper letters could have done. Style, or register, may apply to substantive domains of human communication and action – not to all communication in a medium (cf. Allwood 1976; 2000). Danet concludes that the language of email is in a state of transition. She predicts that informal, partially speech-like email style will increasingly characterize public as well as personal communication. Our normative expectations will change to provide increased legitimacy of a more informal style.

The material analyzed in the present study was chosen partly because it represents messages that are somewhere between private one-to-one messages and messages to public mailing lists, namely email from private people to an unknown “authority” at a city council. The messages are also comparable with a body of traditional letters that were sent for similar purposes. The choice of material also makes for an interesting comparison that might reveal differences due to means of expression. More details about previous research are provided in the study on chapter on email below.

SMS

Short message service, abbreviated SMS, as defined within the GSM digital mobile phone standard, is a service that enables its users to send short text messages to one mobile phone from another, or to a mobile phone via the Internet.

There are two basic ways of composing and sending text messages to mobile phones: typing them on a mobile phone, or typing them on a keyboard for transmission via one of a multitude of web based SMS-service suppliers. The prerequisites for the different modes of composing and sending are slightly different. The amount of effort involved in typing messages and the cost of sending affect when and how often messages are sent.

Common to both ways of composing and sending text messages, the text can be comprised of words or numbers or an alphanumeric combination. Each short message can be up to 160 characters in length when Latin alphabets are used, and 70 characters in length when non-Latin alphabets such as Arabic and Chinese are used²⁸. Non-text based short messages (for example, in binary format) are also supported. These are used for ring tones and logo services for instance.

²⁸ <http://www.mobilesms.com/>



Figure 3. Example of a mobile phone (in this case, a Nokia 6210) displaying a text message on the screen. (Photo by Christer Erling).

Most commonly (cf. Döring 2002a), text messages are created on the small keypad of the mobile phone and are read as text on the tiny screen of the phone. Different models of mobile phones allow for slight differences in the way text is entered. Because simple person-to-person messaging is such an important component of total SMS traffic volumes, anything that simplifies message generation is an important enabler of SMS. The small keypad on the phone has ten keys (for the figures 0-9), as well as two additional keys²⁹. Each key for figures also harbors 3-4 letters in alphabetical order. For example, the key for the figure 4 also holds the letters *g, h, i*, and has to be pressed twice in order to type the letter *H*. To type the five-character word “hallå” (“hello”), the following keys have to be pressed: 42552. However, the keys have to be pressed 14 times, to get the correct sequence of letters (44255555522222). Obviously, this procedure makes text entry laborious and time consuming. In an attempt to overcome this obstacle, most modern mobile phones today have some

²⁹ On a mobile phone which is common in Sweden, for example Nokia 3310, these keys harbor the characters *+ and #↑ respectively.

form of software³⁰, which predicts, or guesses, words from the sequence of keys pressed. Several types of software for predictive text entry have been developed.

Instead of entering text using the “multi-press”, or “multi-tap” technique described above, each number key only has to be pressed once for each letter. The mobile phone has a built-in dictionary, and will try to identify what is being written based on the words already stored in the phone's memory.

As yet, not many studies have been published on the use of text messaging, and even fewer on the effects of SMS on language. Grinter and Eldridge (Eldridge and Grinter 2001; Grinter and Eldridge 2001) report on teenagers' text messaging practices in a study executed in England. Their study shows that SMS is used for arranging and adjusting times to talk, to coordinate with friends and family, and chat. Teenagers find text messaging quick, cheap, and easy to use. Eldridge and Grinter show that most of the teens that they interviewed did not use predictive text entry much because they found it troublesome. The predictive text input system does not accept slang or dialect unless the software is taught to recognize these words by the user. Kasesniemi & Rautiainen (2002) argue that the mobile of a “lazy” user “speaks” only standard language with no personal tones. In addition, they question which effect predictive text entry may have on the development of the Finnish language and how much it will influence the personal style of their users: “Is it the device or the user that will adapt?”

Linguistic features of SMS

Döring (2002a) contends that the technical restriction of 160 characters per message both permits and forces people to express themselves concisely. One may make oneself brief without fear of being perceived as short-spoken. Moreover, one has to be brief not to go beyond the scope of the limited number of characters per message, because each message sent is relatively expensive. In the long run, one will make oneself brief because text input is so cumbersome. Finally, because most SMS communication is interpersonal communication between people who know each other, one can be brief, relying on pragmatic and shared background knowledge.

Features that are characteristic of spoken language, such as dialectal words, interjections and prosody are verbalized and spelled out in SMS.

³⁰ T9 (<http://www.t9.com/>) is an example of predicating software used in many mobile phones today. Other kinds of predictive software are Wordwise and Letterwise from Eatoni (<http://www.eatoni.com/>).

The use of short forms may help to enhance the feeling of directness, unforced and natural spontaneity. Kasesniemi and Rautiainen (2002) report that teens write messages in all lower case or all capitals, and that words are shortened, inflectional endings characteristic of the Finnish language are left out, and word spacing can be left out. They further suggest that the messages often bear more resemblance to code than to standard language, and that the code is not necessarily accessible to an outsider (cf. Sveningsson 2001).

Schlobinski et al. (2001) show that SMS communication shows linguistic characteristics that we are familiar with from chat communication: smileys and other graphic means, “hybridizations” between written and spoken language, varying written language conventions, syntax conditioned by medium (ellipsis). Döring (2002a) made a thorough study of the lexical and syntactical characteristics of SMS communication among German users. Her findings confirm the results of Androutsopoulos and Schmidt (2001), revealing that a great variety of syntactical and lexical reductions are found in SMS communication. The most common syntactical reductions were found to be:

- Deletion of subject (especially subject pronoun):
[Ø] FAND'S AUCH S. SCHÖN
- Deletion of article and possessive pronoun, determinator:
[Ø] Wetter zu pissig!
- Deletion of preposition or contraction of preposition and article:
Sitzen [im] Park, trinken Kaffee.
- Deletion of copula-, auxiliary- or modal verbs (+XP):
[Ø] HEUTE LARA CROFT GUCKEN?, [Ø] Ich dich besuchen?
- Deletion of verb and subject pronoun:
[Ø] Bei dir oder bei mir?
- Telegram style:
Morgen Frühstück?

Döring points out that a proper taxonomy for lexical short forms is lacking. She found two types, acronyms and abbreviations, which were further divided into types.

Acronyms are word forms that are made up from the initial (in some cases the first two or even three) letters in a sequence of words. These will be pronounced letter wise or from the prosodic value. Döring divides the acronyms into *letter acronyms* and *phonetic value acronyms*.

Abbreviations are written short forms of words, typically formed using initial letters, words are pronounced in full form when read out. Döring divides abbreviations into *Conventional abbreviations* and *Unconventional, ad-hoc abbreviations*.

Letter acronyms and ad-hoc abbreviations proved to be most commonly used of the four types. Döring found that out of 1000 words, 30 were acronyms and abbreviations.

Schlobinski et al. (2001) argued that the relatively heavy use of abbreviations they found has a functional explanation: it is a means to compress the text to fit the 160-character format. Further details of previous research on SMS is found in the chapter on mobile text messaging below.

Properties of synchronous modes

Web Chat

A chat room is a Web site that provides a venue for users to communicate in real time. Forums and discussion groups, in comparison, allow users to post messages but lack the capacity for interactive messaging³¹. Most chat rooms do not require users to have any special software; those that do, such as Internet Relay Chat (IRC), allow users to download the necessary software from the Internet free of charge. Chat room users register for the chat room of their choice, choose a user name (a nickname) and password, and log on to a particular room (most sites have multiple chat rooms). Synchronous CMC requires its interlocutors to be online simultaneously. In a real-time chat, a large number of people may participate in written conversation. Similar to a cocktail party, one may overhear, or rather “oversee”, other conversations going on in the chat room. As in a real-world situation, it is difficult to actively take part in more than one or two conversations simultaneously. Inside the chat room, there is generally a list of the people currently online, in addition, users are alerted each time a new person has entered the chat room. To chat, users type their written contributions to the conversation in a text box, hit the enter key or click a send button, and the message will be displayed in the chat window to everybody logged on at the same time.

Bechar-Israeli (1995) suggests that since an individual’s physical existence and identity must be condensed textually into a single line which states one’s nickname and electronic address, the person will attempt to

³¹ It should be pointed out, however, that asynchronous media are not to be regarded as non-interactive by definition.

make these representational elements as prominent as possible. The way to do so is to choose an original nick which conveys something about the person's "self" which will tempt other participants to strike up a conversation with that person. Bechar-Israeli points out the extent to which nicks have become an important part of the electronic self: they are experienced as an extension of the self. Bechar-Israeli further states that the emergent IRC culture is a culture of linguistic virtuosity on the one hand, and of contempt for the rules of the language on the other. It is a culture that provides freedom in abundance to engage in identity games through the use of nicknames.

Linguistic features of web chat

Danet et al. (1997) observe that linguistic features previously associated with oral communication are strikingly in evidence in real-time chat. They further point out that CMC in general is strikingly playful.

Storrer (2001) use data samples from different types of chat services to discuss media-specific forms of turn-taking and the use of deictic expressions in chat communication. She claims that the usage of written language as well as the specific technical setting affects strategies for language processing during chat communication. As a consequence, written chat conversations differ considerably from their spoken counterparts. She discusses and explains some linguistic peculiarities in the dialogues which occurred in chat conversation, and focuses on two phenomena that she claims reveal significant differences between spoken and written dialogues: The organization of turn-taking, and the usage of deictic and local expressions (e.g., "*here*", "*I*", "*me*", "*above*").

Schönfeldt (2001) examined the question how the "typed conversations" in chats that were for a for unspecified topic, as well as un-moderated, can be compared to face-to-face conversations in regard to their organizational structures. The categories developed within the theoretical framework of traditional discourse analysis may be applied in describing the interaction in chats. In applying analytic units used in conversation analysis on a web chat she shows that the organizational structures in chat correspond to patterns of organization in oral conversations. Schönfeldt concludes that chat therefore can be analyzed as a new form of conversation.

Werry (1996) notes that electronic written online interaction makes for less time delay for feedback, compared with traditional written communication. As the participants communicate in real time, they are able to negotiate meaning in another fashion than is possible in traditional letter writing, for example. Werry (1996) argues, though, that the receiver is

usually unable to supply the minimal responses (nonverbal forms such as such as nodding, gaze, and verbal forms, such as ‘mm hm’ and so on), though. The fact that the participants in chat conversations can hide behind text, as it were, enables them to change and play with identities – age, sex, background, and so on, if they feel like it. Witmer & Katzman (1997) point out that:

A key characteristic of virtual reality [is that] CMC can mask personal characteristics and identities of cyberspace travelers to create personal anonymity in a public arena.

In the analyses of web chat below we will get back to the fact that electronic written online interaction makes for less time delay for feedback, compared with traditional written communication, as well as the fact that the participants in chat conversations can hide behind the text. Anonymity and synchronous, multiparty, written interaction are also interesting factors to pursue.

Instant Messaging

Instant messaging systems of various forms have gained much popularity during the past few years. Commercial instant messaging systems such as ICQ and AOL Instant Messenger have attracted millions of daily users in recent years, and the instant messaging phenomenon has also recently attracted researchers (e.g. Nardi, Whittaker et al. 2000; Smith, J.J. et al. 2000).

Instant messaging systems are tools that support awareness of the presence of other specific users and often allow for various types of communication. The systems allow users to have users with whom they wish to communicate in “buddy lists”, and the system alerts them when their friends are online. ICQ allows users to communicate via email, SMS, split-screen chat, and instant messages. The system also allows for transferring files of various types, and sending URL-links. When a message is sent, the system gives visual and/or acoustic notice so that the receiver may read it at once.

Software for systems such as ICQ and AOL Instant Messenger must be downloaded and stored on a particular computer. WebWho, which is analyzed in this study, is web based and requires only a normal web browser and Internet connection.

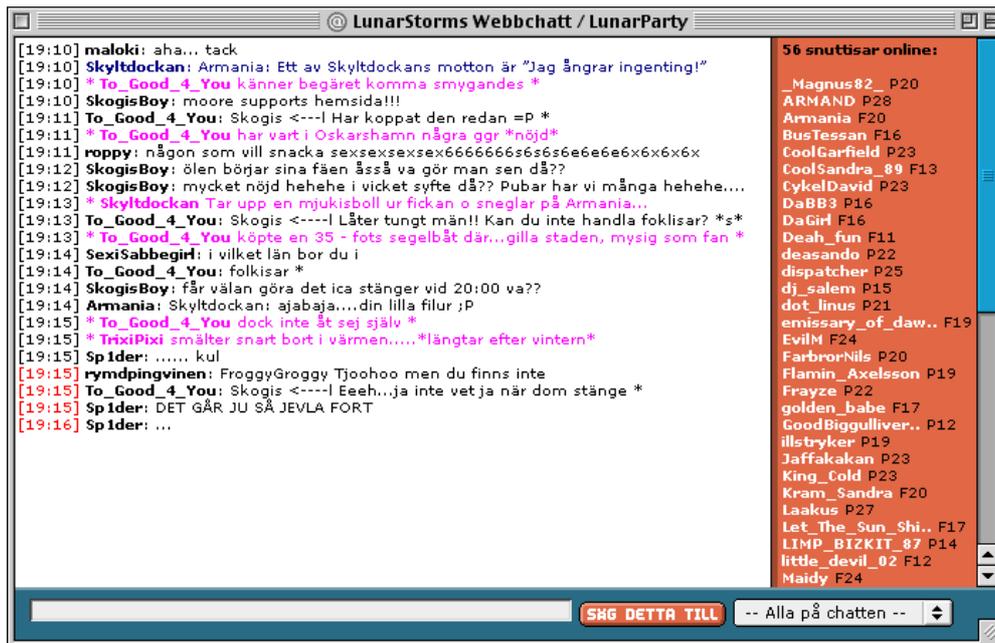


Figure 4. Example of synchronous CMC: a web chat window showing multiple participants engaged in interactive written discourse. The participants type their messages in the textbox in the lower part of this window.

There have been systems designed to support awareness of presence in real time by displaying a map, overlaid with up-to-date location information of people. For instance, ActiveMap (McCarthy and Meidel 1999) is a system deployed in a large corporate office setting. The system is based on active badges with a supporting infrastructure of beacons spread throughout the office environment. When running a custom application on a desktop PC, one can see a schematic map of the offices with information about who is where in the rooms and corridors. Other systems have been created for visualizing the dynamics of electronic communities based on log files (e.g. Donath 1995) but such systems tend to be less useful for supporting synchronous or semi-synchronous activities. Smith et al. (2000) created *Threaded Chat*, a system for real-time visualization of threaded chats between multiple distributed users, somewhat similar to how threads in Usenet newsgroups are organized. Threaded Chat was designed to make it easier for its users to follow the otherwise transient nature of chat or instant messaging. Some online presence information was presented, but there were no fine-grained cues as to the whereabouts of the users in relation to each other or a local physical area. Both these systems were also intended for geographically dispersed users rather than co-located or nearly co-located people.

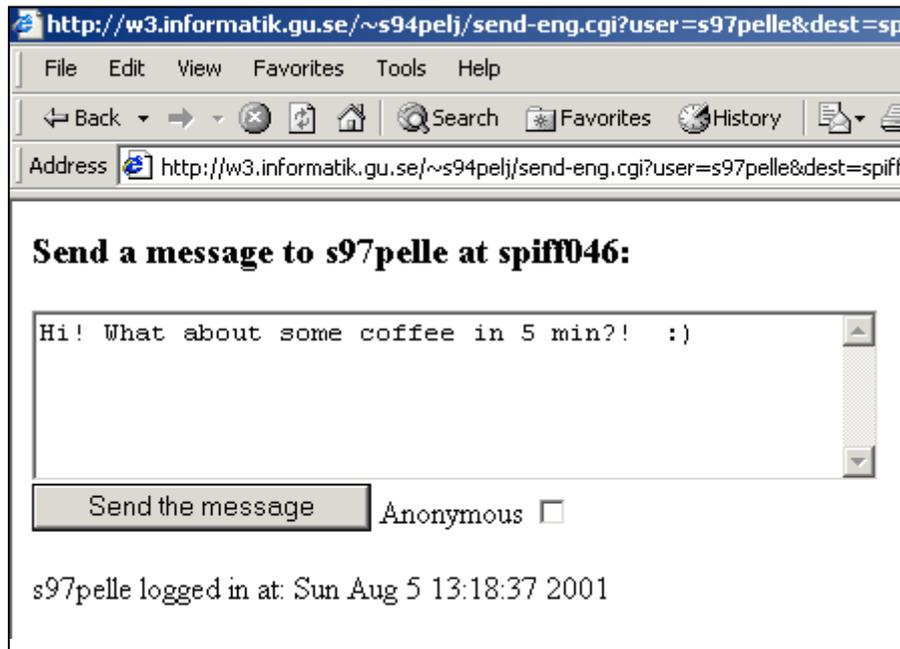


Figure 5. Example of the message window in WebWho. By checking the box next to the send button, the sender has the option of sending the message anonymously.

WebWho displays a schematized view of who is logged in at which computer in a large university computer lab. The study analyzed the instant messages that were sent through the system.

2.6 Chapter summary

The purpose of this chapter was to give a theoretical background to the studies of four modes of CMC that were investigated in the present work. Sections 2.2 and 2.3 reported on theories of human communication in general (Hymes 1974; 1976; Allwood 2000), Section 2.4 dealt with previous research on written and spoken communication, and Section 2.5 reported on theories of CMC. A set of three variables (*Synchronicity*, *Means of expression*, and *Situation*) were introduced and dealt with in each section. It was discussed how the variables condition general human communication as well as spoken and written interaction. The last section of this chapter dealt with how CMC is influenced by the variables of Synchronicity, Means of expression, and Situation. The implications that these have on the various modes of CMC were stated.

The next chapter will present the general methods of data collection and analysis. Specific details concerning data and analyses will be given in each chapter reporting on CMC studies.

The studies below will investigate how written language is used and adapted to the particular conditions of four modes of CMC. The studies will analyze how the variables of synchronicity, means of expression and situation affect how text is used. Questions to be answered are:

- How is written Swedish used and adapted in various modes of CMC?
- In what ways and why do these modes of CMC differ from the norms of traditional written language?
 - Which are the written vs. spoken language features in each mode respectively, and why are these used?
 - Which specific “e-style” characteristics are found in different CMC modes?
- Which variables influence this adaptation?

In the concluding discussion, the question of whether genre/activity or means of expression is the more important factor will be discussed once more. In addition, whether these modes of CMC can be regarded as “a genuinely third medium”, as proposed by Crystal (cf. Section 2.5.2 above), or not, will also be discussed. The often-voiced fears that language is influenced negatively by the use of computer-mediated communication will be addressed.

3 Material & Methods

3.1 Introduction

This chapter describes the general methods that were used to collect and analyze data. Material was collected from four different settings of computer-mediated communication (email, web chat, instant messaging, and SMS), each setting requiring slightly different methods. In this section, both material and methods of analysis are accounted for generally. Specific accounts for material and methods will be found in each section dealing with the different CMC settings respectively. Ethical aspects of collecting and handling data are considered, and the choice of material is discussed.

As mentioned above, computer-mediated communication is used in a variety of modes for varying purposes. Email, chat rooms, various sorts of instant messaging and SMS are all modes that are frequently used today. The choice of modes, or settings, to analyze in this dissertation was made on the assumptions that these are modes that are widely used by many people today. The modes are used for both recreational and “serious”, professional purposes (cf. Herring 1996).

Below follows an account of the method of analysis that is general for all four studies. Next, a brief account of material and methods specific for each study. The chapters in the dissertation that deal with each study separately contain a more detailed account of material and methods specific for each study. The individual modes of communication have been described briefly in the background chapter (Sections 2.5.5), and will be dealt with further in the chapters devoted to each mode respectively.

3.2 Ethical Considerations of Collecting and Handling Data

Before collecting data for each case study it had to be considered which kinds of data would be possible to accumulate and how these should be handled ethically. Established guidelines of how to handle data do not always fit these still rather new modes of communication. Public spaces in which a vast number of participants communicate in real-time by writing messages are a rather recent phenomenon. Mann and Stewart (2000a) outline an ethical framework for qualitative research of internet communication, and state a number of principles of fair information processing online:

- Personal data should be collected for one specific, legitimate purpose
- People should have access to the data collected about themselves
- Existence of data banks should be publicly known
- Personal data should be reasonably guarded against risks such as loss, unauthorized access, modification or disclosure
- Data should be collected in a context of free speech
- Personal data are not to be communicated externally without the consent of the subjects who supplied the data

One has to use common sense in handling data from computer-mediated communication, and apply established guidelines of how to handle data wherever possible. All of the guidelines suggested by Mann and Stewart above have been considered. Information that may identify participants or places is never to be exposed, so the data were made anonymous, participants were given pseudonyms, place names changed, etc. The data included in this study are not to be used by a third party or for any other purposes than those stated when collecting it. As for the collection of material for the case study on email, it originates from a body of incoming communications sent to the city council of Göteborg, Sweden. A city council is an official institution, and Swedish law states that all documents there are to be publicly available, which means that anyone might request copies of these. The database gathered from that material for the purpose of this thesis is not accessible for anyone other than those who have worked with the analysis of it.

In as many instances as possible, the informants' consent has been obtained. Obtaining consent from each participant in a web chat during a

week is nearly impossible. Some researchers argue that the investigator has to participate actively in a chat room, as well as inform the other participants that there is an ongoing data collection in order to justify the use of the material later. The stance taken in the present work is as follows. The material for the email study was chosen partly because it consists of publicly available material, and partly because it could be compared with traditional written letters of the same nature. In addition, it proved difficult to get hold of a suitable corpus of private email. The material poses no other ethical problems than those to which any research needs to adhere.

When entering the web chat, all chat participants were reminded by a system message that they were logging on to a forum that is considered a public space, and that everyone with access to the Internet may read their contributions. Permission to record and analyze the material was obtained from those responsible for the web chat before logging on for data collection, and they were informed in detail about the purpose of the study and how analysis was to be conducted.

The issue of how to deal with nicknames has been discussed in the literature on research ethics and data from the Internet (Mann and Stewart 2000; Sveningsson 2003). Some argue that everything that might identify a user must be de-identified. Nicks are chosen by the participants as a way of signaling something about themselves, and as a means of getting attention through interesting names and eye-catching spelling and orthography. People keep their online identities and often use the same nick over again (cf. Herring 1996; Sveningsson 2001). Frequent participants in certain chat rooms recognize each other, and may even know the off-line person behind the nickname. Nicks are also intertwined with the conversation practices by automatically being added by the system foremost in contribution or used as a device to address other people³². The chat shows who says what to whom, and is thus part of how language is used and adapted to the conditions of the communicative setting. The original nicknames were thus left in the chat corpus used in the present research.

Material from the instant messaging tool WebWho was collected through the systems log of the university computer lab. Students who used the tool were notified that logs were collected for a project, and were aware that their messages were stored in a database. To log on and use the system all users enter their student ID, which may identify the user if compared to data about students that were registered at the time. All messages were made anonymous by giving users pseudonyms.

SMS data was both collected by inviting anonymous informants to share their messages via a web page and by collecting user diaries. Information

³² Werry (1996) uses the term “addressivity”.

about the purpose of the study was provided on the web page, and participants were urged to email the researcher if they had any questions. The informants that contributed by keeping a user diary, forwarding messages, and by being interviewed were naturally informed about the purpose of the study and that their material would be handled confidentially. Everything that was revealed to the researcher in the interviews, or written in the diaries, and the messages that were forwarded were handled strictly confidentially, and participants remained anonymous to everyone but the researcher. These informants were paid for the trouble of keeping a diary and for their costs of forwarding messages.

If a researcher participates in the conversation that he or she is going to analyze, he or she will probably influence topic in the chat room conversation, and to a lesser extent the language that is used in the discussion. Spontaneous, informal, and recreational chat in a web chat will be influenced if participants are informed at various intervals that some researcher is going to analyze their contributions and store the material. For these reasons, and for the reasons given above, permission from those responsible for the maintenance of the chat channel to use the material provided both the necessary and sufficient degree of consent.

3.3 General Method of Analysis

The aim of the analyses contained in this dissertation was to find out how the three dimensions, or variables (taken to condition language use, see the Background chapter above) affect language use in four modes of CMC. The modes represent both asynchronous CMC (email and SMS) and synchronous CMC (web chat and instant messaging). The tool used for text input with the exception of the greater part of the SMS corpus was computer keyboard. Most SMS messages were created on the keypad of a mobile phone. The situations of communication vary, as do the activities for which the interactions take place. Purposes for communication range between information requests, jokes and task solving. The different situations require variations in language use; styles range from formal to informal. For instance, the email messages used here were predominantly sent from people who did not know who the receiver would be, while SMS messages were in almost all cases sent to friends and relatives. Thus the material represents communication on different levels of synchronicity, with differing settings for production and perception, as well as differing

goals of communication, and relationship between communicators. All of these factors had to influence the analysis of language use for this study.

The material from all of the modes of CMC in this dissertation was analyzed automatically with the help of a software tool, TraSA (Transcription Statistics with Automation)³³. TraSA is a computer tool used in corpus linguistics to calculate measurements like the number of *Tokens*, *Utterances*, *Turns*, *Pauses*, *Overlaps*, *Vocabulary Richness*. These measurements are calculated per *speaker*, *transcription*, *section* or combinations of these. TraSA has been used for analyzing the material from all modes quantitatively: word frequency; the occurrence of abbreviations and punctuation marks; mean length of utterance, etc.

3.4 A note on selection of material

There are several reasons for selecting the particular material that is used here to represent material from four modes of CMC. One of the reasons is simply that it was necessary to adhere to ethical methods of collection. It is, for example, technically possible to log a web chat without acknowledging it. It is important to get approval from the people who are responsible for the maintenance of the chat before commencing data collection (cf. Herring, 1996). It is also important to make sure that they are provided with information about the aim of the study and what steps are taken to insure that the chat participants' integrity will be respected.

Another reason is that the material must be available and accessible. For example, despite the fact that SMS is extremely popular, it proved difficult to collect material. It might be the case that SMS is experienced as being even more private than email. Most messages are sent between friends who know each other well, they communicate about things that they do not want any one else to read, even though the content of most messages is seemingly trivial.

The most important reason for collection of this particular data, though, is that the data represents material from four modes of CMC, which are used extensively by many people everyday, and that it is interesting to investigate how language is used in these still rather new modes of communication, and what implications it might have on language in general.

³³ TraSA was developed by Leif Grönqvist at the Department of Linguistics, Göteborg University, Sweden (Grönqvist 2000).

3.5 Representativity of the data

The corpora that have been used in the present study represent use of computer-mediated communication in particular activities. The email corpus represents electronic letters from private individuals to an unknown authority at a city council; the reasons for which range from comments on the city council's web site, complaints about parking facilities, or even simply in order to annoy. The messages were typed in a form at the city council's web page. The web-based form principally looks like the interface for a normal email client, but does not work exactly the same. Customized email clients such as those used on personal computers have personalized preferences stored; the program automatically adds sender address, time of posting, and so on (see, Herring 1996, for example). For that reason, the email corpus might not be representative of all types of email communication.

The web chat corpus represents social, recreational chatting in an open, topic-unspecified chat room. Contributions in the chat room range from, for example, conversation about "serious" matters, to flirtation, or verbal abuse. The corpus of web chat material is likely to be typical of social, recreational written conversation in an open web chat, even though one has to bear in mind that language use may vary depending on which participants are logged in and topics of conversation.

Contributions found in the logs for the instant messaging study vary between anonymous mischief and efficient task solving matters, for instance. The material is first and foremost representative of instant messaging with this particular type of setting. As it functions as an instant means of communication between people who know each other, often to coordinate activities, it is very likely that it is similar to communication on other instant messaging protocols such as ICQ, for instance.

The SMS-corpus constitutes messages that deal with everything from requests for information to chain messages. Although the material represents communication between the individual persons who contributed to this study, the material is broad enough to ensure that it represents language use that is typical for text messaging.

It would have been ideal if the corpora had represented language use in several activities evenly spread across gender and all age groups, as well as covering geographical areas and social status. This study employs as broad user groups and activities as possible under the circumstances at hand, but a broader base than the one for this study would have been ideal. Thus, the results may not be generalized to cover language use for all activities that email, web chat, instant messaging, and SMS may be used. The corpora are

large enough, though, to give interesting insights into how written language may be used in the modes of CMC under discussion.

3.6 Methods and Material in Four Modes of CMC

3.6.1 Analyses of Language Use in Email

Material

The data in this study consists of electronic mail and traditionally written (“pen-and-paper”) letters from citizens to the city council in Göteborg, Sweden. The messages are classified as public material, i.e. they are available to anyone who wishes to take part of the material. The email material consists of 183 messages sent between April 1, 1998 and August 31, 1998. The web page of the city council provides a service for “Questions & Answers”, at which anyone may send messages of any sort with which they hope that the people at the city council might be able to help them.

On the web page, people are asked to fill in a form with the following fields: “Sender” (email return address in order to receive an answer), “Subject/ topic” (what the message is about), and the field “Text” in which the message itself is composed. After filling in the form and writing the message, clicking a “send” button at the bottom of the page³⁴ sends the message. The web based question form basically looks like the interface of a normal email client, but does not work exactly the same way. Customized email clients such as those used on personal computers, or at email services on the web (Hotmail or Yahoo, for instance), have personalized preferences stored that automatically add sender address, time of posting, and so on (cf. Herring 1996, for example).

Material to compare email messages with consists of 41 traditionally written paper letters. Most of the letters are of the same kind as the email messages; open³⁵ letters from citizens asking questions, requesting help, or

³⁴ The web page-form at <http://www.goteborg.se/wwwdb/gbgwww.nsf/fragorochsvar> (retrieved 5, August, 2000).

³⁵ “Open” in this sense means that the letters were addressed to anyone at the city council, and not someone in particular. Some of the traditional letters were addressed to head of the city council in person.

the like. The comparatively low number of traditional letters is due to problems that the staff at the city council archive had for abstracting material out of their filing system. The high word count compensates to some extent for the fewer number of messages.

Table 9. Data from the corpora of email- and traditional letters to the city council.

	<i>Email</i>	<i>Traditional letters</i>
<i>No. messages</i>	183	41
<i>No. words</i>	11,660	17,771

Methods of Collection and Analysis

Both traditional letters and email are asynchronous modes of communication. Both the traditional letters and the email material for this study were collected from city council with the help of the staff from the departments of information and of archives. Because of some obscure regulations at the city council, the email messages had to be printed at the office, and given as paper copy printouts rather than being forwarded electronically. The copies were scanned and stored in digital format, and then archived for proofreading and checking that the scanned versions were entered correctly.

The traditionally written letters that were typed on typewriters or computers were possible to scan and store digitally allowing for automatic analysis. Handwritten letters were copied by manually typing the text in a word processor, and were then saved as electronic text.

The digital copies were analyzed using the automated tool – TraSA - described above. For this study, the software was used for quantitative analyses of the occurrence of abbreviations and punctuation marks, closing and introductory words in messages, mean length of utterance, etc. (see the chapter on email below).

The main focus of analysis was qualitative in the sense that the messages were coded in a number of categories such as main content, or purpose of message, for example complaint, request for information, etc. Qualitative analysis of both email messages and traditional letters also rated them on whether or not they conformed to the business letter template (cf. Danet 2001), focusing on salutations, pre-closings, closings and signatures (which was listed by TraSA).

In order to establish factors influencing how people compose their messages, cross analysis has been done by the gender of the sender, the

status of the message (sent for private or for professional purposes); purpose of communication (e.g. question, complaint, etc.), and topic. The two factors in particular assumed to influence this type of electronic communication were: relationship between communicators and purpose of message (in relation to the variable *Situation*), and production conditions and effort of transmission (in relation to the variable *Means of expression*). As mentioned above, both email and traditional letters are asynchronous, which relates to both the variables *Synchronicity* and *Means of expression*. A detailed description of methods and material are found in the chapter on email, which deals with the analysis of written language use and adaptation in email.

Possible Limitations

Some may regard electronic messages to the city council as not being proper email. Although these share most features with normal email, the messages were created in a web form and did not automatically have information about sender address, time of posting, etc. that would have been added if the message had been sent from an ordinary customized email client. This was kept in mind when analyzing the material, and is one of the reasons why generalizations cannot be applied to all email use.

Perhaps the corpora could have been better matched if the available number of traditional letters had been equal to those of electronic letters. In addition, a larger corpus in general would naturally provide a more reliable result from which to draw conclusions. Unfortunately, the researcher had to rely on city council staff to retrieve the letters, and these were not made available.

Another thing to keep in mind is that the conclusions made about language use in email are based on messages sent in a very particular activity; generalizations that would cover all use of language in email cannot be made on such a basis.

3.6.2 Analyses of Language Use in Web Chat

Material

The material for the web chat study consists of data from two sources. The first is a questionnaire inquiring into students' habits and preferences when communicating on the Internet. The second source is linguistic material logged from one of the chat rooms. The results from the questionnaire proved to be the more popular at the time of the inquiry (May 2000).

All in all, 333 students (164 girls and 169 boys) answered a questionnaire that was sent to five upper secondary schools³⁶ (students aged 16-18), and two lower secondary schools³⁷ (students aged 13-15). The questionnaire inquired into the teenagers' habits and preferences concerning electronic communication. In particular, questions concerning their use of e-mail and chat facilities were included. Of specific interest for this study is the question regarding which languages the informants claim to use in everyday communication and which ones they claim to use in their chat communication. As one of the main aims of this study is to analyze how teenagers adapt their language use to online, real time written communication, they were also asked whether they employ strategies such as emoticons and abbreviations in their chat communication. The students were also asked to give examples that they use in their own communication, both in order to find out which they claim to use and which seem to be in use in the chat room that was logged. One question concerned the chat sites participants considered to be the best and most popular ones, and the reasons for this.

One of the most popular Swedish chat rooms at the time was found to be a web chat maintained by a Swedish newspaper³⁸. It was logged for 119 hours, 39 minutes and 15 seconds. During this period, the total number of messages was 44,380. A robot program, or a 'bot'³⁹, was developed for collecting electronic data. The robot logs on to the chat site acting as a human participant and records both the linguistic contributions in the chat room, as well as logging data such as sender (participant/user) and posting time of each contribution in the chat room in machine-readable format.

Table 10. Data from the web chat corpus.

	<i>Web chat</i>
<i>No. messages</i>	44,380
<i>No. words</i>	410,355

³⁶"Gymnasiet" in Swedish.

³⁷"Högstadiet" in Swedish.

³⁸ <http://nychat.aftonbladet.se/webchat/oppenkanal/Entren.html>

³⁹ Leif Grönqvist wrote the 'bot' programme.

Methods of Collection and Analysis

Teachers at the schools were informed about the study, and asked to distribute the questionnaire in class, collect the completed forms and send them back to the researcher. The students were informed about the study by the teacher and asked to participate. Responses were entered into a database, and the web chat to log was chosen on the basis of the answers. Examples of emoticons (also known as smileys) and abbreviations were stored.

The material from the web chat was stored and analyzed digitally, by means of the automated tool, TraSA (described in Section 3.3). This study required the analysis of strategies found to be used frequently in CMC, and are well documented in previous research (cf. Section 2.5). The material was searched for occurrences and use of emoticons, abbreviations of various kinds, use of ‘all-capitals’, extreme use of repetitive keys (punctuation marks), and asterisks. A detailed description of material and how it was collected is given in the chapter which investigates how written language is used and adapted to the conditions of communication in a web chat.

Not everything could be analyzed automatically. A qualitative analysis was conducted manually, analyzing the occurrence and types of abbreviations, and emoticons. The most frequently used words in the chat material were then compared to a list of the most frequent words in written and spoken Swedish (cf. below).

The material for this study was collected from a web chat, which is a synchronous (or quasi-synchronous (cf. Garcia and Jacobs 1999) means of expression. Automatic logging of all activities (i.e. both linguistic contributions from the chat participants and system information such as time of posting and sender) in the chat has an advantage over copy and paste methods which sometimes have been used to gather material from chat conversations (e.g. Karlsson 1997). Large amounts of data can be gathered, and be handled effortlessly using automatic analysis.

Possible Limitations

Web chat is just one of many types of synchronous, real-time chat. The chosen chat room is also just one chat channel; other channels may show different language use depending on the norms of a certain community, for example (cf. Sveningsson 2001). The logged material represents 119 hours, approximately one week, of communication in one channel, and the

language may be used differently during different periods of time depending on, among other things, which participants are logged on and what the subject matter for discussion is.

3.6.3 Analyses of Language Use in Instant Messaging

Material

The material consists of logs of instant messages, from which a number of messages were extracted for analysis. This corpus consists of a total of 8 255 logged messages sent during the period September 1, 1998, to December 31, 1999. The messages were made anonymous during collection so that the original sender cannot be identified, but the logs still contain the essential text of the messages. The messages were sent in a large computer lab containing workstations in several rooms and on several floors of the building. The corpus was ordered into categories of sender location, whether the sender was in the same room as the receiver or not; and sender status, whether the sender identified him or herself or chose to be anonymous. This cross analysis resulted in six categories (see below). In order to make a closer analysis of what people used instant messaging for, a set of 100 messages were extracted from each category⁴⁰, from a total of 8,255 continuous messages resulting in a sample containing 600 messages. These were analyzed manually for what seemed to be the main topic or reason for communication. Automatic analysis of other measures such as mean length of utterance, word frequency, occurrence of emoticons and other graphical means were based on the total number of messages.

Table 11. Data from the instant messaging corpus.

	<i>Instant Messaging</i>
<i>No. messages</i>	8,255
<i>No. words</i>	111,025

⁴⁰ The decision to extract messages in succession was made in order to follow possible dialogs, and thus get some clues to how content should be interpreted (cf. Nardi et al. 2000).

Methods of Collection and Analysis

The aim of this study is to analyze how awareness of presence, as one of many possible factors, affects the content of messages. Real world facts must be considered, not only the "virtual presence" as visualized through WebWho, but also physical presence when the recipient is located in the same lab room as the sender. The message logs have been divided into the following settings:

- *Co-located*: both sender and recipient were located in the same lab room at the same time, hence, it was possible to have physical awareness of each other's presence, and of course to see one another and talk directly face-to-face (both physical and virtual awareness of presence).
- *Distributed*: sender and recipient were located in different lab rooms in the same building, and used the WebWho tool to trace the presence of each other (virtual awareness of presence).
- *Distant*: the sender accessed WebWho from outside the building, using a dial-up connection, to locate the recipient (virtual awareness of presence). The system only allows for receiving messages from outside the university building, and not for sending messages outside.

In order to analyze whether awareness of presence effects message content qualitative, manual analysis of each message topic was carried out by cross-analyzing three categories of sender location (Co-located, Distributed, and Distant as defined above) and two categories of sender status (Anonymous and Identified). WebWho allows for messages to be sent anonymously; the recipient cannot tell who sent the message. The message window for anonymous messages is quite similar to a system error message window. The body of material did in fact show messages that were sent with non-identified senders.

The instant messaging corpus (8,255 total) has been analyzed automatically using TraSA in order to determine the number of messages sent in the different settings, sender location (in the same lab room, elsewhere in the lab, at home or somewhere else away from the lab), and sender status (anonymous vs. identified). The texts were also analyzed for word frequency, which also applies to "word-like" elements like emoticons. The messages were searched for occurrences of emoticons and abbreviations. The WebWho material was compared with respect to word

frequency to spoken and traditional written language, as well as with corpora of email and web chat.

Possible Limitations

Many factors influence language use. The means of expression used in instant messaging is writing, and the messages are delivered almost immediately making it a synchronous mode of communication. Messages were typed on normal computer keyboards, and read in message windows on the screen. There were no technical limits on how long messages could be. In fact, users endeavored on occasion to find potential limits on message length by asking directly or by telling long stories to attempt to discover these limitations, if any. As stated, the focus of this study is on sender location and sender status. The logs represent messages sent within a synchronous means of expression, which allows the sender to be aware of the receiver's presence. Senders could get this information either from the schematic view of the computer lab, displayed on WebWho, from real-world information by looking about in the physical room, or both. Awareness of receiver's presence does not, however, entail that it is the intended receiver that is actually logged in, and displayed by the system. Someone else might temporarily be borrowing a user ID, or temporarily be borrowing the particular workstation at which someone else is logged on. Moreover, the fact that a message pops up topmost on the receiver's screen does not ensure that it actually is read. The receiver might physically be elsewhere, say on a lunch break, but still be logged into the system. These pragmatic difficulties that the user might experience had to be left uninvestigated, as the gathered material did not reveal such information.

A further aim of this analysis is categorizing the messages for main content, or purpose of communication, supposed by this study to be one of the factors influencing how messages are composed. It was only possible to make guesses about the purpose of some messages. If one had knowledge about the relationship between communicators, contextual and situational information, access to their shared background knowledge, one have more basis for assumptions regarding message content.

3.6.4 Analyses of Language Use in SMS

Material

The material consists of 1152 SMS messages from three different corpora: anonymous entries to a web page form; messages forwarded from informants; and a corpus of private messages collected from friends and family of the researcher. Material also consists of informants' entries in four SMS-diaries, and semi-structured interviews with four informants, which serve as complementary information to support the researcher's own text messaging experience and intuition.

Table 12. Data from the SMS corpus.

	<i>SMS</i>
<i>No. messages</i>	1,152
<i>No. words</i>	17,024

Methods of Collection and Analysis

Mobile text messaging is an asynchronous mode of communication. The production and perception conditions for SMS are particularly constrained by awkward text input and limited message size, compared to the modes of CMC investigated in the other case studies. SMS messages are also mostly sent between people who know each other; situational factors such as relationship between communicators are assumed to have a strong influence on message composition. Data was collected using three methods: through a web-based questionnaire, informants and from friends and family. Informants were asked to answer questions anonymously on a web page form about their use of SMS, and to copy a message – character by character – that they had sent either by mobile phone or web service. 107 people answered questions and shared their messages. The data from this web-based form was entered automatically into a database (see Section 7.7.2). Invitations to join the study were extended by the researcher during interviews for daily newspapers and through a link on the researcher's personal web page. All entries to the web-based data form were anonymous.

The second method used for data collection consisted of recruiting informants to join the research project. They were asked to keep notes on their text messaging in a user diary, in which they filled in answers to

questions about each SMS that they sent or received during a period of a week (or longer if they chose). A preformatted notebook for this was provided. The informants were paid for keeping the diaries. During this period the informants were asked to forward text messages that they had created themselves to the researcher's mobile phone. The forwarded messages were transferred from the phone to a computer using a serial cable and software, and were finally stored in a database. Naturally, the informants were told that they had the choice of forwarding each and every message if they so choose, or of selecting only messages that they felt comfortable sharing. The possible reasons for not sharing some messages were followed up in the interviews. Four informants finished the task – two female and two male, aged 12-25.

Each informant who kept a diary and forwarded messages was interviewed. The interviews were semi-structured, consisting of a number of questions to guide the talk and of free associations and informal chatting in which the informants could comment and add remarks on both the research itself and about their own communication via mobile phones. The interviews took about 1–1,5 hours each, and were audio recorded with the informants' consent. The recordings served as audible research notes. The researcher consciously allowed the topics of conversation in the interviews to stray, as these were intended to be informal. All of the informants' answers were not coded into categories; the main point of the interviews was to follow up the material which the informants supplied, to support assumptions made from taking part of the messages, and to give an opportunity to inform them some more. Special attention was paid however to answers to questions about the informants' use of predictive text entry, type of mobile phone, their main text topics and recipients, whether or not they claimed to be conscious of correct spelling and conform to normative punctuation.

The third method for collecting SMS messages was by asking friends and relatives to part with their messages during a given period of time. This method resulted in the largest of the three data sets: 790 messages were collected from 16 persons (8 female and 8 male).

The data files that made up the SMS corpus were separated into different sets, according to manner of production (on a mobile phone or on a keyboard on the web), and according to informant group. TraSA was used for automated, quantitative analysis, and the results were compared to data from corpora of spoken and traditionally written language, as well as compared to the other CMC modes included in the present study.

Possible Limitations

The study of SMS messages could have done well to have included data from a wider range of people spanning many more users of different age, gender, linguistic and social groupings, or geographical background. It proved difficult to recruit people to part with their messages. One of the reasons that was voiced in the interviews, is that SMS is experienced as being very personal, even more so than private email.

3.7 Overview of Data from Four Modes of CMC

Table 13. Overview of Data from Four Modes of CMC.

	<i>Email</i>		<i>Web chat</i>	<i>Instant messaging</i>	<i>Mobile text messaging</i>
	<i>Email</i>	<i>Traditional letters</i>	<i>Open chat</i>	<i>WebWho</i>	<i>SMS</i>
<i>No. messages</i>	183	41	44,380	8,255	1,152
<i>No. words</i>	11,660	17,771	410,355	111,025	17,024

Table 13 above gives an overview of how much material each of the four case studies consists of.

3.8 Chapter Summary

This chapter provided a general outline of how material was chosen, and which methods were used for collecting and analyzing it. Automatic analysis was used generally to investigate material from all four modes of CMC (email, web chat, instant messaging and SMS) that are part of this dissertation. Each chapter that treats individual case studies will provide more details about the particular methods that are relevant to each study respectively.

In the next section, we turn to the first of the case studies: the use and adaptation of written language in email.

Part II

4 Electronic Letters to Unknown Authorities

4.1 Introduction

This study is an analysis of electronically transmitted letters (i.e. email messages) and traditional paper letters (i.e. handwritten or typewritten letters on paper) from citizens to the authorities of the city council in Göteborg city⁴¹.

The norms for email are still in the process of being established. People seem to be uncertain of how to formulate themselves in this still rather new medium. Most studies of email have analyzed email messages in public mailing lists, in which the messages are one-to-many dialog. This is a study of public one-way, one-to-one email messages, where the receiver is an unknown authority.

The overall purpose of this particular sub study is to try to establish which factors influence how people formulate themselves in an asynchronous, text based electronic medium. What are the differences between traditional paper letters and email? Do email messages to authorities conform to the business template of traditional formal letters, or is it the ease and rapidity of the electronic medium that pose the greater influence on the way senders formulate their messages, or are there combinations of other factors?

⁴¹ Data for this study was gathered in 1998. It was published as an article in 2000 (Hård af Segerstad 2000b). It should be pointed out that since this study was carried out and was published many other studies of email have been published which are not referred to in the present text.

Results from this study confirm results from previous studies of email (Du Bartell 1995; Herring 1996; Danet 2001), suggesting that email messages to authorities are less formal; shorter than the formal business template. Email often seems to serve communicative purposes replacing phone calls (Severinson Eklundh 1994).

Many people today frequently use computer-mediated communication (CMC), both for private and for professional purposes. One of the reasons for using the fast transmitting computers is the convenience it brings: email has a number of advantages compared to traditional written letters. It is an easy, fairly effortless, fast way of getting in touch with people. It is also low in cost. Transmission time is very much shorter than that of the traditional postal service, and the electronic means of expression allows replies to be received within minutes. It is also possible to attach files of various kinds (e.g. audio files, digital images, word processor documents, etc.) to email messages.

As email is an asynchronous, written means of expression, certain demands and constraints are put upon both sender and user. Spoken interaction is multimodal, making use of several channels simultaneously for sending information. Written interaction has to rely on the single, linear channel of vision for communicating textual messages. Previous studies have shown that strategies such as the use of emoticons or abbreviations have been developed to overcome the difficulties of the written medium in order to avoid misunderstandings and ambiguities, and still be able to make use of the speed of transmission that CMC technologies allow for.

It is possible that the evolving possibilities of electronic communication may change the way people approach each other in writing. What situational factors, such as psychological and contextual, influence the way writers compose their electronically transmitted messages? Studies have shown that purposes for communication, topic and medium for communication play a part in the way messages are formulated (Du Bartell 1995; Baym 1996; Hård af Segerstad Forthcoming). Other factors such as the relationship between sender and addressee (Danet 2001), grounding and closure on the actions (Clark and Brennan 1991; Clark 1996) also impact composition. The means for communication one selects to interact through impacts on the potential forms of language which may be manifested (Du Bartell 1995). The ease of access for sending messages and the user's relative anonymity might also influence the way in which electronic communication is formulated.

Other studies have shown that the faster the medium, the more like spoken language the written messages get (cf. Horowitz and Berkowitz

1964). Previous studies of email have shown that messages often tend to be more informal both in terms of composition (salutation and closing conventions) and form (spelling, syntax) (cf. Herring 1996).

4.1.1 Aim of the present study

The overall purpose of this dissertation is to analyze how written language is used and adapted to suit the conditions of a number of modes of CMC. This particular sub study aims at analyzing what the underlying factors might be behind how people compose the electronic messages (email) that they send to an unknown authority at the city council of the city of Göteborg, Sweden. A comparison with traditional paper letters that were sent for the same reasons (see Section 4.6) will be made. The variable Situation, introduced in the background chapter, has to do with relationship between participants. Intuitively, one expects that the relation between sender and addressee will influence message composition and language use; letters approaching “authorities” ought to conform to the formal business template (Danet 2001).

The variable *Means of expression*, also introduced in the background chapter, has to do with production and perception conditions. The ease and rapidity of production and transmission is hypothesized in the present study to make email messages more “speech-like” and less formal than traditional letters, which conform to the business template (see Section 4.3.1). At the same time, email messages are still *written* and rely more on the typed words than contextual information; in this sense email messages tend to be more “written-like”. The written mode may also make people feel that they may remain relatively anonymous and stay “hidden” behind the text, as it were. Most email software automatically includes the sender’s name and email address, which one would expect would lead the sender to omit his or her name in closing the message.

The analysis was made by examining whether the traditional paper letters and the email messages in this study conform to the formal business template; with respect to functional moves such as salutation and closing conventions (cf. Danet 2001), in combination with an analysis of contextual factors such as the status of the sender, the medium, the purpose and topic of communication.

4.2 Previous Studies of Email

4.2.1 Analysis of messages in an electronic Swedish conference system

There has been extensive study of electronic mail as a medium for dialogue. Severinson Eklundh (1994) argues that the computer medium may serve communicative purposes previously reserved for direct, spoken conversation. Her analysis of a body of messages from the Swedish conference system COM (Severinson Eklundh 1987), showed that messages were structured into coherent dialogues in a way distinct from other forms of written communication. Dialogue sequences appear in email simply as a result of linking messages and their replies. The present study is an analysis of one-way, monological messages that consequently do not have any linking between messages. Furthermore, Severinson Eklundh points out that the exchange of email messages may contain a range of speech-like or “conversational” features. This applies both to the character of individual messages and to the structural features of the entire dialogue. Severinson Eklundh states that the email messages in her study typically were short compared to regular letters, and that almost all messages used a direct, informal style of address. These findings can be expected in the present study too.

4.2.2 Analysis of messages in electronic mailing lists

Du Bartell’s study (1995) of the features of the messages in a mailing list suggests that the spoken and written-like characteristics in a written medium result from the constraints imposed by the computer medium – the machine architecture. The computer medium permits texts that seem both written-like and spoken-like. Du Bartell argues further that CMC messages display linguistic characteristics typically associated with spoken language and other forms of written language in addition to linguistic features specific to the medium. Du Bartell explains that we expect written language to be edited, planned, articulated without recourse to non-standard constructions, slang and vulgar expressions. From speech, we expect more or less the opposite: we expect slang, non-standard grammatical constructions, sudden topic shifts and spontaneity. Du Bartell’s contention is that CMC gives us these in writing. Computer-

mediated discourse exhibits the type of grammatical constructions that appear in non-edited non-standard spoken language of face-to-face interaction. As has been pointed out several times above, CMC is used in many activities; to claim that its characteristics are due to the computer medium alone seems to be a rather simplified view of how language use varies. Situational parameters have to be considered too.

Herring (1996) analyzed the schematic organization of electronic messages posted to two academic mailing lists, one mostly male and the other mostly female, in order to evaluate the popularly held view that men and women use email for different purposes (information exchange vs. social interaction). Her results did not support the stereotype, but showed that women and men's messages are structured differently. Female users were found to exhibit alignment, while male users exhibit opposition towards their addressees.

The basic electronic message schema was analyzed into functional moves: salutation, introduction, body, and close. She concludes from analyzing 136 messages that surprisingly few messages were preceded by a salutation (only 13% on average), and fewer yet were followed by a complimentary close or a postscript (Herring 1996).

Both Herring and Du Bartell explain the relative lack of epistolary conventions as being partly due to the fact that a header is added automatically to each message by the electronic mailer, including a separate line for whom the message is "from", whom it is addressed "to", and the date and time of posting. Partly as a result of having a subject displayed, email messages frequently omit even the typical salutations and farewells associated with other modes of communication, regardless of whether the communicators⁴² know each other. Email messages display rather informal register characteristics even between unknown persons (Du Bartell 1995), a fact that is highly relevant for the present study.

Baym (1996) argues that although CMC is written it is marked by many features associated with face-to-face interaction. She takes a more complex view than Du Bartell, and contends that her study of Usenet messages⁴³ showed that email in discussion groups have hybrid characteristics between oral, written, interpersonal, and mass communication. Baym concludes that the message features of her study stem from five interrelated factors, which conform to the variables *synchronicity*, *means of expression* and *situation* that

⁴² Du Bartell employs the terms 'speaker' and 'listener' regardless of means of expression.

⁴³ Usenet is an enormous collection of topically organised discussion groups distributed through the Internet.

were put forward in the background chapter above: the Usenet medium, the institutional context of work, the topic, the participants' gender, and the social context which the participants strive to create.

4.2.3 Analysis of private email

Because of ethical considerations, there have not been many studies which analyze private email. Most studies of email have investigated messages sent in public mailing lists. For ethical reasons, and in order to analyze private one-to-one email messages, Danet (2001) analyzed portions of her own email correspondence. Her focus was on letters that were sent to her by people who did not know her but knew her name and status as a senior university professor. She remarks that:

A writer of a first letter is likely to take special care in its formulation. Traditional norms for letter form are likely to be salient, and writers are likely to be especially conscious of the impression their message may make on the recipient.

Danet analyzed her email messages holistically, and coded them for whether or not they conform to the criteria of the British-American business letter template. Letters were coded for openings and closings, with respect to whether they conformed to the business letter template only if they contained both an appropriate opening and an appropriate closing. The letters were also coded for use of abbreviations, spelling, typography, punctuation, and use of exclamation points. She found that none of the letters of her corpus conformed on all her criteria, and that variability was extreme. Most letters conformed to expectations regarding syntax and vocabulary, as well as those for spelling, typography and layout, but almost none followed paper letter practices regarding openings and closings.

Danet suggests that the new medium invites informality even in business or official contexts. Unlike Du Bartell, Danet argues that this is not just due to the technology per se, but convergence with a general trend that she believes to have been in place already. She remarks that the novel medium can facilitate changes of style and substance in much shorter time than paper letters would have done. Style, or register, may apply to several domains of human communication and action (cf. Allwood 1976; 2000). As has been discussed above, language varies and is adapted according to several interdependent parameters.

Danet concludes that the language of email is in a state of transition. She predicts, among several things, that informal, partially speech-like email style will increasingly characterize public as well as personal communication. How our normative expectations will change to provide increased legitimacy to a more informal style remains to be seen.

4.2.4 Closure on actions

Clark (1996) argues that a fundamental principle of intentional action is that people look for evidence that they have done what they intended to do. He argues that people need closure on their actions, and that to get closure on an action, one looks for evidence that it has succeeded. This principle applies to all kinds of intentional actions. Evidence of success must be valid to be useful; it must be reliable and interpretable. Moreover, evidence must also be easy to get, economical in effort, as well as timely. Without such evidence, one may try the action again, or try to repair what went wrong. People engaged in spoken conversation ordinarily go to some extent to reach joint closure on their actions. See the importance of OCM and ICM features in interaction (cf. Allwood 1976; 2000) as discussed above (see Section 2.2.6). An answer to a question gives evidence that the question is perceived and understood. This applies not only to spoken conversation, but also to written communication as well.

4.3 The present study: Electronic Letters to Unknown Authorities

The analyses of email reported above by Severinson Eklundh (1987; 1994), Du Bartell (1995), Herring (1996), and Baym (1996) all concern email dialogue. Email dialogue consists of an ongoing discussion comprising a series of messages that are interconnected; a person sends a message that is met with one or more replies, sends off another message or replies in his or her turn. Mailing list messages are contributions in a many-to-many interaction. However, the present study deals only with single one-way messages⁴⁴ from private individuals to a remote and unknown “authority”. Like Danet’s study, the material consists mostly of “first letters” from people who do not know the recipient personally. Unlike Danet’s material,

⁴⁴ All email messages were in fact eventually answered, but that material was unfortunately not given for analysis and thus not included in this study.

the senders did not know the recipient by name, but possibly only had a notion of the higher status of the authorities at the city council.

As discussed in the background chapter, written language in general lacks some of the information cues that are conveyed in spoken language. A spoken contribution conveying words such as "It's your fault" may not run the risk of being taken as an accusation or an insult despite friendly intent. The listener takes into consideration more than just the words that are uttered. The speaker and listener's shared background knowledge, the context in which the words are uttered, non-verbal information picked up from the tone and intensity of voice, facial expressions, gaze, gestures etc. can all add up to an interpretation of the utterance as a jesting remark, sarcasm or the like. Had the same contribution been conveyed in nothing but plain text, the receiver would be left with much less information to go by when interpreting the message. Text-only communication without clarifying comments can be very ambiguous and difficult to decipher.

Previous studies (cf. Werry 1996; Witmer and Katzman 1997, for example) of CMC have shown that some distinct strategies in written language in chat rooms have emerged for the purpose of overcoming some of the constraints on written language. Strategies such as the use of:

- Emoticons, in resemblance of facial expressions
- Abbreviations and acronyms
- Words or phrases written in either capitals only or all lower-case
- Extensive use of punctuation marks
- Fonts and colors
- Words or phrases framed in asterisks

The above features are examples of graphic means that are used to enhance and adapt written language and to prevent misunderstanding. Some of these are developments of written language specific for CMC such as emoticons and the specific use of asterisks. Others have been used for the same purpose in traditional written language, for example abbreviations or punctuation marks, but perhaps to a lesser extent than in CMC.

According to Horowitz & Berkowitz (1964), some of the many differences between spoken and written expression are due to the greater ease of speaking one's mind than writing it. Any mode of written communication⁴⁵ that increases the ease of production could be expected to result in the production of cognitive and linguistic material more nearly like that produced in spoken expression (Horowitz and Berkowitz 1964).

⁴⁵ Horowitz & Berkowitz compared handwriting, type writing and steno typing in 1964.

4.3.1 The Swedish business letter template and norms for informal letters

The Swedish business letter template differs slightly from the British-American template as described by Danet (2001). Generally, the standard paper business letter is supposed to be cast in a formal style, using language appropriate to formal situations. The British-American formal letter opens with a salutation or greeting (“Dear Sirs”, etc.), skips two lines between the salutation and the body of the message and between the body of the message and the closing. The letter is signed with the sender’s full name, and perhaps also title and affiliation. Furthermore, it conforms to the normative conventions of spelling, punctuation and orthography (e.g. sentence initial capitalization, no contractions and no typos).

The Swedish formal business template opens with the topic or concern of the letter, often underlined or in bold face (Stenson 1997). It is not Swedish practice to open with a salutation. The sender’s address as well as the receiver’s and the date are often placed at the head of the letter. Like the formal British-American business template two lines are skipped between the salutation and the body of the message and between the body of the message and the closing. It is signed with the sender’s full name and, like the British-American template, perhaps also with the sender’s title or affiliations. The Swedish formal letter also has to conform to the normative conventions of spelling, punctuation and orthography.

On the other hand, informal letters in Sweden, such as letters between pen pals, most often open with a salutation or greeting. The informal *hej* (“hi”), or variants (*hejsan*, *tjenare*, *hallå*) are often used. Danet also notes that in the Anglo-American tradition, personal letters have always been more conversational and informal than business or official ones.

4.4 Material and Method

4.4.1 Material

The aim of this study is to compare email messages and traditional paper letters in order to find out where the differences between these lie, and what the reasons for these differences might be. Material for the study consisted of electronic as well as traditional paper mail from citizens to the

city council in Göteborg, Sweden. Letters of both kinds were sent for approximately the same reasons, and were comparable. Documents at public institutions in Sweden are public material, open to anyone who wishes to examine it. This applies to traditionally produced (paper letters) as well as electronically received documents (email messages).

Email data

The material consists of 183 messages sent between April 1, 1998, and August 31, 1998. The corpus consists of 11,660 words. The web page of the city council provides a service for “Questions & Answers” at which anyone may send messages of any sort for which they hope the people at the city council might be able to provide help. Originally, the service was designed for questions and remarks concerning the web page itself, but people sent messages of all kinds, and it was eventually decided that this service should pose no restrictions of what kind of messages to handle.

Figure 6 is a screen dump of Göteborg’s web page form. People were asked to fill in a form with the following fields: “Sender” (email return address in order to receive an answer), “Subject/ topic” (subject of the message), and lastly “Text” in which the message was composed. After filling in the form and writing the message, clicking a “send” button sent the message⁴⁶. The form looks basically like most email clients, but the return address had to be typed manually in order to get a reply, and was not automatically added as it would have been in a normal customized email client.

Directions and regulations at the city council only allowed the email messages to be printed at the office and sent as paper printouts rather than being forwarded electronically. The copies were scanned and stored in digital format, and the paper copies were kept for proofreading and for checking that the scanned versions were entered accurately.

⁴⁶ The web page-form at
<<http://www.goteborg.se/wwwdb/gbgwww.nsf/fragorochsvar>> (5 August, 2000).

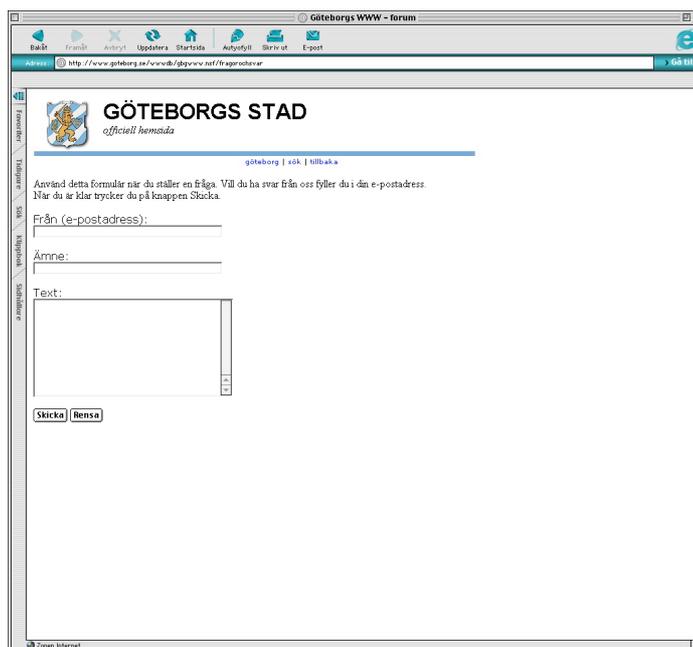


Figure 6. The web page form at the city council, from which the email messages were sent.

Traditionally written paper letters

Data from 41 traditionally written paper letters is used in this study: 7 were hand written, 26 were typed on a computer, and 8 were typewritten. The corpus consisted of 17,771 words. The letters were of the same kind as the email messages: open⁴⁷ letters from citizens asking questions, requesting help, etc. The comparatively low number of traditional letters is due to problems that the staff at the city council archive had for abstracting material out of their filing system. The corpus of traditionally written letters was judged to be large enough to be comparable to the email messages.

The traditionally written letters that were typed on typewriter or computer were possible to scan and store digitally to allow for automatic analysis. Handwritten letters were copied by manually typing the text in a word processor, saving it as electronic text.

⁴⁷ “Open” in this sense means that the letters were addressed to anyone at the city council, and not someone particular.

Table 14. Overview of data from the corpora of email- and traditional letters to the city council.

	<i>Email</i>	<i>Traditional letters</i>
<i>No. messages</i>	183	41
<i>No. words</i>	11,660	17,771

4.4.2 Method

Both traditional letters and email are asynchronous modes of communication. The material for both corpora was collected from the city council with the help of the staff at the information and archives departments⁴⁸. It was stored, and analyzed digitally using the automated tool, TraSA that was described in the background chapter. For this study it was used for quantitative analysis of word frequency, the occurrence of abbreviations and punctuation marks, closing and introductory words in messages, mean length of contribution, etc.

The main focus of analysis was qualitative. Both email messages and traditional letters were analyzed and coded for whether or not they conformed to the business letter template (cf. Danet 2001). Focus was on salutations, pre-closings, closings, and signature.

In order to establish factors that influence how people compose their messages, cross analysis was made of the gender of the sender, the status of the message (private or professional), type, or goal, of communication (e.g. question, complaint, etc.) and topic of the messages.

4.5 Results

4.5.1 Overview of the corpora

The aim of this study was to analyze underlying factors that might motivate how people compose electronic messages (email) that they send to an unknown authority. Situational factors such as gender of sender, aim or type of communication, as well as whether the message was sent for

⁴⁸ My warmest thanks to the friendly and helpful staff at the city council, Göteborg, Sweden!

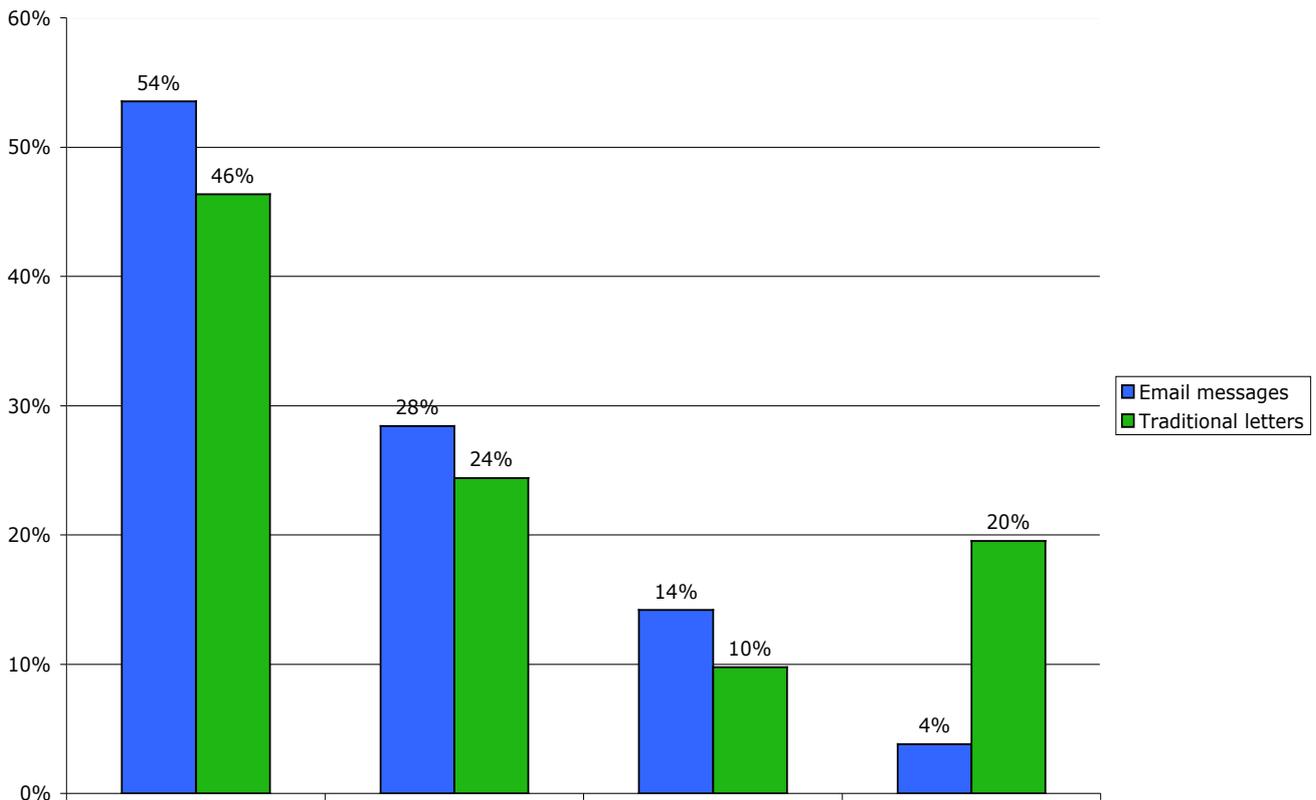


Figure 7. The distribution of messages sent by females, males, unknown, and multiple senders in the email and paper letter corpora (percentage of total)

private or professional reasons were considered. An overview and comparison of the corpora will be presented below.

Sender

An overview (Figure 7 above) of who sent the messages shows that most of the messages in both corpora were sent by men. When a message contained no clues as to gender, it was classified as being sent by an “unknown sender”. Often, the sender’s email address provides clues to the gender of the sender, however, in cases where the sender does not identify himself or herself sufficiently, and the address is cryptic (unrelated series of letters and numbers) the message was classified as “unknown sender”. Some messages may have been sent by someone using someone else’s email account, a fact which must be ignored since there is no way to determine this from the current corpora. A number of messages were sent by more than one sender – a family or two people working together, for instance, and thus classified as “Multiple senders”.

Aim, or type, of message

Contributions in communication are frequently multifunctional; they serve several purposes at the same time (cf. Allwood, Nivre et al. 1993). Both paper letters and email messages were coded for main aim or type. The most common category in the email data was questions (78%). An illustration of such an email message is given in Example 1 below. The reason that only examples are taken from the email corpus to illustrate the various categories is simply that the traditional letters generally are too long to include in this text.

Example 1. jag undrar bara om vandrarhemmet heden
finns kvar???
*[I just wonder if the youth hostel at heden
still exists???)*

The main aim, or type, of traditional paper letters was complaints (44%), illustrated in Example 2., and requests or appeals for help (28%) (Example 3).

Example 2. Varför finns det inga kartor över
Göteborg????? JAg har suttit här och letat
efter karta och adress till Svenska
mässan!!!!
*[Why are there no maps of Göteborg????? I've
been sitting here looking for a map and
address to The Swedish Exhibition and
Congress Centre!!!!]*

Example 3. Kan ni vara snälla och hjälpa mej med namn
och adresser teater- och musikal-skolor i
Göteborg?
*[Could you please help me find name and
addresses theatre and musicalschools in
Göteborg?]*

Most of the messages were questions or requests for information of various sorts. Other messages were statements/ remarks or complaints (Example 4 below).

Example 4. Ämne: Fel på sida.
<http://www.goteborg.se/wwwdb/gbgwww.nsf/soksidor/Adresser?Ope Document>

URL, en ovan ger nedanstående felmeddelande om man använder länken för att skicka mail till redaktionen...

Error 500
HTTP Web Server: Unknown Command Exception
[Subject: Error om web page.

The URL above gives the error message below if you use it to send mail to the editorial office...]

Error 500
HTTP Web Server: Unknown Command Exception]

Suggestions of the type illustrated in Example 5 below, also occurred.

Example 5. Hej!
Jag föreslår att Ni lägger in en länk på sidan om folkomröstningen till de intresseföreningar som initierat och driver frågan. Fler ges då möjlighet att informera sig och sätta sig in i denna mycket komplexa fråga.
[Hi!
I suggest that You add a link to your page about the referendum and the special interest groups who started and are pursuing the question. More people will then be able to be informed and find out about this very complex matter.]

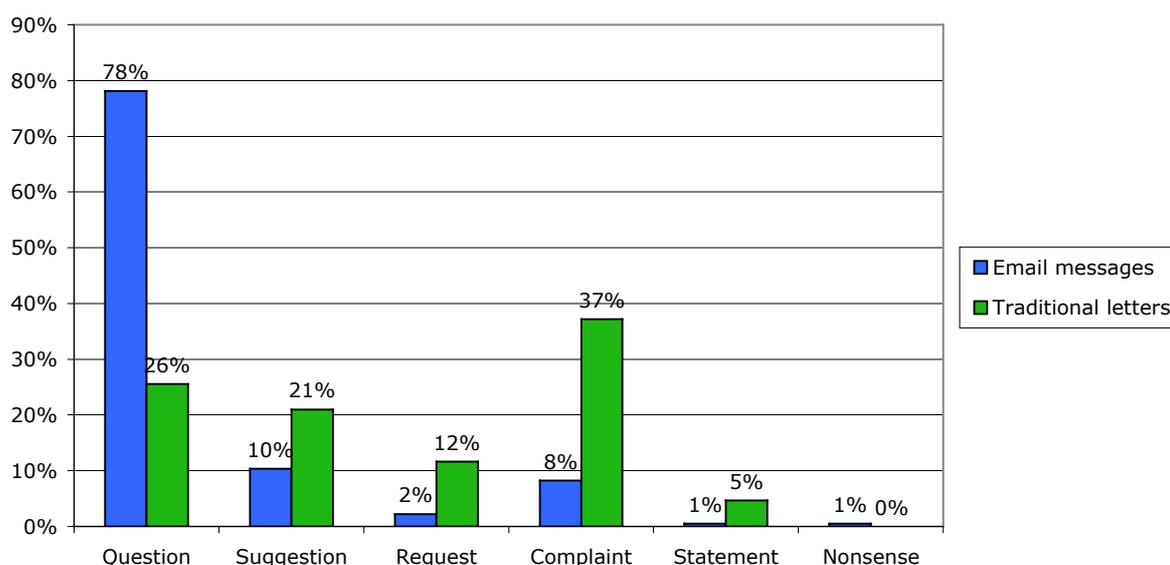


Figure 8. Distribution of aims, or types of message in the email and paper letter material (percentage of total).

The largest category of the paper letters is complaints. In email cases complaints were mostly about information failure on the web page, “technical problems”. Several messages were complaints about matters in the city, opening hours of the museums or parking facilities, for instance. A number of messages contained offers of services or suggestions; people offered to work or to send information that could be useful for the web page or for the people at the city council, or suggested co-operation. A few email messages seemed to be sent in jest, or perhaps just to gauge reaction. An illustration is given below (Example 6)

Example 6. När går nästa spårvagn från mölndal till GBG??
 [When is the next tram from mölndal to GBG (Göteborg)??]

As mentioned above, many of the messages are multifunctional and may, for example, contain both a question and a remark of some kind. These were classified according to which seemed to be the primary reason for communication.

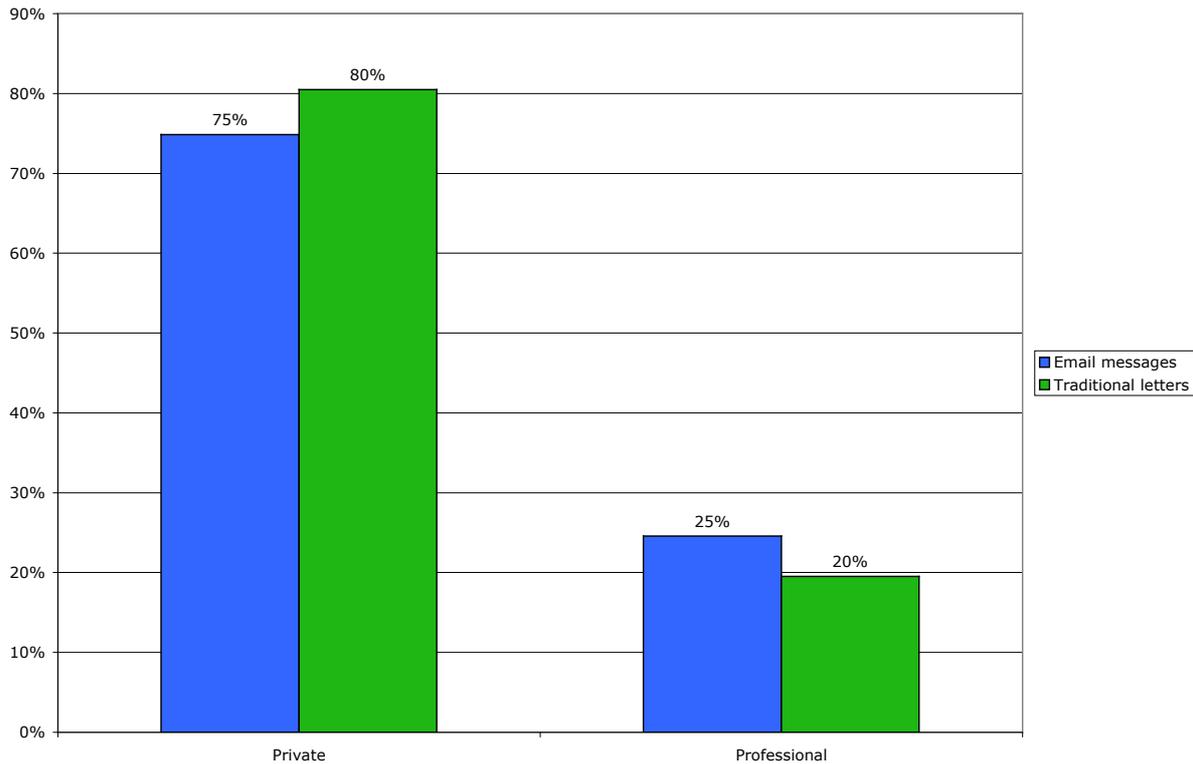


Figure 9. Email messages and paper letters sent for private or professional reasons (percentage of total).

Private or professional reasons for communication

Figure 10 below illustrates that most of the email messages clearly were sent for private reasons, i.e. citizens of Göteborg or others seeking information of various kinds. 75% of the email data, and 80% of the paper letter material, were sent for private reasons. 25% of the email messages, and 20% of the paper letters, were sent for professional, or business related reasons. Whether a message was sent for private or professional reasons was determined according to topic, signature and sender's address. Most of the messages were found to be sent for private purposes.

4.5.2 Analysis of paper letters

Analysis of the paper letters was influenced by Danet's study (Danet 2001). Danet categorized her email into groups conforming more or less to the business template with regards to salutation and closing conventions. Her

email corpus was smaller in size, which allowed for easier scrutinizing of each message. A closer analysis of each letter in this study with respect to spelling, punctuation, occurrence of informal syntax like Danet's would have been interesting; however, the type of analysis that was made in the current study was judged to be sufficient for present purposes.

The email messages as well as the paper letters of this study were analyzed for salutation conventions and whether or not the messages were signed with the sender's name. The Swedish business letter template does not require salutation. A topic like the subject line of the email format is taken to be the equivalent. Four categories emerged:

1. *+Salutation/+signature*
Messages introduced with a salutation and signed with the sender's name.
2. *+Salutation/-signature*
Messages introduced with a salutation, but not signed with the sender's name.
3. *-Salutation/+signature*
Messages left without salutation, but signed with the sender's name.
4. *-Salutation/-signature*
Messages neither introduced with a salutation nor signed with the sender's name.

An overview of how the different categories of salutation and signature were distributed in the corpora is given below.

4.5.3 Salutations and signatures

+Salutation/+Signature

As illustrated in Figure 10, the category +salutation/+signature occupied 48% of the total number of email messages, and 2% of the total number of paper letters. It seems that email follows the US/British letter style, and not the Swedish business letter format. This category of the email messages (+salutation/+signature) covered 48% of all questions, 50% of all suggestions, and 35% of all complaints. The only two instances of emoticons occurred in this category. Both messages were informal in spelling and syntax. Emoticons did not occur at all in the paper letters.

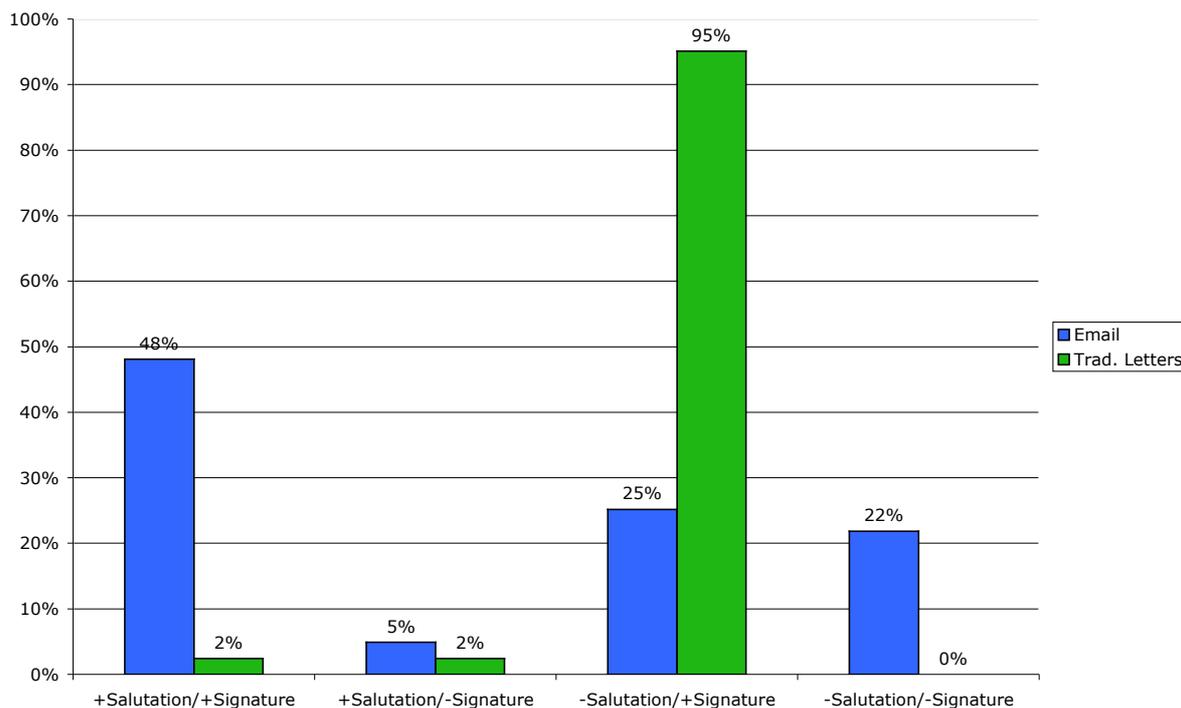


Figure 10. Distribution of email and paper letters into categories of salutation and signature (percentage of email and paper letter corpora respectively).

+Salutation/-Signature

The second category, +salutation/-signature, consisted of 5% in the email material and 2% in the paper letter data. Only 5% of all email messages opened with a salutation but were not signed at all. The two types of email messages in this category (+salutation/-signature) were questions and complaints. This group featured 6% of all questions, 7% of all complaints.

-Salutation/+Signature

The third category, -salutation/+signature. It occupied 22% of all email messages and 95% of the total number of paper letters. Thus, 22% of the email messages had no salutation, but were closed with a signature. This category conforms most closely to the traditional Swedish business letter format, if we allow for the subject line to serve as the subject opener of paper letters.

For this category in the email material we find 22% of all questions, 27% of all suggestions, 50% of all requests, 42% of all complaints.

-Salutation/ -Signature

The last group, -salutation/-signature, consisted of 22% of the total number of email messages and was not represented at all in the paper letter data. These messages conformed least to the business letter template due to having no salutation and no signature. Only four of these had a pre-closing of some sort. This group is the most “email like”. It is also in this category that we find the most unknown senders (11% of all unknown senders). This fact is probably due to the email format, which normally gives name and address of the sender automatically (cf. Herring 1996).

In this category we find 23% of all questions, 22% of all suggestions, 14% of all complaints, and the only clear example of a nonsense message.

4.6 Conclusions from the categories

As the diagram in Figure 10 above illustrates, the email messages collected for this study are distributed over the four categories with great variation, comparable to Danet’s findings (Danet 2001), whereas the paper letters only figured in three categories. In this sample, email style is more varied. Moreover, it seems that the email messages and paper letters show almost opposite features.

None of the traditional letters were categorized into -salutation/-signature or + salutation/- signature. All but two paper letters were consistent with the Swedish business template: not preceded by a salutation but ended with a signature. One of these was written in English, and followed the US/British business letter format. The same category of email messages consisted of 25% of the email messages. The largest category of the email messages, 48%, opened with a salutation and closed with a signature, only 2% (N=1) of the traditional letters did so.

Though these results do confirm the results of previous studies of email, they may seem contradictory at first glance: Herring’s findings from the mailing lists suggested that surprisingly few messages were preceded by a salutation (Herring, 1996) - only 13% on average, where we find 48%. This is somewhat incongruous if the business template referred to requires, as the British-American formal letter, a salutation. The Swedish template does not make that requirement. This suggests that the email messages of this study are less formal than the US/British business letter format, which is in line with previous studies. It seems that they are more formal than “proper private email”, though.

5% of the email messages were preceded by a salutation but not signed with the sender's name – this category scored 0 in the traditional paper letter corpus.

The most “email like” messages, the ones with neither salutation nor signature would perhaps be expected to be more frequent had the messages been posted in an ordinary email client for private use. The messages in this study were composed and posted in an electronic form at a web page where the sender's name was not added automatically. Even so, 22% of all email messages were of this type. This category had no equivalent in the paper letter corpus. It might be the case that people forget to sign on this site because they are used to not signing their own email, or that people are unsure of how the letter will appear when it is received.

The last category described above only appears in the email corpus and does not appear at all in the paper letter data. These facts confirm the argument that the normal email architecture and the information automatically given (“From”, date and time of posting) make it unnecessary to retype such information. In paper letters, on the other hand, it is necessary to include name and address manually if the receiver is to know who the sender is. Nothing gets included automatically.

Previous studies have suggested that email messages are shorter than paper letters (cf. Severinson Eklundh 1994). This study clearly confirms this fact: the mean length of the email messages was 63.71 words, and the paper letters were 433.43 words respectively.

4.6.1 Salutation conventions

The messages were analyzed for different forms for salutation. Of the email messages, 98 opened with some sort of salutation, 86 were sent without salutation. All salutations were informal (variations of “Hej” [“Hello” or “Hi”]) followed by variants of conforming to the Swedish norm for punctuation: exclamation mark, comma, period or no punctuation at all. Note that exclamation points and periods are not normal punctuation for letter salutations in either Swedish or English.

Herring's and Du Bartell's (see Background) respective claims seem plausible: this is a common characteristic most probably stemming from the fact that both topic and sender are announced in the header of each message.

Of the traditional paper letters, only two were opened with salutation, one with the informal “*Hej!*” (“Hi!”). The second one was written in

English for professional reasons, was introduced with “Dear Sirs”, and conformed in all ways to the US/British business letter format.

4.6.2 Closing conventions

44% of the email messages had closing phrases of some sort. 3% were closed with the signature only, and 17% were sent with neither closing nor signature.

The use of abbreviations in closing was common. 19% of the email messages closed with some variant of abbreviating the normative, formal closing phrase (cf. Stenson 1997) “*Med vänliga hälsningar*” (literally: “With friendly salutations”). The abbreviated phrase showed a wide range of variation. Examples are m.v.h., MVH, mvh, M.V.H. None of them is the correct way of signing a formal letter. 23% of the email messages were closed with variants of the formal “*Med vänliga hälsningar*”, for example, “*vänligen*” [“in a friendly manner”], “*Vänlig hälsning*” [“Friendly salutation”].

40% of the traditional paper letters closed with the full formal “*Med vänliga hälsningar*”. None of these closed with any abbreviated form. 24% were only signed with the sender’s name. 8% closed with the date and sender’s name. 28% closed with some pre-closing (“*Tack på förhand*” [“Thanks in advance”], etc.) in combination with the sender’s name.

4.6.3 Word frequency

Concerning choice of words influenced by the situational variable, an analysis of the most frequent words in both corpora shows how the activity of interaction is reflected. The corpora of email and traditional letters to the city council were compared to corpora of spoken and written language at the department of linguistics, Göteborg University⁴⁹. In the list of the most frequent spoken words, these are rendered in their written equivalent, supplemented by an index number or letters in brackets for disambiguation. The second most frequent spoken word, see Table 15 below, [e₀] stands for [är=is, are] and disambiguates it from e.g. [e₁] which indicates hesitation sound. In the same way, [å₀] in the table below stands for [och=and], which disambiguates it from [å₁] which indicates [att=to]. The word [ja{g}=I] is disambiguated from [ja=yes].

The most frequent token in email and traditional letters, as well as the corpus of written language, is the period [.]. The word *jag* [I] is among the most frequent tokens in the email corpus, it occurs both in 6th (*jag*) and 12th

⁴⁹ <http://www.ling.gu.se/projekt/tal/>

place (*Jag*). The equivalent in spoken language is found in place 4 (*ja{g}*). This suggests that email is less detached than traditional letters and written language (cf. Chafe and Danielewicz 1987). That the token *Göteborg* is frequent in the email data reflects the activity at the city council well. As the traditional letters were addressed to the same receiver, one would have expected the same results in that corpus.

Table 15. Word frequency in email and traditional paper letters, compared to spoken and written language.

	<i>Speech</i>	<i>Email</i>	<i>Traditional letters</i>	<i>Writing</i>
1	de(t)	.	.	.
2	e ₀	i	att	,
3	å ₀	på	och	och
4	ja{g}	att	i	i
5	att	och	,	att
6	så	jag	som	det
7	ja	är	för	en
8	som	som	en	som
9	vi	till	till	på
10	inte	det	är	är
11	på	för	av	med
12	man	Jag	på	för
13	i	om	det	av
14	då	en	har	jag
15	du	har	med	den
16	en	'	inte	han
17	ju	Göteborg	de	inte
18	dom	med	om	till
19	men	Hej	ett	--
20	den	av	den	Var
	<i>Speech</i>		<i>Writing</i>	

4.6.4 Mean length of messages

When comparing the mean length of messages, it was found that the mean length for email messages was 63.71 words, compared to 433.43 words for paper letters. Both modes of communication were asynchronous, but people probably spend more time on a paper letter than on email messages. The difference is most likely due to the variables means of expression and situation. Electronic text provides the sender with relative

anonymity. The messages were posted on the web page form while senders were searching the web page for info, they had perhaps not planned to write an epistle. If people are willing to go through the trouble of finding pen, paper, envelope, stamp and so on, they almost certainly have planned to communicate in another way than they would just surfing the net. Another factor connected with the variables of means of expression and situation is the type of Internet connection the senders were using at the time. Surfing the net on a slow modem while the phone bill is ticking is different than being hooked up using a broadband connection.

4.7 Conclusions

The aim in this study is to analyze electronically transmitted written messages from citizens to the city council of Göteborg, Sweden, and thus try to establish which factors related to the variables synchronicity, situation, and means of expression influence people's electronic message composition to unknown authorities and how these affect communication. It was hypothesized that situational factors such as the purpose for communication, topic and medium for communication play a part in the way messages are formulated. Other situational factors were the relationship between sender and addressee. Factors related to means of expression were the asynchronous mode of communication, and lack of technical feedback on grounding and closure on the actions, as well as the ease of access for sending messages and the user's relative anonymity. All these were hypothesized to exert influence the way in which electronic communication is formulated.

Results confirm suggestions from previous studies as well as hypotheses in the present study: norms for email are still in the process of being established. People are uncertain as to what conventions to use and what the effects of their messages will be. People juggle with both written and spoken conventions when formulating electronic messages. Are emails to be like speech, formulated according to telephone behavior (but without visual and vocal cues) or according to written norms of traditional letters? The styles of the messages in this study range from formal, letter like messages, to informal messages with all the features to be expected of email messages, a fact which supports the findings of Danet (2001). The email messages studied here are clearly shorter than the traditional paper letters in concurrence with evidence from Severinson-Eklundh (1994).

Just about 48% of all the messages in this analysis kept to the traditional Swedish way of writing informal letters; they were introduced with some sort of salutation and concluded with some kind of closing convention. Only 5% of the messages were introduced with a salutation of some sort and concluded with no closing convention at all. 25% of the messages were not introduced with any kind of salutation convention but concluded with closing conventions of some sort. 22% of the messages were neither introduced with a salutation nor concluded with any kind of closing convention.

Herring (1996) and Du Bartell (1995) explain the relative lack of epistolary conventions found in the messages of their analyses as being due in part to the architecture of email programs. Email technology in itself is easily accessible; typing messages and transmission occur rapidly. Sending email is also low in both cost and effort. The increasing accessibility of email seems to make people less wary of having their say in writing; the rapidity and ease of sending e-mails may influence the way people write, and what they write about.

Email is still a written mode of communication, and allows the sender to remain relatively anonymous, while struggling with the monomodality of written language.

Danet's (2001) suggestion, which is of a more social/psychological nature, is that the relationship between the sender and the receiver is a factor that might influence how people compose their messages. First letters to an authority should conform to the formal norm of letter writing. This proposition is perhaps not confirmed in this study; perhaps due in part to cultural differences in letter writing. The senders' ages were not available for reference in this study, and gender was difficult to establish in some cases. Nevertheless, the age and gender of sender affect writing style.

The closure time (Clark 1996) of email is uncertain. The sender does not know in advance whether she or he will receive a reply, nor is it clear in this study whether or not the answers were sent at all. A couple of messages were sent twice, indicating an uncertainty of a closure on the action.

The choice of style may not always be apparent to the writer. One of the reasons seems to be the purpose for the communication. If the purpose is to obtain help or information, there is a tendency to mind one's language and send "correct" messages. If the purpose is to complain or remark, there seems to be a tendency towards being less polite and formal as evidenced in lack of salutations or closing conventions. Anecdotal evidence from the staff at the city council suggests that email often replaces

telephone calls. They also suggested that the types of messages they get through email are a different sort than traditional paper letters. It seems that people feel they can hide behind the text and computer screens, remaining relatively anonymous. The ease and rapidity of sending email causes people to send messages which they probably would not have bothered to send had they been forced to find pen, paper, envelope, stamps, post box, and so on and so forth. Moreover, when people sent the email messages used in this particular study, they were already online surfing the net. Below is an example that illustrates this beautifully, and which must not be withheld from the reader. This particular letter would most probably not have been sent as a traditional paper letter:

Example 7.

Email message

Från: Stocckholmarn

Ämne: Hata lantisar

Translation into English

From: the Stocckholmer

*Subject: Hate country
bumpkins*

Ni götborgare är jävla
lantisar. Ni kan sluta va så
jävla kaxiga. Ni är en liten
skitstad mot Stockholm.
Älska Stockholm

*You gothenburgers are
real country bumpkins.
You better stop being so
damn cocky. You are a
hole compared to
Stockholm.
Love Stockholm*

This could well be seen as evidence that the accessibility and least possible effort it takes to contact someone in writing produce messages of a different type than the norm for traditional letters.

4.8 Chapter Summary

This purpose of this study is to analyze email messages and traditional paper letters that were sent from citizens to authorities at the city council of Göteborg in Sweden. The investigation attempts to establish which factors, in relation to the variables synchronicity, situation, and means of expression, influence language use and message composition in electronic communication with unknown authorities. Material was collected and analyzed both automatically and manually. Situational parameters such as

gender of sender, aim or topic of message were investigated. Comparison of salutation and closing conventions formed the basis for analysis of whether messages conformed to the formal business template or not. In order to find out how activity influenced language use and choice of words, word frequency email messages and paper letters were analyzed and compared to spoken and written language.

Results show that email messages show a wider variety of styles than paper letters. They were also found to be 6.5 times shorter than paper letters; mean length of message was 63.71 words for email and 433.43 words for paper letters. This difference was judged to be due to the variables of means of expression and situation, rather than synchronicity as both modes of communication were asynchronous. The paper letters were found to be more formal both with respect to opening and closing conventions, and regarding use of abbreviations and graphical means like emoticons.

5 Written Conversation in a Web chat

5.1 Introduction

This sub-study is the second among those that comprise this dissertation. It is an investigation and analysis of Swedish chat participants' use and adaptation of written language to the conditions of synchronous real-time chat conversation. A questionnaire was sent to students from 14 to 19 years of age in a number of Swedish schools, inquiring into the habits and preferences of their Internet communication. A chat room, chosen on the basis of the results of the questionnaire, was logged for approximately 120 hours. The material was stored and digitally analyzed. Results show that Swedish teenagers not only copy the strategies frequently used for verbal conversation in international chat environments; they also adapt the Swedish written language to meet the limitations and advantages of online communication. Strategies such as the use of emoticons and repetitive keys were found to be similar to those used in international chat environments. Abbreviations and acronyms based on the Swedish language were used, in addition to the ones copied from international chat environments. Swedish teenagers who use writing as a conversational tool in Swedish chat rooms write in Swedish for the most part, but they also use English under certain conditions. There is a blend of languages where English phrases turn up among Swedish utterances as fixed and common expressions.

Popular opinion assumes that computer-mediated communication will have, or is already having, supposedly negative effects on the language that we use in our daily life (cf. Crystal 2001; Kasesniemi and Rautiainen 2002,

among many others). Communication via e-mail and conversation in electronic chat environments of various kinds are becoming increasingly popular activities, especially among teenagers. People who criticize CMC seem to believe that popular activities like those will inevitably lead to harmful and negative effects on language. Language is supposedly in the process of being turned into something sloppy, less expressive and generally worse than it used to be in the good old days. Young people, especially, do not know how to express themselves or how to use “correct” language, or - perhaps even worse - they do not know correct grammar or how to spell! History has shown that new technological developments, such as the introduction of the telephone or the television, for example, often have been met with this sort of skepticism (cf. Baron 1984; 2000). The attitudes towards language use and behavior of younger generations are often just as skeptical.

Written conversation in electronic chat environments is an apparently important activity, which is part of the lives of an increasing number of teenagers. Teenagers might perhaps be the stereotypical chat addicts, but people who are out of their teens frequent chat rooms. This is an analysis of how chat participants use and adapt their written language to suit the conditions of written online conversation. As has been discussed above, written conversation suffers from the time consuming labor of typing and it misses out on many of the extralinguistic qualities of spoken face-to-face interaction. Furthermore, it cannot altogether satisfactorily replace spoken language; it is used for other purposes. The adaptation of language to suit the conditions of communication in which one is involved is interesting in its creativity and spontaneity. Electronic, real-time chat conversation has also been claimed to be a virtual playground for testing identities and social roles (cf. Danet, Ruedenberg-Wright et al. 1997, for example).

The main questions to be investigated in the present study concern which strategies are developed and used by Swedish chat participants, in order to adapt Swedish written language to the conditions of real-time written conversation. Are there any adaptations that can be viewed as specifically Swedish, compared to findings in chat rooms with international⁵⁰ participants? Are all strategies adjusted to the Swedish language, or are some just adopted unaltered from international chat rooms?

⁵⁰ “International chat environments” refers to chat channels which people from any country or linguistic background may participate in, not ones particularly set up for Swedish participants. The language used in “international chat environments” is mainly English.

A brief repetition of the short presentation of what real-time chat is, and its conditions provided in the background chapter (see Section 2.5.5 above), will be given below, together with findings from previous studies of the linguistic characteristics of chat.

5.2 Real-time Chat

5.2.1 Real-time Chat

Chat is a synchronous mode of CMC. Synchronous CMC requires its interlocutors to be online simultaneously. A chat room is a web site that provides a venue for communities of users to communicate in real time. All participants in an interaction must be electronically present at the same time, and messages are immediately transmitted through the intermediate servers to all participants, wherever they may be (Paolillo 1999). Forums and discussion groups, in comparison, allow users to post messages but do not have the capacity for interactive messaging (Herring Forthcoming). There are several forms of real-time chat that have slightly different conditions regarding, for example, text-only vs. graphic means such as capacity for digital images to be sent, or if messages are sent in their entirety vs. if other participants may see messages character-by-character. Examples of chat systems are IRC (Internet Relay Chat) (cf. Werry 1996; Paolillo 1999), text-based virtual realities such as MUDs (Multi-User Dungeons, or Domains (cf. Cherny 1999), web chat, and ICQ chat. Most chat rooms do not require users to have any special software; those that do, such as IRC allow users to download it from the Internet. In a web chat, the users log on to chat room of their choice, choose a user name (a nickname) and password, and log on to a particular room (most sites have multiple chat rooms).

In a real-time chat, a large number of people may participate in written conversation. Similar to a cocktail party, one may overhear, or rather “oversee”, other conversations going on in the chat room. As in a real-world situation it is difficult to take part actively in more than one or two conversations simultaneously. Inside the chat room generally there is a list of the people currently online; users are alerted each time another person enters the chat room. To chat, users type their contributions in a text box, hit the enter key or click a send button. The message is then displayed in the chat window to everybody logged in at that time. Bechar-Israeli (1995)

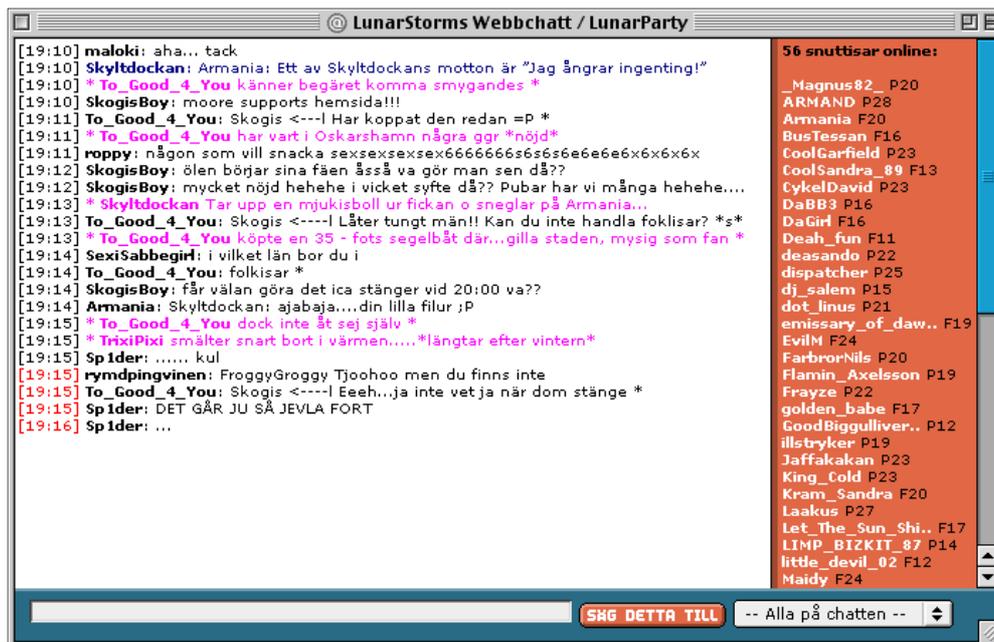


Figure 11. Example of synchronous CMC: a web chat window showing multiple participants engaged in interactive written discourse. The participants type their messages in the textbox in the lower part of this window.

argues that since a person's physical existence and identity must be condensed textually into a single line which states one's nickname and electronic address, the person will attempt to make these representational elements as prominent as possible. The way to do so is to choose an original nick which conveys something about the person's "self" and which will tempt other participants to strike up a conversation with that person. Bechar-Israeli points out the extent to which nicks have become an important part of the electronic self: they are experienced as an extension of the self. Bechar-Israeli further argues that the emergent IRC culture is a culture of linguistic virtuosity on the one hand, and of contempt for the rules of the language on the other. It is a culture that provides freedom in abundance to engage in identity games through the use of nicknames.

The particular chat system that was investigated in the present study was a web chat. Figure 11 above illustrates an example of a web chat window. A web chat is a quasi-synchronous system (cf. Garcia and Jacobs 1999), which requires that both sender and addressee(s) must be logged on simultaneously. The messages that are posted in the chat window are ephemeral, scrolling up and off participants' computer screens as new messages replace them. Messages are posted and read by multiple participants (many-to-many). It has a high degree of anonymity, is typically social in function and informal, or even non-serious in tone. In a way it is

the ideal arena for play with identities, social roles and language: it contains a high incidence of flirting and phatic (empty, social) exchanges. Studies have shown that participants are most often persons between the ages of 18-25 (Reid 1991; Werry 1996; Danet, Ruedenberg-Wright et al. 1997).

5.2.2 Linguistic Characteristics of Chat

Danet et al. (Danet, Ruedenberg-Wright et al. 1997) argue that linguistic features previously associated with oral communication are strikingly in evidence in real-time chat. They further point out that CMC in general is remarkably playful.

Storrer (Storrer 2001) used data samples from different types of chat services to discuss media-specific forms of turn-taking and the use of deictic expressions in chat communication. She claims that use of written language as well as the specific technical setting affects the strategies for language processing during chat communication. As a consequence, written chat conversations differ considerably from their spoken counterparts. Storrer discusses and explains some linguistic peculiarities in (Herring Forthcoming) dialogues which occurred in chat conversation, and focuses on two phenomena that she argues reveal significant differences between spoken and written dialogues: organization of turn-taking (also discussed by Herring 1999), and usage of deictic and local expressions (e.g., *here, I, me, above*).

Schönfeldt (2001) examined the question of how “typed conversations” in chats with an unspecified topic, as well as being un-moderated, can be compared to face-to-face conversations in regard to their organizational structures. The categories developed within the theoretical framework of traditional discourse analysis may be applied in describing the interaction in chats. In applying analytic units used in conversation analysis on a web chat she shows that the organizational structures in chat correspond to patterns of organization in oral conversations (cf. Werry 1996). Schönfeldt concludes that chat therefore can be analyzed as a new form of conversation.

Werry (1996) notes that electronic written online interaction makes for less time delay for feedback, compared with traditional written communication. As the participants communicate in real time, they are able to negotiate meaning in another fashion than is possible in traditional letter writing, for example. Werry (1996) notes, though, that the receiver is usually unable to supply the minimal responses (nonverbal forms such as nodding, gaze, and

verbal forms, such as ‘mm hm’). The fact that the participants in chat conversations can hide behind the text, as it were, enables them to change and play with identities – age, sex, background, and so on, if they like (cf. Danet 1998). Yates (1996) argues that CMC as a mode of communication is very complex. As with both written and spoken discourse, it is affected by social structural and social situational factors that surround and define the communication taking place. Further, he states that:

What is yet to be made fully clear is the extent to which human beings in specific social and cultural settings can develop and enhance their communication through the use of CMC.

However, just as CMC cannot be regarded as a homogenous mode of communication, different chat systems and chat forums must to be viewed in their own rights. Herring (Forthcoming) agrees with Cherny’s (Cherny 1999) conclusion that the norms for discourse in a social MUD are not the same as those for Internet Relay Chat, despite the fact that both are synchronous chat environments and make available similar communication commands. Paolillo (1999) argues that, since IRC messages are typed at a keyboard, there is a tendency to use conventions of written English, particularly spelling. Paolillo found a number of distinctive IRC spelling practices that have emerged. The practices of substituting the letters u and r for the English words you and are, and substituting z for s, especially in word-final position, are three such IRC spellings. Paolillo contends that all three spellings diverge from standard written English, and so would be considered “vernacularizing” changes.

5.2.3 Constraints on Written Communication

As discussed above, traditional written language provides fewer clues to guide message interpretation, compared with face-to-face interaction. Relying on the single modality of vision, written conversation has to be more explicit than spoken face-to-face conversation in order not to be too ambiguous. Baym (1996) argues that:

In written media...most of the nonverbal cues of oral interactions are not available. There can be no body movements, vocal tone, rate, or volume... [T]he single greatest difference between writing and oral interaction, especially face-to-face conversation, is that participants do not share a common physical or temporal context. As a result, writing is generally more explicit than oral language, as it must make manifest all that would be apparent from the context.

In spoken language, nonverbal indications are simultaneously given with the words uttered. Face-to-face interaction employs many modalities at the same time to signal information, while written communication has to rely on the single modality of vision. Consequently, one of the problems of writing is the fact that it not only takes a good deal longer to type what you would like to express than it takes to utter it in speech, one also misses out on the possibility to convey (consciously or subconsciously, voluntarily or involuntarily) simultaneous clues of being ironical, nervous, angry and so on. Written communication also typically lacks information regarding the sender's age, sex, social background, all of which are provided by vocal and visual cues in face-to-face interaction. In traditional written language, strategies which attempt to minimize these difficulties have already existed for a long time; we use capitals to indicate stress or give emphasis to a word, abbreviations to save effort, punctuation marks to make reading easier and convey the sender's intention, and so on. When new demands are placed on written language by online communication, the need to be able to convey more information than what can be done by means of traditional written language emerges. Consequently, strategies to support this need develop. The limitations of written interaction can be characterized chiefly by the lack of immediate feedback from the other participants (cf. Herring 1999), lack of simultaneous non-verbal cues, and the effort it takes to write or type one's contribution instead of uttering it. These problems may lead to ambiguities and misunderstandings in text-only environments. The mode of CMC, which the present analysis deals with, is text-based, and the participants read each other's messages rather than listen to them.

5.2.4 Advantages of written communication

Written contributions are not just constrained by the features described above, there are also advantages to this mode of communication. First, written contributions can be preserved in a fashion not normally possible in spoken interaction. In chat communication, participants can relate to

things contributed earlier in a way that is different from spoken language: scrolling the screen to read and reread the messages previously transmitted. The potentiality to read and reread written conversation provide participants with the possibility to engage in multiple conversations simultaneously. Herring (1999), following Walther (1996), calls this “hyper personal interaction”. She sees this as a distinct advantage of the medium.

CMC thus enables a greater intensity of interaction – that is, more interactions concentrated within a single temporal frame - than is possible face-to-face. To appreciate this, one has only to compare a typical IRC session with a face-to-face social event, such as a party, to realize the limits on multiple simultaneous interaction in the latter.

However, it might be questionable to label this type of interaction "intense interaction", as it is not clear how more interaction makes interaction more intense. Still, the potential to be able to relate to contributions made earlier in the conversation(s) makes it plausible. Keeping track of all the different conversations going on at the same time might be confusing to the novice. Werry (1996) notes that

It is no doubt only because utterances have a longer half-life on the screen than they do in oral discourse that the users are able to keep track of the separate IRC conversations going on around them at all.

Electronic written online interaction makes for less time delay for feedback, compared with traditional written communication. Since participants are communicating in real time, they are able to negotiate meaning in another fashion than is possible in traditional letter writing, for example. Werry (1996) points out that the receiver is usually unable to supply the minimal responses (nonverbal forms such as nodding, gaze, and verbal forms, such as ‘mm hm’ and so on), though. The fact that the participants in chat conversations can hide behind the text enables them to change identity, as it were – age, gender, and background - if they feel like it. Witmer & Katzman (1997) point out that:

A key characteristic of virtual reality [is that] CMC can mask personal characteristics and identities of cyberspace travelers to create personal anonymity in a public arena.

The issue of relative anonymity in CMC will be revisited below.

5.2.5 Strategies to adapt to text-only limitations

Previous studies⁵¹ have shown that strategies have been developed to overcome the limitations of text-only communication, some of which are probably rather well known by now. Cyber communicators use emoticons to convey non-verbal signals. Emoticons can be described as keyboard characters combined to resemble facial expressions to indicate the mood in which something was said or how something is to be interpreted. Words or phrases framed in asterisks have the same function. Various kinds of abbreviations are used extensively to save time and effort. All-capitals are used to indicate prosodic features. A convention of indicating the intended addressee has emerged in chat rooms, in order to avoid misunderstandings. Extreme and unconventional use of repetitive keys or punctuation marks of all sorts, has also been found to be used extensively in chat (cf. Witmer and Katzman 1997). It seems likely that these strategies have other functions than just saving time and effort; this issue will be revisited in the present study below.

5.3 Material & Method

5.3.1 Material

The material for this study consists of data from two kinds of sources. The first is a questionnaire inquiring into students' habits and preferences of communicating on the Internet. The second source is linguistic material that was logged from one of the chat rooms that the results from the questionnaire indicated as the most popular at the time of the inquiry (May 2000). Before logging material, permission to do so was obtained from the persons responsible for the maintenance of the chat room.

All in all, 333 students (164 girls and 169 boys) answered a questionnaire that was sent to five upper secondary schools⁵² (students aged 16-18), and two lower secondary schools⁵³ (students aged 13-15). The questionnaire inquired into the teenagers' habits and preferences of electronic communication. Questions concerning their use of email and chat

⁵¹ Cf. Werry (1996).

⁵² "Gymnasiet" in Swedish.

⁵³ "Högstadiet" in Swedish.

communication were included. Most of the questions only required choosing one from a list of short answers: *'yes'*, *'no'*, *'daily'*, *'once a week'*, etc. Only a limited number of items required the informants to provide written justification for answers. The students were asked whether they have access to Internet and how often they use it, whether they frequent chat rooms, whether they have their own e-mail address and how often they send e-mail messages. Of specific interest for this study was the question regarding which languages the informants claimed to use in everyday communication and which ones they claimed to use in their chat communication. Because one of the main aims of this study is to analyze how teenagers adapt their written language to online communication, they were asked whether they employ strategies such as emoticons and abbreviations in their chat communication. They were also asked to give examples of these strategies from their own communication. One question concerned which chat sites they considered to be the most popular ones, and arguments for why they felt this way. The choice of which chat site to use to log material for this study was based on these results.

The students were all informed that the survey, conducted at Göteborg University, aimed at finding out in what ways and how often they communicate electronically and what languages they use – and how they feel about it. The students were also informed that the questionnaire was to be treated absolutely anonymously and that participation was voluntary.

One of the most popular Swedish chat rooms at the time was found to be a web chat maintained by a Swedish newspaper⁵⁴. It was logged for approximately 120 hours (see table 16 below), both in order to get a large corpus of material and in order to view the spread of activity during the same period. The material was logged in 1999, between Wednesday 25 August at 2.07 p.m., and Monday 30 August, at 1.46 p.m. A robot program, or a 'bot'⁵⁵, was developed for collecting electronic data. The bot logged on to the chat site acting as a human participant and recorded both the linguistic contributions in the chat room, as well as data such as sender (participant/user) and posting time of each contribution in the chat room in machine-readable format.

During this period, 4,293 chat participants ("unique pseudonyms") from 278 different domains posted 44,380 contributions, which amounts to 410,355 tokens⁵⁶ totally. Quantitative analysis of the presence of strategies

⁵⁴ <http://nychat.aftonbladet.se/webchat/oppkanal/Entren.html>

⁵⁵ Leif Grönqvist wrote the 'bot' program.

⁵⁶ Somewhat simplified, a token is a sequence of letters surrounded by spaces, which the automated search tool TraSA is able to find. Emoticons are treated as tokens by the tool and are also included

outlined above (Section 5.2.5), as well as the occurrence of the examples that were given in the answers to the questionnaire, was made using the automated search tool, TraSA. Quantitative comparisons with Swedish spoken language and written language corpora⁵⁷ were made in order to relate chat conversation to spoken and written modes of communication.

Table 16. Data from the web chat corpus.

	<i>Web chat</i>
<i>No. messages</i>	44,380
<i>No. words</i>	410,355
<i>Hours logged</i>	119 hours, 39 minutes and 15 seconds

Table 17 below illustrates a sample of the log. The first column represents time of posting, the second represents the participant (nickname), in some cases the intended receiver was indicated with an addressivity marker (Pär1a>>>>Anonym79), where Anonym79 is the receiver). The third column represents the participant's message. The logged material included information about the senders' domain, but that was omitted for the sake of further anonymity for the participants.

Table 17. Sample of the log.

<i>Time of posting</i>	<i>Nickname of participant</i>	<i>Message</i>	<i>Translation into English</i>
14:07:02	Felicia, 29	<i>*trist*....</i>	<i>*boring*....</i>
14:07:07	Max	<i>hur är det JEN</i>	<i>how's things JEN</i>
14:07:19	mac	<i>nina : hur gammal är du</i>	<i>nina: how old are you</i>
14:07:23	Pär1a>>>>Anonym79	<i>Har du några egna hästar eller ??</i>	<i>Do you have any horses of your own or what??</i>
14:07:27	isola	<i>icq kanske</i>	<i>icq perhaps</i>
14:07:41	Anonym79>>>Tobbe.s	<i>var tog du vägen .</i>	<i>where did you go.</i>
14:07:41	JEN	<i>Hej MAX, VAR KOMMER DU IFRÅN ??</i>	<i>Hi MAX, WHERE ARE YOU FROM ??</i>
14:07:41	Sebbe>>Anonym 79	<i>Hej vem är du ?</i>	<i>Hi who are you ?</i>
14:07:43	Rocko>>>>>>josse	<i>PUNKAR SVIN</i>	<i>PUNK SWINE</i>
14:07:50	Fresh>>>Pär1a	<i>*ler* ...alright .. ja fattar nu ... *ler*</i>	<i>*smiles* ...alright .. I understand now ... *smiles*</i>

⁵⁷ The spoken language corpus (Kernel Corpus - adult 1st language Swedish) and written language corpus at the Department of Linguistics, Goteborg University, Sweden, consisting of approximately 1.2 million words each.

< <http://www.ling.gu.se/SLSA/gbgcorpora.html> >

5.3.2 Method

The chat material was stored and analyzed digitally, by means of the automated tool, TraSA, described in the chapter on Material & Methods above⁵⁸. The software allows for quantitative analyses of, for example, word frequency; number of utterances per participant; the occurrence of abbreviations and punctuation marks; and mean length of utterance. This study required the analysis of occurrences of emoticons, abbreviations of various kinds, use of ‘all-capitals’, extreme use of repetitive keys (punctuation marks), and asterisks.

A qualitative analysis had to be conducted manually, analyzing the occurrence and types of abbreviations, and emoticons. A comparison of the most frequent words in the chat material with the most frequent words in spoken and written Swedish (cf. Section 5.4 below) was then made.

5.4 Results

Results from the questionnaire that directly concern the teenagers’ adaptation of their written language to online communication dealt primarily with the language(s) that the informants claimed to use in everyday situations, as well as the language(s) they claimed to use in chat conversations. Furthermore, attention was paid to questions inquiring into whether they use strategies such as emoticons and abbreviations including the examples they gave. Other questions gave information of a more indirect nature, which is just as interesting but perhaps not as relevant in this analysis. The question of which were the most popular chat channels was naturally relevant for the choice of which chat room to log.

Investigating the differences between answers given by girls and by boys would also prove interesting in the long run, but was not germane to the present analysis.

The present study’s aim is analyzing the language employed by Swedish chat participants in a Swedish chat room. One should bear in mind that verifying informants’ age was impossible; the only information about chat participants’ age that could be reliable came from the questionnaire. Not only is age of participants in a chat room impossible to know with total certainty, gender of participants is also difficult to determine as text-only

⁵⁸ TraSA was developed by Leif Grönqvist, M. Sc in Computing Science. Some information can be found at: <<http://www.ling.gu.se/SDS/multitool/related.htm>>, or at Leif Grönqvist’s homepage: <<http://www.ling.gu.se/~leifg>>

conversation allows the senders to take on new identities. As suggested by Witmer & Katzman (1997):

Impersonating a member of the opposite sex is a fairly common practice in the world of CMC.

Herring (2000) found that to be less common than claimed. If the information concerning participants' age and gender that was sometimes found in the chat material is to be trusted, the nicknames suggest ages somewhere between 13 and 30 during the particular week that was logged.

Results from quantitative analyses of both the questionnaire and the logged material will be presented below. Although the questionnaire did not specifically inquire about anything other than emoticons and abbreviations, the analysis of language used and the analysis of the following strategies will be given: emoticons, asterisks, 'all-capitals', extensive use of punctuation marks (or repetitive keys), and abbreviations.

A comparison of the most frequent words in the web chat to spoken and written language follows the explanation of the table below. The most frequent spoken words are rendered in their written equivalent, supplemented by an index number or letters in brackets for disambiguation. The second most frequent spoken word, see Table 18 below, [e₀] stands for [är=is, are] and disambiguates it from [e₁] which indicates a hesitation sound. In the same way, [å₀] in the table below stands for [och=and], which disambiguates it from [å₁], indicating [att=to]. The word [ja{g}=I] is disambiguated from [ja=yes].

Interestingly, the most frequent token in written conversation in web chat is the word *du* [you]. This fact reflects the activity in which the communication takes place, which in the case of web chat is the situational influence of social-recreational chatting. It also shows the importance of topic, or perhaps goal of interaction, in the situation: the words *du* [you] and *jag* [I] are what people talk about.

Table 18. Comparison of the most frequent representations in Swedish spoken language, written language and chat communication. Approximate English glosses are given.

<i>Frequency order</i>	<i>Speech</i>	<i>Translation</i>	<i>Web chat</i>	<i>Translation</i>	<i>Writing</i>	<i>Translation</i>
1	de(t)	it	du	you	.	.
2	e ₀	am/are/is	?	?	,	,
3	å ₀	and	jag	I	och	and
4	ja{g}	I	,	,	i	in
5	att	that, to	är	am/are/is	att	that, to
6	så	so	.	.	det	it
7	ja	yes	det	it.	en	indefinite article
8	som	that/which/who	som	that, which
9	vi	we	som	that/which/who	på	on
10	inte	not	!	!	är	am/are/is

As the comparison in Table 18 above shows, chat conversation seem to be about relationships; the “primary discourse situation” (1st and 2nd person singular); getting to know other participants and expressing yourself. It is about you and me. The frequency order of Swedish chat conversation naturally also includes graphic representations like the frequency order of Swedish written language.

The most frequent pair (combinations of two tokens) in the web chat material was som vill [who wants]. This, and variants of it, is a very common chat opening: Is there anyone who wants to chat with me? The examples below, taken from the chat material, illustrate the frequent opening phrase.

Table 19. Comparison of the most frequent pairs in Swedish spoken language, written language and chat communication. Approximate English glosses are given.

<i>Speech</i>			<i>Web chat</i>		<i>Writing</i>	
<i>Pairs</i>	<i>Translation</i>		<i>Pairs</i>	<i>Translation</i>	<i>Pairs</i>	<i>Translation</i>
de(t) e ₀	it is		som vill	who wants	det är	it is
e ₀ de(t)	is it		är du	are you	för att	because
att de(t)	that it		? sanna	? sanna (nickname)	det var	it was
men	but it				att det	that it
de(t)			jag är	I am		
e ₀ ju	why ⁵⁹				är det	is it
	(interjection)		du ?	you ?		
så att	so that		, hallå	, hello	i en	in a
de(t)	it was				att han	that he
va(r)			det är	it is		
å ₀ så	and then		sanna,	sanna, (nickname)	i den	in the
de(t)	this				som en	as a
här			hallå ?	hello ?		
å ₀ de(t)	and it		har du	have you	men det	but it

Example 8. HALLÅ någon som vill...

[HELLO anyone who wants to...]

Example 9. Hej! är det någon som vill chatta med mig?

[Hi! is there anyone who wants to chat with me?]

Example 10. Någon som har lust att chatta en stund??

[Anyone who feels like chatting for a while??]

Example 11. Är det någon som vill.

[Is there anyone who wants to.]

The other common pairs suggest a reflection of the activity in a web chat: it is social-recreational and basically concerns relation and interplay with others and the expression of the self.

⁵⁹ Confirmational interjection used to indicate surprise or indignation placed first in English phrases, example: 'Why, you can play the piano!' ('Du kan ju spela piano!').

Frequenting chat rooms

Results from the questionnaire show that most of the informants that responded to this item claim that they frequent chat sites a couple of times a week. There seems to be a slight difference between girls and boys:

Table 20. Results from the questionnaire: “How often do you frequent chat rooms?”

	<i>Girls</i> (% of all girls)	<i>Boys</i> (% of all boys)	<i>Total sum</i> (% of both girls and boys)
<i>No answer</i>	60.9%	55.0%	57.9%
<i>Every day</i>	0.6%	7.6%	4.2%
<i>A couple of times a week</i>	28.0%	34.3%	31.2%
<i>Never</i>	0%	2.3%	1.2%
<i>A couple of times a month</i>	7.3%	0%	3.6%
<i>Every 3rd month</i>	3.0%	0.5%	1.8%

Answering the question whether they find chatting important (not represented in a table), 53% answered *No*, and only 7.8% answered *Yes*. Their explanation as to why this is so was frequently given as “better to meet people in the real world”, or “you lose your REAL friends”. Still, although as many as 60.9% chose not to answer this question⁶⁰, we can see in the table above that as many as 31.2% visit chat rooms a couple of times a week. Of the informants in this analysis, boys seem to frequent chat rooms on a daily basis more often than girls. 2.3% of the boys claimed never to chat at all.

Languages used

The majority of the informants, not unexpectedly, claim to use Swedish in everyday situations, and Swedish and English in chat conversations. Remaining percentages are combinations of other minority languages in Sweden, such as Swedish + Arabic, Swedish + Greek.

⁶⁰ One can only speculate about what the possible reasons for not answering the question of whether they considered chatting to be important or not. Perhaps some informants simply ignored it, or they felt they had no opinion in the matter.

Table 21. Results from the questionnaire: Languages claimed to be used in everyday situations.

	<i>Girls (164)</i>	<i>Boys (169)</i>	<i>All (333)</i>
<i>Swedish only</i>	65.2%	40.2%	52.5%
<i>Swedish + English</i>	15.2%	10.0%	12.6%
<i>Swedish + English + German</i>	0.2%	0.3%	2.7%
<i>Swedish + English + French</i>	1.8%	0%	0.09%

52.5% of the informants claim to use only Swedish in everyday situations. It is possible that some of the informants took the question of language use in ‘everyday situations’ to mean even languages that they are taught in school, and thus included them in their answer. The question should perhaps better have been formulated like “Which languages do you speak with your family and friends” to captured the purpose better.

Table 22. Results from the questionnaire: Languages claimed to be used in chat room conversations

	<i>Girls (164)</i>	<i>Boys (169)</i>	<i>All (333)</i>
<i>Swedish only</i>	23.7%	15.9%	19.8%
<i>Swedish + English</i>	20.1%	15.9%	18.0%
<i>English only</i>	12.2%	2.9%	2.1%
<i>Swedish + English + German</i>	12.2%	1.2%	1.2%

A comparison of the answers to which languages the informants claim to use, show that only 19.8% use Swedish only and almost as many (18.0%) use the combination Swedish + English in chat conversation. The language used in the chat room data in this study was mainly Swedish. Apart from loan words (in some cases with the English spelling intact, in other cases adapted to Swedish spelling), English phrases show up here and there, sometimes in the middle of a Swedish sentence. Note that the examples below are taken out of their original context, and they did not appear in succession in the chat data.

Table 23. Examples of messages in a web chat written in a mixture of Swedish and English.

<i>Time of posting</i>	<i>Participant's nickname</i>	<i>Message</i>	<i>Translation into English</i>
01.07.20	Darth Olsson	Helloo allibadi hur e de i dag?	<i>Helloo everybody how's things today?</i>
14:44:40	G.B	Critical information check	<i>Critical information check</i>
01.11.40	Little Boy Lost	fru hjärterdam...120 mil busstripp...Låter höjdare om det...;)	<i>mrs queen of hearts... 120 mile bus trip...Doesn't sound too good...;)</i>
18.10.30	PeeWee	this sucks	<i>this sucks</i>
22.17.12	Ellen (16)	whatever!	<i>whatever!</i>
16.06.55	Blackboy	Whats up	<i>Whats up</i>

The above examples show that both nicknames and contributions might consist of either a mix of Swedish and English, or of pure English. The spelling is rendered exactly as it was in the chat material.

5.4.1 Strategies

Emoticons

Of the informants answering the questionnaire, only about 16.2% claimed to use emoticons often in their chat conversation. 11.7% claimed to use it sparingly, and 33% claimed to never use emoticons at all. 39% did not answer the question, whether this is due to the fact that they previously answered No to the question if they frequent chat rooms at all, or because they ignored the question is not known.

The emoticons that were found to be most frequently occurring in the web chat corpus were also given as examples in the questionnaire material. Emoticons that were actually and frequently used were found to be the simplest ones, i.e. the ones that are likely to be understood by most users and which can also be typed quickly. The most common emoticon in the chat material is the simple eyes and smiling mouth [:)] of a happy face. It is easy to type quickly and presents no problem of interpretation.

Table 24. Emoticons given as examples in the questionnaire and found in the chat material

<i>Examples from the questionnaire</i>		<i>Findings in the chat material</i>			
<i>Tokens</i>		<i>No. of occurrences</i>	<i>Token</i>	<i>No. of occurrences</i>	<i>Token</i>
:)	8)	94	:)	1	===)
:-)	x)	84	:-)	1	=0)
=)	:-S	81	=)	1	=-)
:-(*U*	29	:-(1	;-*
;))	=o)	23	;))	1	: '(
@-}-	:-	15	~~~~{~~@	1	:(c)
:0	:/} {O}<	7	=(1	:=)
(*) (*)	:)=	7	:(1	:-))
o=3	=`(`	6	:))	1	:-)<
(*)	:-9	5	=)	1	:)))))
:(;-	4	:-)	1	--{-@
-(.)	>=)	3	;-)	1	,-)
:°)	=D	3	:0)	1	(:
:°D	=0)	3	(.)(.)	1	@->--.
:-O	=P	1	:(((1):)
:*	:/				
:~(0<:-)				
<#-)					

The elaborated emoticons given as examples in the questionnaire are hardly ever used in actual written conversation. The problem is that the use of intricate symbols demands that specific background knowledge that has to be shared by both sender and receiver in order to function as intended; to convey information simultaneously in a quick and easy manner, as well as saving time and effort. Consequently, the emoticons that are in fact most frequently used are the simplest ones, which are likely to be understood by most users. It is easier for the receiver of written messages to get the information, when general or more specific knowledge is shared (Levelt 1981). The examples below were taken from the web chat corpus, and illustrate how emoticons were used in web chat contributions.

Example 12. jo det har jag ;))
[yes I have ;)]

Example 13. Finns det någon som vill????? :-)
[Is there anyone who wants to????? :-)]

Example 14. `ahh...det säger du, berätta =)`
`[ahh...is that so, tell me =)]`

Asterisks

Another strategy to indicate an action, a mood or how things are to be taken, is to frame words or phrases in asterisks. These are mostly positioned at the end of a contribution. Several answers from the questionnaire suggest that the use of framing words indicates or expresses an emotion or feeling. The asterisk strategy suggests an attempt to signal non-verbal (facial or vocal) cues, which would inform the receiver about the sender's mood, etc. Like the abbreviations, words within asterisks might be in English just as well as Swedish. Words within underscores, `_laughter_`, have the same function⁶¹. These are 'poor man's emoticons' – i.e. when the user is unclear about how keyboard characters are combined to resemble facial expressions. It might also be quicker and easier to put a word in asterisks and, in that way, make the receiver know in what way the contribution should be taken. It could just as well be seen as yet another variant to convey extralinguistic information. It might be the case, though, that those who touch-type might have trouble getting used to typing that much punctuation at once, whereas those who look at the keyboard and use two fingers to type would not necessarily have any more problem with emoticons than real words.

If compared with the occurrence of emoticons (Table 24 above), it seems that the method to indicate action or emotions by means of asterisks is the more popular. It is much more frequent; there were 87 instances of the most frequent emoticon [:)], while the most frequent asterisk-framed action occurred 814 times [`*!er?`].

⁶¹ This habit seems to bear traces from “pre-windows” based formatting of computer typed texts.

Table 25. The most frequent asterisk framed words in the chat material.

<i>No. of occurrences</i>	<i>Token</i>	<i>Translation</i>
814	*ler*	smiles
102	*gone*	[Eng.]
80	*s*	s=smiles
60	*kram*	hug
50	*suck*	sigh
45	*skrattar*	laughs
43	*spanar*	searches for
38	*snyft*	sob
36	*l*	l=ler (smiles)
35	*skratt*	laughter
33	*puss*	kiss
27	*ensam*	alone, lonely
22	*nyfiken*	curious
22	*ler	smiles
18	*borta*	gone
16	*rodnar*	blushes

Examples that were taken from the web chat log below illustrate how asterisks were used in contributions in this particular chat.

Example 15. Hmmm.....synd!!.....hur gammal är du
då???.....snälla svara!!*ler*
[Hmmm.....pity!!.....how old are you then
???.....please answer!!*smiles*]

Example 16. Det låter ju underbart...har du någon
kille..?...man måste ju fråga...*nyfiken*
[That really sounds wonderful...do you have
a boyfriend..?...one has to ask...*curious*]

Example 17. whooop!!!! Sleep tight...*waves*

The proportion of tokens within asterisks amounts to roughly 1% of the total number of tokens in the chat material (total number of tokens 410,355).

Whole Messages Typed in Capitals

Typing words or phrases in capitals seems to be an attempt to convey prosodic features such as stress and emphasis on words. Typing which uses

nothing but capitals is equivalent to shouting and extensive use of all caps might be taken as rude behavior. The questionnaire did not inquire specifically into the use of capitals, so a more exhaustive analysis is needed to investigate the contexts in which these expressions written in capitals occurred.

Table 26. The most frequent words in ‘all-capitals’ in the chat material. Approximate English glosses are given.

<i>No. of occurrences</i>	<i>Token</i>	<i>Translation</i>
890	DU	YOU
763	JAG	ME, I
658	ÄR	AM/IS/ARE
647	HATA	HATE
554	VILL	WANT/S
530	I	IN
490	DET	IT/THAT
418	SOM	WHICH/WHO/THAT
283	PÅ	ON, AT
255	NÅN	SOMEONE*
249	INTE	NOT
227	MED	WITH
224	OCH	AND
217	NÅGON	SOMEONE
212	DÅ	THEN
204	HALLÅ	HELLO
203	OK	OK
192	EN	A/AN
182	HAR	HAS/HAVE

A comparison of the most frequent words of the chat material and the ones written in capitals only show that these items address the same topics; it is still about you and me. Interestingly enough, the word “HATE” does not occur in the frequency list of words written in small caps, which indicates that it has to do with shouting or being abusive and employs the strategy of capitals to make the message come through more clearly, as Example 18 below shows. Another illustration of a message typed in all capitals for emphasis or to indicate shouting is illustrated in Example 19.

Example 18. SKAFFA ER EN “RIKTIGT TJEJ”.. SÅ JÄVLA MKT
TREVIGARE JAG LOVAR ER... STICK JAG HATAR ER
[GET YOURSELVES A “REAL GIRLFRIEND”...SO DAMN
MUCH NICER I PROMISE YOU...PISS OFF I HATE
YOU]

Example 19. HAAAALLLLLLLÅÅÅÅ!!!!
[HEEEELLLLLLOOOOO!!!!]

The proportion of tokens typed in all-capitals in this particular chat material amount to roughly 5.6% of the total number of tokens.

Extensive use of punctuation marks

There was an extensive use of various kinds of punctuation marks in the data, which seem to be used to express attitude, ask questions or generally ‘make oneself heard’. All kinds of variations and combinations were found.

The use of punctuation marks conforming to conventional, traditional norms of written language was naturally frequent in order to express questions, exclamations and make statements. Use of a single question mark showed up 13,361 times, single comma 10,282 times, single period 8,854 times, ellipsis (three periods) indicating thoughtful pause 4,970 times, single exclamation mark 4,677 times. So far, nothing exceptional is noted. However, looking further down the frequency order, examples like the ones in Table 27 below appear.

Table 27. Examples of extensive use of punctuation marks

<i>No. of occurrences</i>	<i>Token</i>
558	>>>
510	!!!!
464	>>
411	?????
22	!!...
19	!!!!!!!!!!!!!!!!!!!!!!!!!!!!
9	??

In the chat data, >>> is often used to indicate the intended addressee (cf. Werry 1996). Example 20 below was taken from the web chat log, and illustrates this use. A plausible guess at the reason for repeating a character in extremis, in some cases up to as many as 315 question marks in a row, might be that the use stems from both the wish to make oneself heard (or seen, rather) in the crowd, and from the fact that it requires little more effort to type a hundred question marks than it does to type one. It is just a matter of pressing down the key a bit longer.

Example 20. >>>Armageddon var kommer du ifrån
[>>>Armageddon where are you from]

Example 21. Hallå!!!!!!!!!!!!!!! Var är alla pratglad
människor??????
[Hello!!!!!!!!!!!!!!! Where are all you
talkative people??????]

Abbreviations

The strategy that perhaps was most interesting to analyze in this study was the use of different types of abbreviations. Results from the questionnaire show that 10% of informants use abbreviations every time they chat, 20% indicated “sometimes”, and 25% indicated “never” (43.2% did not answer this question). Some of the answers suggest that people answering No to the question of frequenting chat rooms, answered that they never use abbreviations, simply because of the fact that they never chat at all.

As indicated in Table 28 below, some of the abbreviations are the same as those used in traditional writing. Others seem to have been developed specifically to suit the needs of online communication. Results from the questionnaire show that only a minority claimed to be using abbreviations frequently in their chat conversation. The examples they gave are based both on English and Swedish as used in traditional writing. Interestingly enough, Swedish teenagers do not just copy behavior from international chat rooms that they have visited: the influence of these make them create new abbreviations based on Swedish, as well as ones made up by analogy with what those based on English look like. A number of different types of abbreviations emerged, both in the examples from the questionnaire and in the chat material:

- Acronyms, abbreviations made up from the first letters in a phrase (e.g. *istf* = i stället för [in stead of]).
- Numbers representing the sound value of a syllable in combination with letters (e.g. *3vligt* = trevligt [‘nice’]).
- Letters representing the sound value of a syllable in combination with other letters forming an abbreviated representation of a word (e.g. *CS* = (vi) ses [‘see (you)’]).

These are creative and innovative adaptations of written language for online written conversation; moreover it is an adaptation to suit the needs of the activity of written online conversation. Creative unconventional

abbreviations, together with other strategies, also seem to function as a marker of the sender's ability to master cyber communication, as it were. It marks the sender's identity and seems to represent the sender's belonging in the cyber community. This aspect will be revisited in the discussion section.

Table 28. Examples of English abbreviations and the explanations given by the students in the questionnaire

<i>Abbreviations</i>	<i>Full phrase</i>	<i>Abbreviations</i>	<i>Full phrase</i>
L.O.L	laughing out loud	a'r	are
OK	cool	M	Male
ASL	Age sex land	Fem	Female
S	Smile	SM	secret message
How RU	How are you	ROFL	rolls around the floor laughing
Gr8	great	thnx	thanks
U	you	ppl	people
brb	be right back	wb	welcome back
STATS	all facts about a person	j/k	just kidding
wanna chat	do u want to chat	g2g	got to go
w8	wait	any1	anyone
cu	see you	no1	noone
l8r	later	u2	you too
cya	see you	OKI	OK
I luv ya	I love you	dta	data (?)
wana	want to	later	see you later
ya	you	yer	your

Table 28 above illustrates examples of English abbreviations and their interpretations, as provided by the informants in the questionnaire. Swedish users have most probably observed their use in chat rooms where the English language dominates, and then copied the use to Swedish chat sites.

Table 29 shows examples of abbreviations that were given in the questionnaire: both conventional abbreviations and unconventional abbreviations that were developed to suit the conditions of written online conversation.

Table 29. Examples given in the questionnaire of innovative and abbreviations established in Swedish standard written language.

<i>Innovative abbr.</i>	<i>Full phrase</i>	<i>Translation</i>	<i>Traditional abbr.</i>	<i>Full phrase</i>	<i>Translation</i>
asg	asgarvar	laughs hard	ngn	någon	someone
iofs	i och för sig	strictly speaking	Ngra	några	some ones
iaf, if	i allafall	anyway	gbg	Göteborg	Göteborg
é	är	is	sv	svenska	Swedish
d	det	it	bla	bland annat	among other things
cs	(vi) ses	see you	t.ex.	till exempel	for example
lr	eller	or	ngt	något	something
B.S.D.V	Bara Så Du	Just So You	t.om	till och med	even, including
	Vet	Know			
p	på	on, at	Etc.	et cetera	
QL (ql)	kul	fun	m.m	med mera	and more
3vligt	trevligt	nice	m.a.o.	med andra ord	in other words
tebax	tillbaka	back	mkt	mycket	a lot
oxå	också	too	ibl	ibland	sometimes

A comparison between the examples given in the questionnaire and those that were found to be employed in the chat material shows that all innovative abbreviations that were exemplified were actually used, albeit with different orthographic form.

As Table 30 below suggests, Swedish chat participants create abbreviations in analogy with what is commonly found in chat rooms where English dominates. Single letter forms, where a word is truncated to represent the sound of a single letter when pronounced, like Werry's (1996) findings in French chat rooms, are often found. Acronyms, numbers representing the sound value of a syllable in combination with letters (e.g. *3vligt* = *trevligt* [nice]) are two common generated abbreviations. Unorthodox spelling may save time and effort, as well as provide an interesting touch (cf. "tebax [back] is ordinarily spelled 'tillbaks', see Example 23 below). These often save keystrokes, but the principle of economy does not always apply.

Table 30. The most frequent abbreviations used in the chat material

<i>No. of occurrences</i>	<i>Innovative abbr.</i>	<i>No. of occurrences</i>	<i>Traditional abbr.</i>
224	oxå	74	GBG
101	oki	60	gbg
62	Oki	56	ngn
47	é	43	mm
16	P	42	Gbg
10	iofs	37	ngt
10	if	26	bla
10	d	19	tex
5	tebax	19	Tom
5	OKI	18	etc
4	É	8	MM
4	ql	6	Ngn
4	p	5	BLA
4	OXÅ	4	tom
4	D	4	NGN
3	asg	4	Mm
3	IF	3	TEX
2	Oxå	2	TOM
1	cs	2	Ngt
1	Tebax	1	ngra
1	QL	1	bLA
1	If		
1	ASG		

Unconventional spelling may sometimes require just as many keystrokes, or sometimes even more than the conventional spelling. Example 22 below requires the shift key to be pressed, so typing the abbreviation this way does not save effort exactly. In most cases, the innovative spelling seems to be closer to the pronunciation than the traditional spelling is. The examples below illustrate how abbreviations occur in the chat material.

Example 22. *hade himla **QL** [=KUL]där förra helgen*
 [*had so much **FUN** there last weekend*]

Example 23. nåja, **tebax** [tillbaka, tillbaks] till ditt
 problem...*ber om ursäkt*
 [well, **back** to your problem...*sorry*]

Example 24. Miles..kul..? Mest trubbel denna veckan...suck...var **iofs**. [=i och för sig] på nattklubb i onsdags. Det var kul, kom i säng runt 4.00....fortfarande trött..*fniss*
*[Miles..any fun..? Mostly trouble last week...sigh... anyw. was on a nightclub last wednesday. It was fun, got to bed around 4.00....still tired..*giggle*]*

Example 25. Kram på dig **mé** [**med**]. Så du talar både svenska och spanska?
[Hugs to you too. so you speak both Swedish and Spanish?]

The proportion of abbreviations of the total number of tokens in the chat material amounts to roughly 0.02% (total number of tokens 410,355), with the reservation that there might occur other types of abbreviations not detected by the search tool.

5.5 Conclusions

The aim of this study was to analyze how Swedish web chat participants' use and adapt their written language to suit the conditions of written real-time conversation. As mentioned in the introduction, written conversation suffers from the time consuming labor of typing. Written conversation also misses out on many of the extralinguistic cues given in spoken face-to-face interaction. Written conversation has to be more explicit if senders want the receiver to get their meaning; written messages give no other clues than the text itself; the lack of information can render messages ambiguous and easily misunderstood. Several strategies were found to have been developed to minimize misunderstandings and the effort of typing, such as abbreviations, and all-capitals. Previous studies have shown that chat participants frequently use features that can convey information to guide the interpretation of written contributions, such as emoticons, asterisks, as well as extreme use of punctuation marks to express emotions and represent prosodic information. This study aimed at analyzing which strategies were used and developed by Swedish users for the adaptation of

Swedish written language to suit the conditions of real-time chat. It seems likely that the language use is not only an adaptation to the particular conditions of real-time chat, but also an adaptation to fit in with the chat culture.

Results showed that Swedish chat participants not only copy the strategies employed in international chat environments, but also adapt the Swedish written language to meet the conditions of online communication. Strategies such as the use of emoticons and repetitive keys were found to be similar to the use in international chat environments. Creative and new abbreviations of various kinds based on Swedish were used in addition to conventional established abbreviations as well as the ones copied from international chat environments. We saw also that Swedish chat participants use Swedish for the most part in Swedish chat rooms, but that they also use English under certain conditions. Code switching was sometimes found, where English phrases mix in among Swedish utterances as common fixed expressions.

It can be argued that apart from the fact that abbreviation strategies function to save time and effort as well as minimize ambiguity, that being able to master ‘the code of the cyber community’ not only makes the user a rapid communicator, but also tells the other participants something about the user. Just as non-verbal cues in face-to-face interaction can give information about the sender’s age, gender, social background, and so on, a chat participant is able to convey something about himself or herself. Showing off one’s ability to use strategies to express oneself tells other participants that the sender is a skilled chatter. Being able to present yourself as someone who masters the code, establishes the chatter’s identity and group membership. The text-only format also allows users to hide behind text and take on another identity. Chat rooms can act as virtual playgrounds in which people can try out a new identity and play with social roles, as well as foster creativity and language play. The material from the chat room showed that an extremely frequent topic of conversation was “looks and location”: people wanted to have detailed descriptions of how other people look and where they live. Perhaps they do this to find out if another person is a possible partner in the real world or maybe just for the thrill of knowing. Hiding behind the text allows for behavior that some participants would perhaps never try in the real world. In text-only conversation being sexually inviting or abusive, for example, is not dangerous – you are safe behind your screen and can say things you would never say face-to-face. Unlike interaction in the real world, one does not

have to stand up for what one claims in the cyber world⁶². People most likely adapt their way of talking to the conversation they observe. This goes for spoken face-to-face interaction as well as for adopting the use of strategies to express oneself in text-only conversation.

According to Ferrara et al. (1991), when confronted with a new communicative situation, people draw on previous knowledge of partially similar activities. Experience and knowledge from communication in other media are brought into new communicative activities. Cyber communication offers new aspects of communication which are not shared with any other type of interaction, just as it has qualities of communicative activities that one is already familiar with.

In short, this study suggests that Swedish chat participants conversing in electronic chat rooms adopt strategies that they observe in international chat rooms, and adapt these to suit the Swedish language. In using written language as a conversational tool, users draw on their previous knowledge of strategies employed in traditional written language to minimize time and effort when typing. This supports the claim of Ferrara et al. (1991): when confronted with a new communicative situation, people draw on previous knowledge of partially similar activities to form an amalgam:

...Competent users of a language have an extended language repertoire, and when new situations arise, they create new appropriate language varieties: They form hybrids.

Connecting the results from this analysis with Yate's (1996) call for investigating "[t]he extent to which human beings in specific social and cultural settings can develop and enhance their communication through the use of CMC", suggest that Swedish chatters import the main concept of using the strategies which have been discussed in this paper, and adapt these to the Swedish written language.

⁶² The chat participants quite certainly do not want their parents to know how they behave and what they talk about in the chat rooms.

6 Instant messaging with WebWho

6.1 Introduction

WebWho (Ljungstrand 1999; Ljungstrand and Hård af Segerstad 2000) is a lightweight, web-based awareness tool that shows a schematic view of the workstations in a large university computer lab. It visualizes who is currently logged in and where. As the system is reachable through a web page the ease of deployment is enhanced, as well as the accessibility of the system. WebWho explicitly conveys place information (i.e. real world user location) and provides a good overview for students to find the whereabouts of each other at-a-glance, as well as to find unoccupied computers in the lab (see Figure 12). In addition to the schematic overview of the lab, WebWho provides simple messaging services: normal email can easily be sent using a shortcut to the user's default email program, directly selectable from a pull-down menu in the web page. In the same way, students can easily look at someone's entry in the on-line student catalog, showing photos and personal contact information, as well as access that person's home page (see Figure 12). The system also has a function for sending short messages that instantly pop up topmost on the recipient's computer screen. The WebWho system was mainly intended to support collaboration and coordination between distributed users, primarily within different rooms in the lab building, but also for people situated elsewhere, such as students with Internet access at home. Although WebWho has a number of different - but related - functions, this study will primarily examine the instant messaging part, and in particular how different factors affect the content of the instant messages. A brief description of previous studies of other instant messaging systems will follow below. An

introduction of how WebWho works will be found after that. In order to provide a background for the present investigation, findings from previous studies of CMC, as well as features of different modes of communication, will be sketched. Method and material will then be accounted for, followed by results and a discussion.

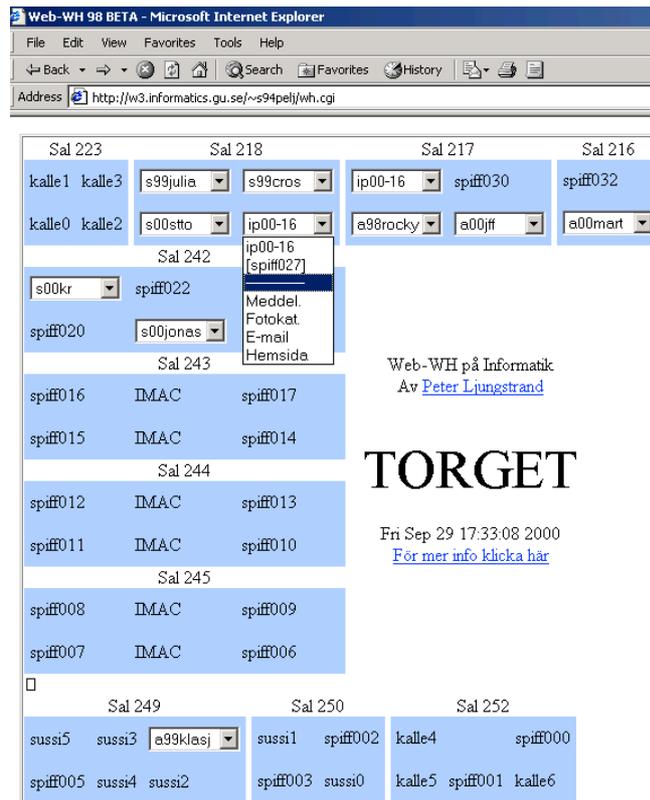


Figure 12. A WebWho screendump with one user's pull-down menu activated. The shaded boxes symbolize different rooms in the lab, each with two, four or six workstations, spatially organized as illustrated on the screen.

6.2 Awareness Tools

6.2.1 Instant Messaging and Education

WebWho allows for sending instant messages to specific workstations in a university computer lab. The sender of the message most probably knows the identity of the person logged in onto a particular workstation, because

the students collaborate on mutual assignments. Instant messaging systems of various forms have gained much popularity during the past few years. Commercial instant messaging systems such as *ICQ* and *AOL Instant Messenger*⁶³ have attracted millions of daily users in recent years, and the instant messaging phenomenon has also recently attracted researchers (e.g. Nardi, Whittaker et al. 2000; Smith, J.J. et al. 2000). In addition to desktop-based, on-line services, many millions of short messages (SMS) are being sent daily to and from mobile phones throughout the world.

WebWho is deployed at a university, in an educational setting, but it was not designed to specifically support education *per se* in a traditional sense. Research within the field of computer-supported collaborative learning (CSCL) and similar areas has investigated the use of CMC tools to support correspondence education, for instance (e.g. Wolz, Palme et al. 1997). The present investigation has not specifically looked at instant messaging systems, perhaps partly because of the somewhat chaotic nature of such systems. CMC systems within CSCL research generally support a learning situation where one or more teachers are “in control of” and direct the learning situation for and with the students, or at least they assume a teacher–student relationship in some fashion. WebWho does not have such built-in assumptions; it is used by students without any teacher interference or control, for whatever purposes the students might see fit. In this sense, WebWho is more similar to generic communication tools such as email and telephones than to systems specifically designed to support an educational setting.

6.2.2 Awareness Support and Active Maps

Research related to awareness of other users in one form or another has gained much attention within the CSCW and HCI (Human-Computer Interaction) communities. However, most systems proposed to support awareness have involved some overhead in order to work: specific hardware and software to be installed, login procedures, the need to explicitly state one’s current activity, etc. (e.g. Dourish and Bellotti 1992; Tollmar, Sandor et al. 1996; Pedersen and Sokoler 1997; Erickson, Smith et al. 1999). This is understandable given the experimental nature of these systems, but most end-users such as students simply want usable systems that work right away, without any hassle. A major requirement when

⁶³ ICQ (“I seek you”) was described in chapter 2. for further information, see <<http://www.icq.com/products/whatisicq.html>>, and <<http://www.aim.com/index.adp?promo=208886&aolperm=h>> (both URLs were retrieved 18 November 2002).

designing WebWho was that it should not use any custom software at all on the client side, and that the user's explicit involvement should be kept at a minimum, especially in terms of updating profiles, etc.

WebWho is a lightweight service that relies on readily available server status information, which is refined and visualized in a way that is easily accessible for individuals from any workstation with a web browser. No explicit actions (except the normal login procedure for Windows or Linux) are needed on behalf of the students to make their on-line status available to others. WebWho can be seen as an extension of the Unix command *who*, with a graphical interface to display location and messaging support. Indeed, the *who* command is one of the sources of information for the WebWho scripts on the server.

There have been other systems designed to support awareness of presence in real time by displaying a map, overlaid with up-to-date location information of people. For instance, ActiveMap (McCarthy and Meidel 1999) is a system deployed in a large corporate office setting. The system is based on active badges with a supporting infrastructure of beacons spread throughout the office environment. When running a custom application on a desktop PC, one can see a schematic map of the offices with information about who is where in the rooms and corridors. WebWho provides a much more lightweight and simplistic solution, with less location granularity, but much easier deployment.

Other systems have been created for visualizing the dynamics of electronic communities based on log files (Donath 1995) but such systems tend to be less useful for supporting synchronous or semi-synchronous activities. Smith et al. (2000) created *Threaded Chat*, a system for real-time visualization of threaded chats between multiple distributed users, somewhat similar to how threads in Usenet newsgroups are organized. Threaded Chat was designed to make it easier for users to follow the otherwise transient nature of chat or instant messaging (cf. discussion of the relative persistency of textual communication in the background chapter). Some on-line presence information was presented, but there were no fine-grained cues as to the whereabouts of the users in relation to each other or a local physical area. Both these systems were also intended for geographically dispersed users rather than co-located or almost co-located people.

Churchill and Bly (1999) presented a study of how a text-based virtual environment (such as a MUD) can be used to support communication among non-co-located colleagues. They concluded that text-only communication could offer a high degree of richness of expression, despite the fact that it lacks almost all the visual and auditory cues known to be

important in face-to-face collaboration. This suggests that even a very simple instant messaging tool like WebWho still can be used for very expressive communication. Similar results were found by Mitsuoka et al. (2001) in a recent study of i-Mode-based mobile phone services for university students in Japan.

A major difference between WebWho and many other awareness systems is that WebWho is primarily place-centered and only secondarily person-centered. As the students do not have their own personal workstations but rather have to share the existing ones with all the other students, they typically occupy different physical locations in the lab from one time to another. This is not the case with systems designed to support awareness of people's presence at a typical workplace, no matter if the system is intended to support people, usually in one office (Tollmar, Sandor et al. 1996; Erickson, Smith et al. 1999; McCarthy and Meidel 1999) or distributed people at different geographical locations (Dourish and Bellotti 1992; Smith, J.J. et al. 2000). With these systems, a person described by the system is generally also associated with only one place (desk, room or cubicle). With the Media Space system (Dourish and Bellotti 1992), it was possible sometimes to see more than one person at a time, but the system was installed in an environment where people had their own offices and tended to move around a relatively low number of places. Media Space makes it easier in case one wants to find coworkers, rather than finding students moving from place to place all the time. In the university computer lab, students can be logged in practically anywhere; there are often no "typical" places to look for them, at least not among the workstations. There are far more students than workstations, and the students often have to reserve a specific computer to use during a specified time slot, and then perhaps move to some other place to continue their work. It is important to students who want to engage in face-to-face interaction with their friends and classmates (as is often the case) to find out not only if the person they are looking for is in the lab, but also where that person is actually located. The rooms and the locations of the workstations in the lab are static, but the places where the students log in are not. Therefore, it seemed logical to have a schematic view of the workstations rather than a list of the currently logged-in students (as in the Unix who command) as the basis of the system.

6.3 WebWho in Use

Students use WebWho (as well as email, ICQ, mobile phones and face-to-face spoken interaction) to communicate and coordinate their actions both for work purposes and for social activities. When using a web browser to display who is in the lab, and where, the system takes a “snapshot” of all users that are currently logged in at that moment. This information is mapped to a schematic view of the lab rooms and displayed in a web browser, arranged as a map of the building (Figure 12 above). This schematic view allows students to find the whereabouts of their friends in the lab, as well as to find unoccupied computers at a glance, without having to physically search through the different rooms and floors in the building. When sending instant messages, the sender can choose to be anonymous to the recipient but this requires that he or she consciously and manually check a box before sending the message, see Figure 13 below.

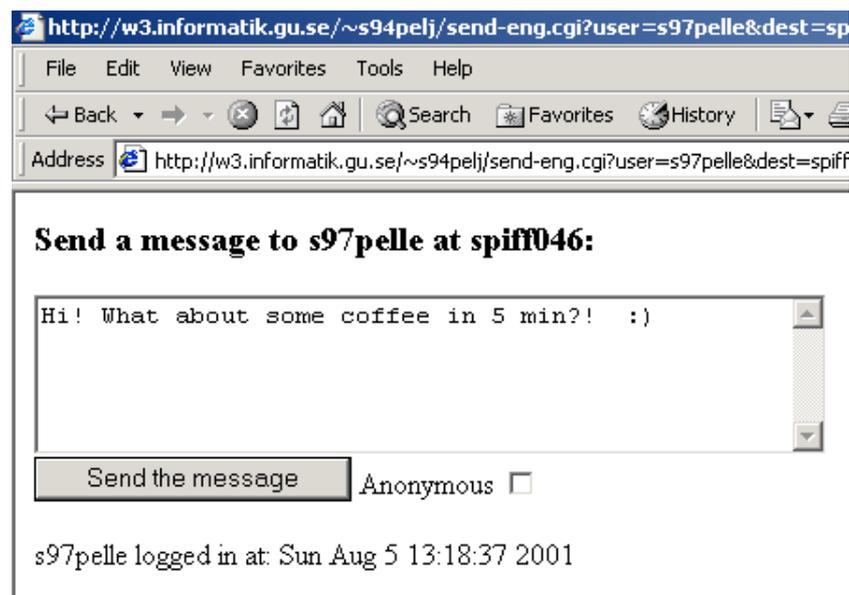


Figure 13. Snapshot of message window in WebWho.

The default case is messages with identified sender; the user name is automatically added to the message if the sender is logged in.

The university computer lab in which WebWho is used consists of a large number of rooms in one building, each with approximately six computer workstations. Most rooms are on the ground level in the building but some are located at higher levels. Since there are many more students than

workstations (some 700 students share the 120+workstations) the computer lab tends to be very busy at times; it is not always possible for students who are working together to get computers that are located right next to one another in the lab. WebWho allows them to locate easily the whereabouts of each other in the computer lab, and to communicate one-to-one in a near-synchronous fashion using the instant messaging system. The fast network makes the transmission time only a fraction of a second. The visualization of the computer lab supports the sender's awareness of the recipient being logged in at a specific workstation prior to sending an instant message.

WebWho has been on-line since November 1997. Even though WebWho was never publicly advertised, students have used it extensively. The instant messages sent through WebWho typically have a coordinating nature; they are used to arrange physical meetings (e.g. lunch, breaks for smoking, etc.), and for short questions and answers (e.g. regarding details of how to solve a particular programming assignment). In contrast to what one might expect, students do not perceive WebWho as an intrusion of privacy. During the years WebWho has been running, not one single student has complained about the fact that his or her on-line presence and very precise location is available to anyone on the web. Perhaps this is due to the fact that there is an option to block annoying incoming messages, though hardly used at all, and that the system is based on server data that are already publicly available to the students. Also, parts of the system (in particular, the on-line photo catalogue) are unavailable from computers with IP-numbers outside the university network domain. There might be students who dislike the system but who still for some reason have not complained, but nothing has yet indicated this.

In general, privacy issues are a major concern for awareness systems, and in particular for systems conveying detailed user information to just about anyone who is watching, which is one way to interpret how WebWho works. However, a detailed discussion of privacy issues was not the purpose of this study. For a more thorough discussion of privacy and awareness of presence as well as a system designed to take this into account, see Godefroid et al. (2000). Ethical considerations were considered in the background chapter above.

The primary aim of the present study is to examine if and how awareness of the recipient's presence affected the content of instant messages. To facilitate this, logs of instant messaging communication were collected during an extended period of time. The sender's and recipient's true identities were replaced with unique, but untraceable identification codes. It was hypothesized that being aware of the recipient's actual and

immediate presence would affect the topic of the messages. The location and identity of the recipient was given through his or her login name, as visualized through the WebWho web page.

Awareness of presence is one of several possible factors that influence what people write about in their messages, and the one main concern in this particular study. As has been proposed in this dissertation, other factors that influence communication are related to the variables *synchronicity*, *situation*, and *means of expression*. Texts in instant messages are certainly the products of written means of expression. The mode is asynchronous; the texts convey fewer cues to guide interpretation and are more of an effort to produce, compared to the ease of production, multimodality and synchronicity of spoken face-to-face interaction. Networked communication is a tertiary means of expression that gives less obvious clues to identity. Situational parameters such as relationship between communicators and their shared context and background knowledge are also influencing factors according to which communicators can be assumed to adapt language use and topic of conversation.

6.4 Computer-Mediated Communication

6.4.1 Features of CMC and factors that influence language use

In order to provide a background to the present study, a brief sketch of some findings from previous studies of CMC will follow.

CMC is the communication produced when human beings interact with one another by transmitting messages via networked computers. Most CMC today is text-based communication, typed on a computer keyboard and read as text on a computer screen (Herring 2001). Different modes of CMC set the stage differently for the type of communication to take place; users communicate either asynchronously, for example via email, in which the receiver of the message does not have to be on-line in order for the communication to take place, or they communicate synchronously, or to put it more accurately: near or quasi-synchronously (cf. Ferrara, Brunner et al. 1991; Garcia and Jacobs 1999), like in a web chat or IRC, which require all participants to be logged on at the same time.

Many studies have claimed that different types of CMC, such as chat and email, are hybrids between speech and writing (cf. Du Bartell 1995, among several others). Email is said to take on some of the features of

spoken interaction because of the speed of message exchange, but employs many features of traditional writing as it is still in fact a written medium (Maynor 1994). Studies comparing different kinds of writing techniques have shown that the faster the written medium, the more like spoken language the written messages get (cf. Horowitz and Berkowitz 1964). Horowitz and Berkowitz' study argues that the language in steno typed messages show more features that are normally associated with the spoken register, than messages that are handwritten or typed on a typewriter. Du Bartell (1995) concludes from a study of features of messages in a mailing list that the spoken-and written-like characteristics in a written medium result from the constraints imposed by the computer medium - the machine architecture. This view seems somewhat simplified, as it does not take into account situational parameters such as relationship between participants or goal of interaction, among several factors. Herring (Herring 1996) points out that email messages seem to get more informal in terms of both composition (salutation and closing conventions) and form (spelling, syntax) because most email software automatically prompts information about the user and details of posting.

CMC messages display linguistic characteristics typically associated with spoken language and other forms of written language in addition to linguistic features specific to the medium. Maynor (1994) argues that the syntactical features of "e-style" sometimes reflect informal habits of speech when pronouns, subject, verbal auxiliary, copula or modal are often omitted. Du Bartell (1995) claims that we expect written language to be edited, planned, articulated without recourse to non-standard constructions, slang and vulgar expressions. From speech we expect more or less the opposite: we expect slang, non-standard grammatical constructions, sudden topic shifts and spontaneity. The same kind of argument is found in Baym (1996), who claims that although CMC is written it is marked by many features typically associated with face-to-face interaction.

Previous research has shown that the purposes of communication, topic and means of expression play a part in the way messages are formulated (Du Bartell 1995; Baym 1996; Hård af Segerstad 2000b; Hård af Segerstad 2001). Situational parameters, such as the influence on messages by the relationship between sender and recipient, for example whether the sender and receiver know each other and their social status, was discussed by Danet (2001). Partly as a result of having a subject displayed, email messages frequently omit even the typical salutations and farewells associated with other media, regardless of whether the speakers know each other. Email messages do display rather informal register characteristics,

even between persons unknown to one another (Danet 2001). Recent research has shown that instant messaging takes this even further, in that much of the context of a short message can be left out from the text (Nardi, Whittaker et al. 2000).

For this particular study, it is of interest to find out whether the sender's awareness of the receiver's presence influences the content of messages.

6.4.2 Modes of Communication

Traditional written communication is typically asynchronous, assuming non-presence and interaction with severe time delay. Writing is a means of expression that is monomodal, employing vision only, and its production is physically laborious and time consuming. Written communication employs a graphic system for representing only the vocal aspects of spoken interaction, employing 28 alphabetic graphemes in Swedish to represent it, not counting other prosodic phenomena such as emphasis and tone of voice, etc. Handwritten messages provide clues to the sender's identity from the handwriting itself. Computer written messages give fewer clues. The effort of sending traditional mail, "snail mail", involves the trouble of putting a piece of paper in an envelope, addressing it and stamping it, carrying it to the post office, etc. Not to mention the time delay of having to wait several days for it to reach its destination (cf. Maynor 1994)!

Like traditional mail, email is an asynchronous messaging system without the need for visual confirmation of the receiver's presence at the time of transmission. Email communication shares many features of traditional written communication: it is indeed written using the same graphic system and monomodality as traditional writing. The interaction suffers much less from time delay, though, and the ease of access for sending messages is considerably greater.

The relative user anonymity of computer-mediated production and transmission might also play a part in message composition. Web chat is a near-synchronous (cf. Ferrara, Brunner et al. 1991; Garcia and Jacobs 1999) conversational tool in which chat participants know that other participants are logged on. This is also true for WebWho. Web chat participants very often have no personal knowledge of the users, in contrast with WebWho interlocutors who very often are classmates and friends. The time delay is much less severe compared to email interaction, and in this respect comes closer to spoken communication. Chat systems and instant messaging systems both require synchronous participation. The interaction is only near-synchronous, though; the messages have to be typed first and then transmitted, whereas telephone and face-to face

interaction are fully synchronous modes of communicating. Chat participants do not share the mutual context of a computer lab, which is the normal case for WebWho communication.

Default spoken face-to-face interaction is multimodal, sending and receiving information through both visual and vocal modalities. The effort of production is minimal and the exchanges rapid. Awareness of presence increases the more synchronous the communication gets. It is a matter of discussion whether web chat is more synchronous than instant messaging communication: both modes require both the sending and the receiving party to be logged in simultaneously to succeed. Web chat normally allows both parties to read the whole conversation in a window on the screen, whereas instant messages, such as ICQ messages, often come one by one.

Simply viewing and comparing modes of communication in a linear fashion along the synchronicity continuum does of course not give us the whole picture of what is going on in communication. As previously mentioned, a whole range of other factors, as have been introduced in the present dissertation, are also important for how we engage in the action, what we talk or write about and how we use language.

6.4.3 Awareness of presence in WebWho

Because awareness of presence in the instant messaging system WebWho is a feature that is unique among the modes of CMC investigated in the present dissertation, this particular study primarily aimed at finding out whether awareness of presence affects message content. It was also interesting to investigate what the students were using the instant messaging service for. Does the awareness of the receiver's presence seem to affect the purpose of the messages? The WebWho message log was compared to corpora of spoken interaction and traditional written language, as well as to other modes of CMC that are of (slightly) different settings, namely email and web chat. In what aspects does the instant message mode of communication differ and show similarities with speech, traditional writing, email and chat?

It was predicted that the WebWho tool would be used by students to coordinate their work for group assignments and to coordinate social activities, such as coffee breaks and activities outside the university. Being aware that the receiver is actually there to read the message at the time that the sender transmits it was taken as a main factor that would affect the topic of the message. The identity of the receiver was revealed through the login name, a student ID, whose referent was most probably known to other students within an assignment group. These would be impossible for

someone outside the university environment to decipher. The predictions concerning how awareness of presence affects the content of messages had to consider real-world facts: not only the “*virtual presence*” as visualized through WebWho, but also *physical presence* when the recipient is located in the same lab room as the sender. The message logs were divided into the following settings, based on the messages sent between students.

- *Co-located*: Both sender and recipient were located in the same lab room at the same time, hence it was possible to have physical awareness of each other’s presence, and of course to see one another and talk directly face-to-face (both physical and virtual awareness of presence).
- *Distributed*: Sender and recipient were located in different lab rooms in the same building, using the WebWho tool to locate the presence of each other (virtual awareness of presence).
- *Distant*: The sender accessed WebWho from outside the building, using a dial-up connection or the like, to locate the recipient (virtual awareness of presence). The system only allows for receiving messages from outside the university building, and not for sending messages outside.

As mentioned above, WebWho allows for messages to be sent anonymously. That means that the recipient cannot tell who sent the message, and the pop-up window is quite similar to a system error message window. The log showed messages that were sent with non-identified senders. An analysis of these anonymous messages is especially intriguing: what kind of messages did the senders choose to send anonymously, and why were they sent? A cross analysis of *sender location* and *sender identity*, whether the sender identify himself or herself or choose to send anonymously, was made for this purpose.

6.4.4 Data Collection and Analyses

Different types of logs were gathered during several years of WebWho usage, examples of these are frequency of use of the overview page visualizing the lab, frequency of use of the instant messaging service, usage from within or without the university network domain. A number of messages from the instant messaging service were extracted. The resulting corpus, the WebWho message log, that was analyzed for this study consisted of a total of 8,231 logged messages sent during the period September 1, 1998–December 31, 1999. The messages were made

anonymous during collection so that the original sender cannot be identified. The logs still contain the essential text of the messages.

6.4.5 Quantitative Analyses, Automatic Measures

The WebWho corpus was analyzed automatically using TraSA, as described in the background chapter above, in order to find out the number of messages sent, word frequency (also applicable to “word-like” elements like emoticons), sender location (same lab room, other lab rooms, at home or elsewhere) and sender status (anonymous vs. **identified**). The texts were searched for occurrences of emoticons, in order to investigate a feature that has been proven to be CMC specific. Automatic analysis of other measures such as mean length of utterance and word frequency was based on the total number of messages. The WebWho material was also compared with respect to word frequency to spoken and traditional written language, and with the corpora of email and web chat gathered for the present dissertation.

6.4.6 Qualitative, Manual Analyses

In order to analyze whether awareness of presence seemed to affect message content, qualitative manual analysis of message topic was carried out. Cross-analyzing three categories of *sender location* - *Co-located*, *Distributed* and *Distant* - as defined in Section 6.5.3, and two categories of *sender status* (and receiver status) - *Anonymous* and *Identified* - resulted in six categories. In order to perform a closer analysis of what people used instant messaging for, a set of 100 messages from a set of continuous messages was extracted from each category⁶⁴, resulting in a sample of 600 messages out of 8,255 total. These were manually analyzed for apparent main topic or reason for communication.

Table 31. Data from the instant messaging corpus.

	<i>Instant Messaging</i>
<i>No. messages</i>	8,255
<i>No. words</i>	111,025

⁶⁴ The decision to extract messages in succession was made in order to follow possible dialogs, thus getting some clues to how the content should be interpreted (cf. Nardi et al. 2000).

Figure 14 below gives an overview of the distribution of the different message categories. The messages may in many cases contain several topics, but were categorized for topic by what seemed to be the main content. Categories for topics⁶⁵ were, for example, task related, social coordination⁶⁶, mischief, meta-comment, etc. A number of categories emerged and will be described and exemplified below.

6.5 Results

6.5.1 Comparison with Speech and Writing, and with CMC

A comparison of the most frequent pairs of words in the WebWho data and corpora of spoken and traditionally written Swedish reflects both the activity in which the linguistic interaction took place (Allwood 2000), and the means of expression through which they were sent. The contents of the instant messages sent via WebWho naturally concern the activities in which students were involved; the most frequent pairs of tokens (two words occurring together) reflect this.

Table 32 below illustrates the most frequent pairs in a comparison across five modes of communication: traditional writing, email, web chat, instant messaging, and speech. Traditional written language and spoken language corpora were taken from The Swedish Spoken Language Corpus at Göteborg University⁶⁷. The email and web chat corpora were data collected for the purpose of the present dissertation. The most frequent spoken words in Table 32 below are rendered in their written equivalent, supplemented by an index number or letters in brackets for disambiguation. The second most frequent spoken word, [*e*₀] stands for [*är*=is, are] and disambiguates it from [*e*₁] which indicates “hesitation sound”. In the same way, [*ä*₀] in the table below stands for [*och*=and],

⁶⁵ Topic, activity or “functional moves” (cf. Herring 1996).

⁶⁶ “Social coordination” refers to coordinating social activities, such as going for lunch together or the like. In the present study the concept was not used to indicate social coordination in the sense of coordinating the social interaction in a communicative situation, in terms of feedback signals (cf. Chapter 2) by which the speaker and listener monitor each other’s involvement in the interaction.

⁶⁷ <http://www.ling.gu.se/SLSA/SLcorpus.html>

which disambiguates it from [*å*₁] which indicates [*att*=to]. The word [*ja*{*g*}=I] is disambiguated from [*ja*=yes]. The token *sanna*, which occurs in the web chat material, was a nickname that was extremely frequent in the chat data.

Table 32. A comparison of the most frequent pairs across five modes of communication.

<i>Writing</i>		<i>Email</i>		<i>Web chat</i>		<i>IM</i>		<i>Speech</i>	
<i>Token</i>	<i>Transl.</i>	<i>Token</i>	<i>Transl.</i>	<i>Token</i>	<i>Transl.</i>	<i>Token</i>	<i>Transl.</i>	<i>Token</i>	<i>Transl.</i>
det är	it is	i göteborg	in göteborg	som vill	who wants	/ /	/ /	de(t) e0	it is
för att	because	. jag	. I	är du	are you	går det	does it work	e0 de(t)	is it
det var	it was	hej jag	hi I	? sanna	? sanna	. Jag	. I	att de(t)	that it
att det	that it	för att	because	jag är	I am	det är	it is	men de(t)	but it
är det	is it	tacksam för	grateful for	du ?	you?	är det	is it	e0 ju	is why (interj.)
i en	in a	tack på	thanks in	, hallå	, hello	-)	-)	så att	so that
att han	that he	jag undrar	I wonder	det är	it is	JAVA JAVA	JAVA JAVA	de(t) va(r)	it was
i den	in the	på förhand	in advance	sanna ,	sanna, ,	:)	:)	å0 så	and then
som en	like a	jag är	I am	hallå ?	hello?	, jag	, I	de(t) här	this
men det	but it	jag har	I have	har du	do you have	Hej !	Hi !	å0 de(t)	and it

The frequency list is slightly modified, as digits were found among the 10 most frequent pairs in the email data, which made it difficult to compare with data from the other modes. The digits appeared in the 2nd, 8th, and 10th place in frequency order. Leaving the digits in the list would be like comparing a particular word that is frequent to a whole class of words (i.e. digits).

From the contents of the messages one may deduce that the sender and receiver know each other, and frequently draw upon shared background knowledge, as the contents can only be interpreted if the context is shared (see Section 6.6). Thus, the content and form are structured for that purpose. Content, or topic in the different kinds of data, reflects the various activities that the messages have been used for. As was shown in

the chapter on web chat, many of the messages in the web chat material seemed to be contact seeking (“who wants to chat?”, “is there anyone who **wants to...**”), and very often concerned what people looked like and where they lived.

Such messages were not found in the WebWho data presumably because most of the students who use the instant messaging system know each other. There simply is no need to ask about a person’s looks or where he or she lives, as the sender probably already has such information prior to sending a message. It is probable that the anonymity of a web chat calls for explanation of things that are apparently of such importance in young people’s minds, and that anonymity allows for things to be communicated that would never be verbalized in a face-to-face or a telephone conversation. None of the messages in the email corpus (see the chapter on email) dealt with contact in this personal way, but are of a more business-like matter. The data were taken from a corpus of electronic communications to a city council, and thus concern what can be expected of it: pleas for information and advice in legal matters, messages dealing with the information structure of the web page of the city council, for example.

Returning to the comparison of frequent pairs (Table 32 above), it can be deduced that the large corpora of spoken language and traditional written language share five out of ten most frequent pairs. The most frequent combination is the same in both modes (“de e” and “det är” both corresponding to “it is”, respectively). This combination of words was also found among the ten most frequent in both the web chat and the instant messaging data. Traditional written language shares one item with the email corpus (“för att” = “because”). The instant messaging corpus is the only one that contains emoticons among the ten most frequent pairs.

Comparing the WebWho messages with a Swedish email corpus (Hård af Segerstad 2000b) and a Swedish chat study (Hård af Segerstad 2000a; Fortchoming) with respect to the occurrence of emoticons, shows that the proportion of emoticons is different in the different settings. For instance, the most “basic” emoticons :-) and :) appear in 6th and 7th place, respectively, in frequency order for the WebWho corpus, whereas emoticons did not occur until the 461st place in the chat corpus. The email corpus as a whole only showed two occurrences of emoticons. The development of emoticons, by now a well-known feature in CMC, serves to convey paralinguistic information and to enhance the written mode of communication, by conveying more than traditional written language normally allows for (cf. Du Bartell 1995).

Instant messaging, such as communication via WebWho, is spontaneous, rapid, and has a number of features in common with the spoken register; it is playful, dynamic and speech-like (cf. Horowitz and Berkowitz 1964; Bolter 1991).

6.5.2 Physical and Virtual Awareness of Presence

Messages from the first group, “Co-located”, were all sent within the same lab rooms in which the senders of messages could be physically aware of the receiver. Eight percent of all messages were sent in this setting (692 messages).

Two categories form the second group: the first, “Distributed” messages were sent between different lab rooms 70% of the total number of messages (5,675 messages), the second, “Distant”, messages were sent from somewhere outside the university computer lab, (1,864 messages), for 22% of the total. Since Distant messages are often sent using a dial-up connection, the senders had no physical evidence of the recipients’ presence, but were informed of their presence virtually through the WebWho web page.

From the analysis of the messages, it was revealed that the topics are different in the three settings. On the whole, messages sent between lab rooms (Distributed) tended to be related more to social coordination, such as coordinating for going to lunch together. This group also contains more work-related messages with assignment coordination content. Messages sent within the same rooms tend to be more mischievous in nature. Anonymous messages made up 13% of the sample; Figure 14 shows the proportions of messages with identified and anonymous senders in the topic categories.

6.5.3 Message Categories

The following section describes the categories in the different settings and illustrate with examples from the log. Diagrams illustrating cross-analysis of sender location, sender status and message topic are found below.

Task Related

Task related messages concerned collaborative assignments, or individual assignments on the course curriculum. The majority of the task-related messages with identified senders were sent in the *Identified Distributed* category, while the minority was sent within the same lab rooms, in *Identified*. Example 26 illustrates a message that was coded as task related, and that was sent from a user who was identified and located in another room than the receiver (Identified Distributed).

Example 26. ska vi fortsätta med 10:an?
[shall we get on with number 10?]

This is not particularly strange, rather it was what could be expected: people communicate about the task they are working on. When there is an

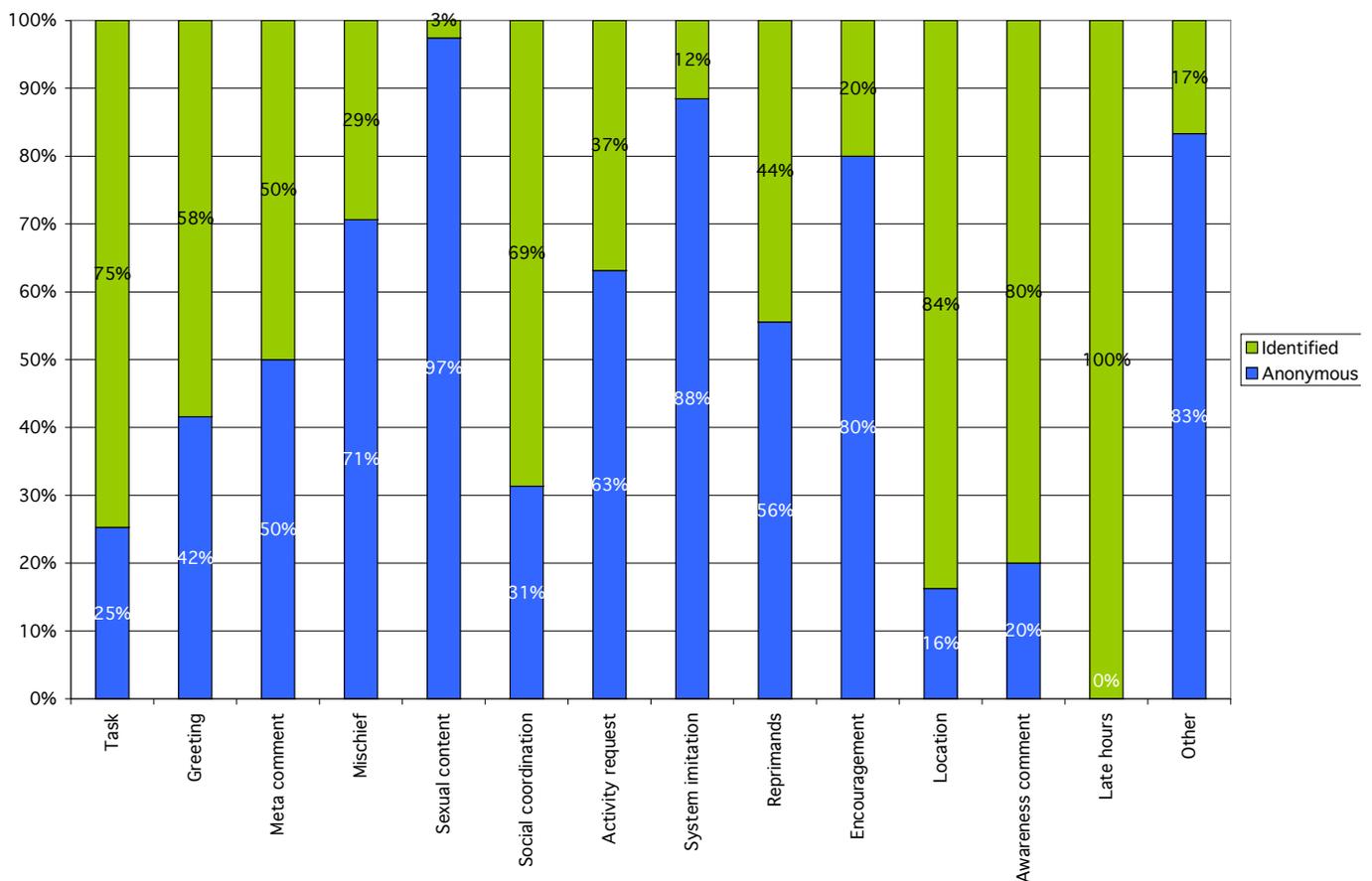


Figure 14. Message categories with proportion of messages that were sent by identified or anonymous senders.

opportunity for face-to-face interaction in the same lab room, there is no need to send textual messages: it is faster and easier to talk about assignments. This theory is supported by the fact that collated messages make up only 8% of the total.

The majority of the anonymous messages were sent between lab rooms (*Anonymous Distributed*), while the minority was found in *Anonymous Co-located*. Sending this type of message anonymously seems a strange way to cooperate on group assignments, no matter where you are located; however, some examples of this were found. Below is an example of a message that was sent by an anonymous sender to a person located in the same room (example 27).

Example 27. Hur går det för er? Behöver ni hjälp så kan ni posta tillbaka. Hej svejs!
 [How are things working for you? If you need help just mail back. Bye!]

Greeting

Messages that were coded to be greetings simply were greetings, or message openings, of various kinds. Identified senders within the same lab room sent the majority of the greeting messages (*Identified Co-located*), while the minority of greetings were sent between lab rooms (*Identified Distributed*). Example 28 below is a greeting message that was coded *Identified Co-located*.

Example 28. GODMORGON, ALLA PIGGA OCH GLADA
 IDAG?????????????
 [GOODMORNING, EVERYBODY ALERT AND HAPPY
 TODAY?????????????]

Intuitively, this seems contradictory to what one could expect: why send a computer-mediated message to say hello when you are located in the same room, where you could just turn your head and say hi? As noted by Nardi et al. (2000), instant messaging is often used for informal communication; it would be considered lunacy to deliver a “Good morning” message in email - you can never tell when a person is going to check his or her inbox - but people appreciate a quick instant message greeting. One of their informants explained: “It’s a nice way of saying hi without being intrusive” (Nardi, Whittaker et al. 2000). An American informant resident in Sweden

remarked that walking into a lab and yelling ‘good morning’ in Sweden might get you committed to an asylum.

The majority of the anonymous greeting messages were sent between people within the same lab room (*Anonymous Co-located*). These are perhaps just a way of checking the receiver’s reaction, or may be similar to the type of greeting noted by Nardi et al. The minority of the messages were sent as *Anonymous Distant*, which seems a strange thing to do, as the sender gets no feedback at all of his or her action because neither is it possible to witness the receiver’s reaction nor to get a written reply. Example 29 below illustrates a greeting message that was coded *Anonymous Co-located*.

Example 29. Hej era döingar!
[Hi there you stifffs!]

Meta Comment

Messages that were coded as meta-comments concerned comments on the system itself, and functioned as tests of what one could use it for. The majority of the messages with identified senders that commented on the message in itself or the messaging system were sent from outside the building (*Identified Distant*), while the minority was in *Identified Distributed*. Both are relevant since the messages concern the messaging system itself, as shown by example 30.

Example 30. Fick du mitt meddelande ? Kan du vara snäll
och bekräfta med ett mail.
[Did you get my message ? Could you please
confirm it with an email.]

The same goes for messages sent anonymously: the majority of meta comments were sent in *Anonymous Distant* and the minority in *Anonymous Distributed*. Note that it is difficult if not impossible to reply to an anonymous sender and confirm that the system works. Example 31 illustrates a meta-comment sent by an anonymous sender located somewhere outside the university network.

Example 31. Nu ska vi se om detta funkar.... säg till
om det kom fram :-) //Förnamn
[Now let’s see if this works.... tell me if
it reached you :-) //First name]

Mischief and Nonsense

Messages that were coded to be mischievous or just nonsense had varying types. Identified senders within the same room (*Identified Co-located*) sent the majority of the messages with mischievous content, while the minority was from rooms in the same building (*Identified Distributed*). Example 32 below shows a message that was coded *Identified Co-located*.

Example 32. MÖGEL...
[MOLD...]

The majority of the anonymous mischievous messages were also sent within the building (*Anonymous Distributed*); the minority of these were sent as *Anonymous Distant*. An example of a nonsense message sent by an anonymous sender located outside the university network is shown in Example 33 below.

Example 33. blääääääääää!!!
[yuuuuuuuuuuck!!!!]

It does not seem particularly strange to send mischievous messages anonymously, regardless of location. The number of messages sent anonymously for this purpose is clearly higher than the ones with an identified sender. It seems that when no one knows who you are or where you are located, it is much easier and perhaps also more tempting to be naughty.

Sexual Content

The only messages that concerned sexual insinuations, or rude language, in combination with an identified sender, were sent in within the same lab room (*Identified Co-located*). An illustration is given in Example 34. It was written in English, and it could be argued that using another language than your mother tongue is to push responsibility away, as it were. However, there was no evidence (apart, perhaps, from the limited mastery of spelling and preposition use) that the sender knew any languages other than English.

Example 34. very good you sexgod.what about sex on the
toilette? or are you not as horny as i am

None of the messages in the corpus seemed to be of a really harassing nature; rather they seemed to have a friendly, playful tone. A plausible explanation is that the interlocutors knew each other and also wanted to see or hear the receiver's reaction later to the mischievous messages.

It seems that this kind of message is easier to send when you are anonymous: the majority of these were sent from home, or other locations than in the computer lab (*Anonymous Distant*), but almost as many were sent in the category *Anonymous Co-located*. The minority were sent in the *Anonymous Distributed* category. The example below (35) was coded *Anonymous Distant*.

Example 35. Tryckte Du henne med de stora bröstena
tillslut igår? :0)
[Did you pump the one with the big boobs
yesterday? :0)]

Social Coordination

Messages concerning social coordination concerned, for example, coordinating lunch breaks or invitations to go partying. The majority of these were sent in *Identified Distant*, from home or locations other than in the computer lab. The minority of messages with identified senders were sent in *Identified Co-located*, not unexpectedly. In this setting it is easier and more convenient to conduct these matters in spoken face-to-face interaction. Example 36 illustrates a typically coordinating message.

Example 36. Nu går jag strax. Hur gör vi imorgon inför
kvällen: höras av, träffas, present etc.
[I'm leaving soon. What about planning for
tomorrow night: get in touch, meet, present
etc.]

Strangely enough, some messages sent for social coordination were sent anonymously. It seems difficult to arrange something when you do not even know who sent the message. The majority of these were sent from places outside the computer lab, *Anonymous Distant*. Messages of this kind sent in *Anonymous Co-located* are perhaps more easily explained by the possibility to check the receiver's reactions; an example is given in Example 37:

Example 37. Its time!!

A possible explanation why people send messages coordinating social activities anonymously could be that the sender knows that some previous, shared background knowledge is being activated by the message, and thus it is not a complete waste but a funny way of contacting one another, perhaps experimenting with what the messaging system can do.

Activity Request

Messages that concerned requesting information about what others were up to were coded “activity request”. An equal number of messages was sent by identified senders in the same lab room (*Identified Co-located*) as from locations outside the building (*Identified Distant*), with a minority being sent from within the building (*Identified Distributed*). An example is given below; it was sent by an identified sender located in a different room from the receiver.

Example 38. Hej! Vad har du för dig? /Förnamn
[Hi! What are you doing? /First name]

Messages requesting information about what someone is doing seem to function as mockery, “I can see you”, to lead the receiver to believe that he or she is being watched. The majority of activity requests with anonymous senders were sent from other places than from the computer lab (*Anonymous Distant*). Some were sent within the same lab room (*Anonymous Co-located*), perhaps corresponding to the same function as just mentioned, or to something like “quit it”, depending on the context. Out of the total number of messages of this kind, most were sent anonymously, which seems to confirm this suggestion. Example 39 illustrates an activity request from an anonymous sender located outside the computer lab.

Example 39. din lilla bonde vad gör du
[you little peasant what are you doing]

System Imitation

System imitation messages were typically sent in an attempt to make the receiver believe that the computer system was breaking down or something that is equally frightening when working on some important task. The messages were all written in English in order to be more convincing. When

sent with an identified sender, the only messages imitating various system directives were found in the *Identified Co-located* category. This suggests that the sender wants to witness the receiver's reaction. Example 40 was sent in the *Identified Co-located* category.

Example 40. System is going down for reboot within 5 second..

Most of the messages imitating system directives were sent anonymously, and they were sent both from within the building (*Anonymous Distributed*) and from other locations (*Anonymous Distant*). A system imitation message that was sent in the *Anonymous Distant* category is illustrated by example 41.

Example 41. Fatal error! Please close all programs and restart Windows.

The minority of these were found in *Anonymous Co-located*, which is the category one would have expected to be the most frequent, because it would resemble real system messages insofar as these messages are not sent by some person with a student identity.

Reprimands

Reprimands were messages telling people to “shut up and get to work”. These were, perhaps naturally, sent most of the time by identified senders within the same lab room (*Identified Co-located*), as other people working nearby probably got disturbed sometimes. This category is exemplified in example 42 below.

Example 42. Vad är det du sysslar med egentligen?????
[What are you doing really?????]

Equally naturally, none were sent between the lab rooms. Strangely enough, some were also sent from other locations where no such interference is possible. Perhaps a student who knows from experience that the receiver is likely to be chatty and disturbing, sent a message just to annoy and to keep in touch, with the function “I’m here, I’m working and keeping an eye on you”. Perhaps for the same reason as reprimands with identified senders, the majority of the anonymous messages of this kind were found in the *Anonymous Distributed* category. Some were sent anonymously within the same lab room (*Anonymous Co-located*), and a few

from other locations. Example 43 illustrates a reprimand in the category *Anonymous Co-located*:

Example 43. shut the fuck up

Encouragement

Messages coded to be encouraging were messages with identified senders sent from outside the computer lab (*Identified Distant*), perhaps by some student in the same work group as the receiver. None were sent either within the same lab room, or between rooms.

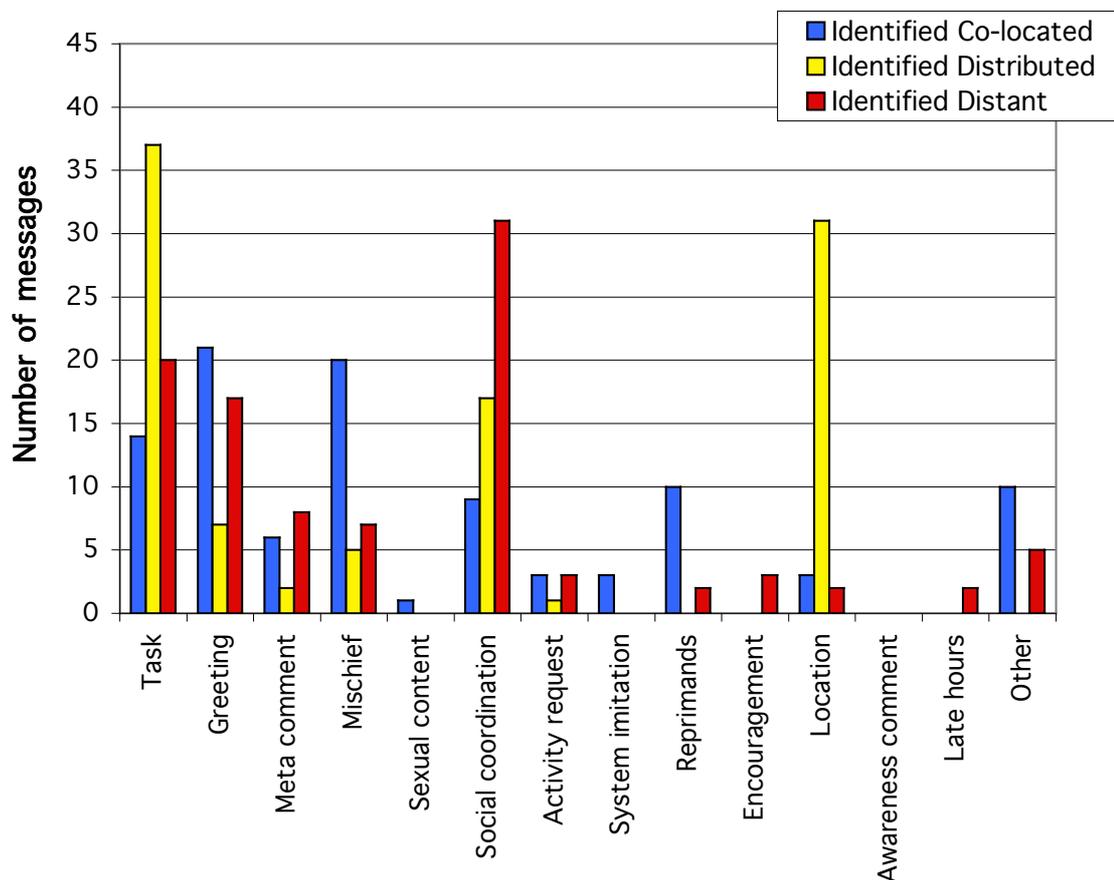


Figure 15. Messages with identified sender. We can clearly see that the purpose, or function, of the messages are different in the three settings. On the whole, messages sent between different lab rooms (Distributed) are more about social coordination, such as coordinating going to lunch together. This group also contains more work-related messages with assignment coordination content.

Example 44. Java nisse, kämpa på, vi på xgatan stödjer dig
[Java gnome, keep struggling, we here in xstreet support you]

A number of encouraging messages were sent anonymously. Example 45 illustrates an encouraging message that was sent anonymously within the same lab room. Strangely enough, despite being made anonymous by the sender, it was signed with what appears to be sender name.

Example 45. jobba på :-)förnamn
[keep it up :-)first name]

Again most of the anonymous encouraging messages were sent from other locations outside the lab (*Anonymous Distant*). Some were sent within the lab rooms, but none at all were sent between rooms.

Location

Messages coded to concern location inquired where people were located, and about possible un-occupied computers. Most of the messages in this category were sent by identified senders between lab rooms (*Identified Distributed*). This is what was to be expected; the students coordinate their whereabouts and try to get workstations close to one another if they collaborate on mutual assignments or just like to sit close together. A few messages were sent within the same rooms, and fewer yet were sent from other locations. An example coded *Identified Distributed* illustrates this category:

Example 46. Sitter du på 55:an? Förnamn
[Are you at number 55? First name]

Some messages concerning location and computer whereabouts were sent anonymously. These were of the type “who is logged on to this particular computer?” (the students’ ID may not always give clues to the real person’s identity - sometimes a temporary ID (to be used during a certain course) is used for logging in). The message in example 47 was sent by an anonymous sender located outside the university network.

Example 47. Jaså, det är där ni sitter...
[Aha, that’s where you are...]

Naturally, there were messages concerning location that were sent within the same lab room, as there is physical awareness of both fellow students' location, un-occupied computers and who the actual physical person logged on to a particular computer is.

Awareness Comment

Messages commenting on being aware that a person was located some place were coded *Awareness Comment*. Messages of this type were not sent by identified senders at all. The only anonymous messages that concerned this topic were sent within the same lab room (*Anonymous Co-located*), or between the lab rooms (*Anonymous Distributed*). This suggests that the sender wanted to see the receiver's reaction or just to keep in touch, or inform the receiver that the sender is on location. An example illustrating the category *Anonymous Co-located* is given below:

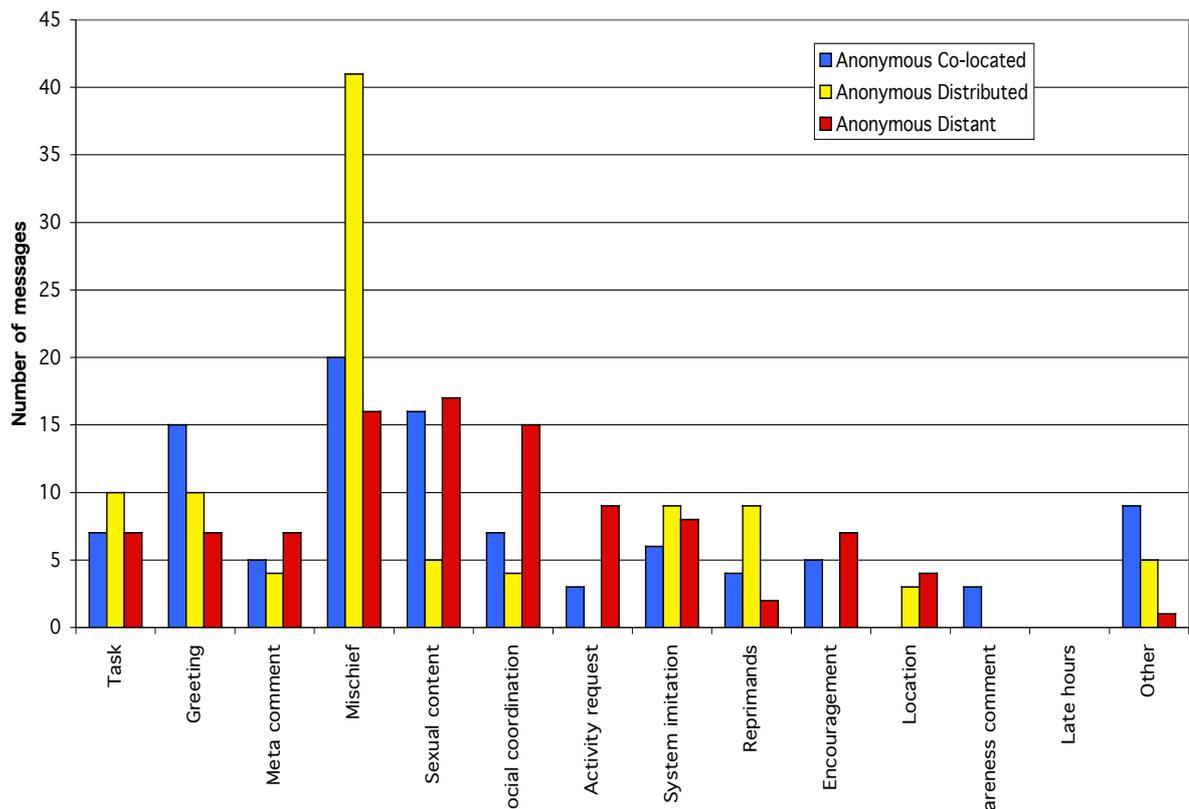


Figure 16. Messages with anonymous sender. There is a high proportion of mischievous messages sent within the same lab rooms, in which the sender has physical evidence of the recipient's presence.

Example 48. Jag ser dig...
[I can see you...]

Late Hours

A few messages were sent asking what a person was still doing in the lab, or “go home now”, “does your mother let you stay out this late”. They all fell into the category Identified Distant - probably, some student checked the lab via the web browser from home.

Example 49. Hörru din galning, nu får du allt gå hem å
lägga dig... Klockan är ju inte barnet
längre...
[Hey you maniac, you better go home and go
to bed now... The night is not young
anymore...]

No messages at all of this type were sent anonymously.

Other

Messages falling into the admittedly vague category of “other” did not match any other category. Examples are:

Example 50. Olli har alltid feeeeeeeel!!!!
[Olli is always wrooooooong!!!]

Example 51. Hobbes Hub dominerar
[Hobbe's hub dominates]

These seem to be comments of some kind. This type of message is obviously very hard to interpret without relevant contextual knowledge. The majority of these messages were sent within lab rooms with identified senders (*Identified Co-located*). Some as yet uncategorizable messages were sent anonymously within the same lab room (*Anonymous Co-located*) and fewer still were sent between lab rooms (*Anonymous Distributed*). For a further discussion of the problems of interpreting how a message is meant, see the following section.

6.6 Discussion

The aim of this study was to attempt to find out if awareness of the recipient's presence affects messages and how. It was hypothesized that the sender's knowledge of the recipient's actual and immediate presence, visualized through the WebWho web page, would affect the topic of messages sent using the instant messaging system. Other factors, such as the electronic means of expression, user location and status, level of synchronicity, etc., were also taken to influence messages. The WebWho message log was compared to linguistic analysis of other modes of CMC (on email and chat). The log was also compared to corpora of traditionally written language and of spoken language in order to find out if there is any correspondence with findings from previous studies stating that the more synchronous and fast the text communication gets, the more the interactions carry features associated with spoken interaction. It had to be kept in mind, though, that the different forms of communication studied here are used for different activities.

While there are many factors that influence message composition, awareness of presence seems to be one of them, as well as the purpose of the interaction and the nature of the particular means of expression, it appears that there is evidence for the conclusion that instant messages sent via WebWho - with respect to topic – were indeed affected by the senders' awareness of both the receivers' physical presence and of virtual presence. Much of the results follow what could be anticipated intuitively, and confirm hypotheses about what the instant messaging system is used for. It was found to be used extensively for collaborating on mutual assignments and for coordinating social activities. 15.8% of the messages of our sample were task related, and 13.8% were related to social coordination.

Instant messaging using WebWho is also used for playful behavior; 18.2% of the messages in the sample were nonsense or mischief, and another 6.5% were mischief with sexual content. The latter two categories form 24.7% together - which makes it the largest category in the sample. Sending mischievous messages anonymously seems like a good way to avoid "being caught", for example. It is possible that text-only communication in combination with the virtual awareness of presence increases the temptation to toy with the possibilities of the messaging system. Awareness of physical presence might also make people send messages anonymously, perhaps to be able to view the effect of the text in person. Messages of the meta-comment-type often explored the possibilities of the system and reflect the users' unfamiliarity with what it can do and their curiosity to find out what they can actually use it for.

Eleven messages were sent but never received, because of the fact that the system does not allow users located in the computer lab to send messages to people outside the building. The users were able to receive messages from outside the building, though, and the message composition of those was characterized by comments on the communication *per se*. Example 52 below illustrates a message where someone tried to send messages to people outside the building (by manually altering CGI parameters in the URL string):

Example 52. 'Får du detta skall du inte vara glad, för
då kan jag terra dig järnet i
fortsättningen :)
*[You shouldn't be glad if you get this
because then I can harass you big time in
the future :)]*

So far, most results confirm the hypotheses and intuitive statements above. Some results, though, seem to be contrary to expectation. Most of what are intuitively contradictory results are found among the messages that were sent anonymously. At first, it seemed pointless and strange to discuss mutual assignments anonymously, or to send anonymous greetings, comment on the system and ask for feedback without telling who you are, receive encouragement from someone who is unwilling to give away their identity, or to try to coordinate your social interaction without giving deictic reference. Shared background knowledge, previous experience and contextual information, which are outside the messages themselves, surely account for relevant explanation in most cases. The receiver can probably often guess who the anonymous messages are from. Human interaction may not always be overtly rational or may have different goals than one expects, though, and everything cannot be accounted for using reason.

A special type case of deploying the anonymity feature is found in the several examples of anonymous messages that were signed with the sender's name, seemingly eliminating the whole idea of anonymous interaction. Why do people explicitly make the message anonymous, by consciously making it appear anonymous to the receiver, and then sign the actual message with their names in the end? Out of a total number of 1,067 anonymous messages, 19 messages were signed with the sender's name (1.8%), the majority of these were sent in the *Distant Anonymous* category (14 messages). In some cases, if one had had access to contextual cues and other background information, one could perhaps guess that the sender

wished to make the receiver believe that someone else sent it, like in this example:

*Example 53. tyvärr lunch stängt, åter 990402 Mvh
Förnamn Efternamn
[sorry closed for lunch, back April 2, 1999
Best regards First name Last name]*

This message was an ironic imitation of a message sent by the department secretary, who the sender feels takes unnecessarily long lunch breaks (the message was sent on February 2, 1999). In most cases the messages seem just ordinary, and probably reflect the same idea as noted in Nardi et al. (2000) - it is a nice way of saying hi without being too intrusive, even though making it anonymous seems like a strange thing to do. Here is an example:

*Example 54. Gomorron!/ Förnamn Efternamn
[Good morning!/ Last name]*

Another way of using the instant messaging system which seems counter-intuitive, was when people communicate anonymously about their mutual assignments. Unless the receiver really does know who it is from, it seems difficult to be able to collaborate on a mutual task, when the person who keeps sending you messages is anonymous.

*Example 55. nää nu får du lösa det här
[noo you've got to solve this one now]*

The messages of the rather large category that was labeled “Other”, which includes messages that did not fall into any other category, might in many cases be explained and filed away in the various categories respectively, if one had had access to the senders’ knowledge and context. Much of what we communicate, be it in speech or writing, would be pretty unintelligible if it were cut out from the situation and the context in which it occurred. Shared background knowledge of facts and situations, our experience of how the world works and how communication works give us clues as to how to interpret messages. Having had access to those particular situations or minds, for that matter, would certainly have cleared things up in the slightly messy “Other” category. For instance, lack of the above mentioned cues for guiding message interpretation makes it impossible to know for certain that all of the messages with sexual content were playful and

friendly. But this also holds true for any communication in any other situation: it is difficult to tell exactly what other people mean us to believe. This pragmatic issue is indeed intriguing, but was not the purpose of the present study and will thus have to be left for the time being.

6.7 Chapter summary

The study presented here deals with how awareness of presence affects content of instant messaging sent between students using WebWho, an easily accessible web-based awareness tool. WebWho visualizes where people are located in a large university computer lab and allows students to virtually locate one another and communicate via an instant messaging system. As WebWho is designed to be accessed through any web browser, it requires no programming skills or special software. It may also be used from outside the computer lab by students located elsewhere. The sender's user name is normally automatically added to the instant messages, but the messages can also be sent anonymously. A key question here is if the sender's conscious hiding of his or her identity seemed to be reflected in the content of anonymous messages, and how these differed from those with identified senders. Awareness of presence was one of several factors influencing message composition, both content and structural aspects. At this stage, the study has focused primarily on examining how different factors affect the content of the messages. Cross-analysis of messages for content in relation to parameters such as sender location (co-located, distributed and distant) and sender status (anonymous vs. identified) was done in order to find out whether awareness of presence seems to be an influencing factor. Computer-mediated communication (CMC) is often claimed to be a sort of hybrid between spoken and written interaction (cf. Ferrara, Brunner et al. 1991, and several others). Messages that were sent using the instant messaging tool in WebWho were compared with data from other types of CMC (email, chat) and also with corpora of spoken language and traditionally written language. The aim of the study was primarily to investigate awareness of presence affects on instant messaging, and only secondarily to investigate spoken vs. written features of the texts. Results show that awareness of both physical and virtual presence affects the content of the messages, and that these factors affect the text differently. Sender status, the nature of the computer-mediated medium, and the written mode shape the messages as well. Results show that the

students use the messaging system to support collaborative work and coordinate social activities, and extensively for playful behavior.

7 SMS – Text Messaging via Mobile Phones

7.1 Introduction

In this chapter the use and adaptation of written language to suit the conditions of text messaging via mobile phones – SMS (Short Message Service) – are analyzed. The main focus of the study is on factors that influence written language in these circumstances; what characteristics the resulting texts show in comparison with traditionally written language and spoken language. Comparisons of SMS texts with the characteristics of texts written in email, web chat and instant messaging are also made.

7.2 What is SMS?

The basic concept of cellular phones began in 1947 when researchers started developing crude mobile (car) phones⁶⁸, but it was not until around 1982 that mobile phones, as we know them, were first used. The mobile did not immediately become a device “for the masses”, and it was not widely spread in Sweden until the end of the 1990’s. It was at that time the mobile phone industry developed its Short Message Service (Crystal 2001), but it did not become popular until about 1998.

SMS, as defined within the GSM digital mobile phone standard, is a service which enables its users to send short text messages to one mobile phone from another, or to a mobile phone via the Internet. The text can

⁶⁸ <http://www.affordablephones.net/HistoryCellular.htm> (June 15, 2003)

be comprised of words or numbers or an alphanumeric combination. Each short message is up to 160 characters in length when Latin alphabets are used, and 70 characters in length when non-Latin alphabets such as Arabic and Chinese are used⁶⁹. Non-text based short messages (for example, in binary format) are also supported. These are used for ring tones and logo services, for instance⁷⁰.

Messages are typically sent from one mobile phone to another, although it is also possible to send from web based SMS-services on the Internet, and even over landline phones (Döring 2002a; 2002b).

The first SMS was sent in December 1992, from a PC (Personal Computer) via a Short Message Service Center (SMSC) to a mobile in the Vodaphone GSM network in the United Kingdom⁷¹. SMS is a “byproduct” of mobile systems using bandwidth that is left over, hence the limited number of characters. Shortis (2001) points out that phone companies introduced SMS initially as an afterthought; another gimmick to help beat the competition.

Today text messages are far more popular in Europe than making voice calls, and they far exceed voice calls by volume⁷². In May 2001, about 19 billion SMS messages were sent worldwide; 10 billion of these were sent in Europe. By the end of 2001, approximately 80% of the Swedish population owned a mobile telephone (Williamsson 2001). During 2001, more than 1 billion SMS messages were sent in Sweden; this is more than double the amount of messages that were sent during 2000, during which 473 million text messages were sent. 85% of these were sent from private accounts.

SMS was introduced in the United States during 2002. The reasons why it has not been widely used there until now are many, among them, the different standards among telecom operators. Operators have to assure SMS interoperability technology and sign interconnect agreements with their rivals to drive adoption of the messaging technology (Hoffman 2002).

A study of text messaging in Germany showed that during the year 2000, the average German mobile owner sent about 35 text messages per month (Döring 2002a). A survey spread across the whole of Sweden (Telia/Temo 2002) revealed that 24% of Swedish users send between 1-5 messages a day, and that 55% send between 1-5 messages per week. Döring’s study further shows that text messaging is used by young people; about 74% of 12-13 year olds owned a mobile phone in 2001, up from

⁶⁹ <http://www.mobilesms.com/> (June 15, 2003)

⁷⁰ <http://www.mobilesms.com/whatis.asp> (June 15, 2003)

⁷¹ http://www1.slb.com/smartcards/news/02/sct_smsanniv1902.htm (February 4, 2003)

⁷² <http://www.nokia.com/aboutnokia/compinfo/history.html> (October 3, 2002)

about 50% in 2000. People who use SMS intensively are mostly teenagers and young adults. These groups send more text messages on a daily basis than they send letters, email or make phone calls. More than 100 million text messages per month are sent in Germany (GSM Association, 2001).

Shortis (2001) argues that the resulting popularity of text messaging is a clear recognition of its social and technical capacity to stand in both where communication would have existed previously (e.g. making arrangements) and also to encourage interaction where it never used to exist at all, pre-face-to-face conversation primers (*“cu in a da common room.. cant wait to hear bout....”*) and post-face-to-face conversation interactions (*“that was lovely 2 c u... spk soon”*). Pre-face to face conversation primers, such as suggested by Shortis above, have perhaps not been as abundant before the introduction of mobile text messaging, but it seems likely that the same kind of communication could have been sent through email or over the phone. Whether SMS takes over functions of other modes of communication, or whether it complements other modes, will be discussed in Section 7.9.

7.3 Production and perception conditions of SMS

7.3.1 Text input

Most commonly (Döring 2002a; 2002b), text messages are created on the small keypad of the mobile phone and read as text on the tiny screen of the mobile. Different models of mobile phones allow for slight differences in the way text is entered. On older mobiles, text is entered with the so called “multi tap”, or “multi press” technique. The small keypad on the phone has ten keys (for the figures 0-9), as well as two additional keys⁷³. Each key for figures also contains 3-4 letters in alphabetical order. For example, the key for figure 4 also holds the letters *g,b,i*. In order to type the letter H, the key has to be pressed twice. To type the five-character word “hallå” (“hello”), the following keys have to be pressed: 42552. However, the keys have to be pressed 14 times, to get the correct sequence of letters (4425555522222). Obviously, this procedure makes text entry laborious and time consuming. In an attempt to overcome this obstacle, most

⁷³ On a mobile phone which is commonly used in Sweden, for example Nokia 3310 (the phone to the right in Figure 17), these keys contain the characters *+ and #^ respectively.

modern mobile phones today have some form of software⁷⁴ which predicts or guesses, words from the sequence of keys pressed (see below, Section 7.3.3).



Figure 17. Examples of types of mobile phones. Note the difference in screen size. (The keypad of the phone to the left is hidden under a lid). The pictures were downloaded from the photo galleries of the respective mobile phone producers' web pages: <www.sonyericsson.com> and <www.nokia.com>.

7.3.2 Screen size

Screen size also makes a difference in how the message is read, and thus has impact on which messages will be successful. Examples of this are messages which consist of text in combination with ASCII-characters. This type of message is typically a sort of “chain message”, in analogy with chain letters, which are passed on from one user to another. Figure 18 below shows an example of a “chain message” which is not possible to

⁷⁴ T9 (“Text on Nine Keys” <<http://www.t9.com/>>), is an example of prediction software used in many mobile phones today. Other kinds of predictive software are *Wordwise* and *Letterwise* Eatoni (<<http://www.eatoni.com>>).

view on a phone with the smallest type of screen. The ASCII-art in the “chain message” in the example below depicts a bear that is looking for the receiver in different directions, asking itself where the receiver is. The screen views in Figure 18 represent what is displayed on the screen of the mobile phone. To see the whole message on a mobile phone, one has to scroll between each of the screen views manually (in this case four screen views). A mobile with a smaller screen would cut the picture, and perhaps only show the upper part of the picture shown as Screen view 1, in Figure 18 below. The lower part of what is depicted as Screen view 1 in the figure below would show up as Screen view 2, making the message impossible to decipher.

Screen view 1
0"!0 ((T)) () Var är du?
Screen view 2
0"! "0 ((T)) () var...?
Screen view 3
0"!0 ((T)) () inte här...
Screen view 4
0"! "0 ((T)) () Aha, här är du!!

Figure 18. An example of a chain message sent via SMS. Screen size affects how a message is displayed, and may thus affect meaning. The text translates as follows. Screen 1: *where are you?* Screen 2: *where...?*, Screen 3: *not here...*, Screen 4: *Aha, here you are!!*

7.3.3 Text input using predictive software

Because simple person-to-person messaging is such an important component of total SMS traffic volumes, anything that simplifies message generation is an important enabler of SMS. Several types of software for predictive text entry have been developed. T9 is one of the most widely used types. Instead of entering text using the “multi-press”, or “multi-tap”, technique described above, with predictive text each key only has to be pressed once for each letter. The mobile phone has a built-in dictionary, and it will try to identify what is being written based on the words already stored in the phone's memory. The word shown on the screen changes after each keystroke, trying to predict the most likely word with the particular sequence of keys entered so far. If there are several possibilities when the word is typed, the user has the possibility to select the one that was intended. The predictive text input system does not accept slang or dialect unless taught by the user. As Kasesniemi and Rautiainen (2002) argue that, the mobile of a “lazy” user “speaks” only standard language with no personal tones. There is also a possibility to type in new words using the multi-tap technique to input them into the memory. Kasesniemi and Rautiainen ponder what effect predictive text entry will have on the development of the Finnish language, or what influence it will have on the personal style of their users. Is it the device or the user that will adapt?

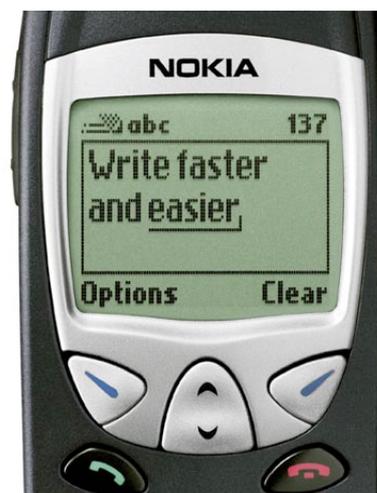


Figure 19. Predictive text entry displayed on the screen of a mobile phone.

The text input can be set to either multi-tap or predictive text entry mode. Once mastered, text input with predictive text entry saves time and effort

when composing messages. Depending on which system for typing one uses, the impact on the messages one writes is different. The software can also be set to either use capital letters or lower case letters only. It can also be set to take care of beginning each new sentence with a capital letter, without requiring the user to prompt it. The software judges a sentence to be finished when a full stop (and a blank space?) is entered, and begins the next word entered with a capital letter, assuming it to be the beginning of a new sentence. This may sometimes lead to mistakes and “typos” which are technology driven, as when a full stop is entered to mark an abbreviation that is not stored in the memory of the mobile. The software will regard this as the end of a sentence and begin the next word with a capital letter, even though that is in the middle of the intended sentence. We will come back to this issue when analyzing the SMS-corpus (Section 7.8.4).

Frequent use of text messaging eventually makes the user more skilled, and he or she may not perceive text entry as being cumbersome, as the beginner most often does.

7.3.4 Texting via web based SMS services

Short Message Service is also offered by a multitude of web-based providers. The prerequisites for composing and sending messages on the web are slightly different from creating messages on a mobile, and affect the composition of messages.

There are several services for composing and sending SMS messages on the Internet, three of the dozens of such sites are: <www.mtnsms.com>, <www.freesms.net>, <www.hooya.se> (see Figure 20 below). Many telecommunications companies offer the service to their subscribers. These services often require the user to become a member in order to take advantage of a personalized interface. The user is able to store the most frequently used telephone numbers in order to send messages conveniently without having to type the numbers every time a message is sent. Messages may often be stored in an out-box. Some providers allow replies to be sent to the user’s personalized in-box. The messages are typed on a computer keyboard, which allows for much faster and more effortless typing even than when using the text input system of newer models of mobile phones. Sending text messages via web based SMS services is often free of charge, in exchange for the user having to put up with being shown commercial ads. The charge for sending SMS from a mobile phone in Sweden is about 1.50 SEK (about €0.16, see below) per message. Message size is sometimes limited to 120 characters.

Short messages are not sent directly from sender to recipient, but always via an SMS Center instead (a store and forward service). Text messages are



Figure 20. An example of a webbased SMS-service. The service is often free of charge, and the text is entered on a computer keyboard. The number of characters per message varies between 120 and 160.

sent to an SMS-center (SMSC: Short Message Service Center) in the operator's network, and are delivered to the recipient's mobile when it is switched on and connected to the network next. Some mobile phones may be set to display confirmation of message delivery. Some web based SMS suppliers also provide this service. This means that unlike paging or email, users do not simply send a short message and hope that it gets delivered. Instead, the sender of the short message can receive a return message notifying them whether the short message has been delivered or not, depending on type of mobile phone as well as type of network. Whether the message is actually read, and not just received, is another issue, however.

7.4 Previous research on SMS

7.4.1 Accounts and popularity

Grinter and Eldridge (Grinter and Eldridge 2001) report that sending text messages in the UK is relatively expensive (between €0,05 and €0,20/message), while receiving is free of charge. The telecom companies in Sweden charge approximately 1.50 SEK (approximately €0,16/message) for each SMS sent, depending on what kind of account the sender holds. There are many different kinds of mobile phone accounts, and many teenagers hold so called “top-up cards”, or “pay-as-you-go”. Top-up cards are prepaid; charges for voice calling or sending SMS messages are deducted from the debit account. “Regular” accounts are charged like a phone bill for landlines; the user receives a bill once a month. During 2001, 50.4% of all GSM accounts were top-up-cards (Williamsson 2001). These are popularly given the teenagers by parents in order to avoid the nasty surprise of being sent a bill later, and to keep the users in control of the money spent on telecommunication. Studies have shown that teenagers in the UK are much more aware of the costs of their text messaging than teenagers in the US (Wakeford and Kotamradju 2002). In the US, “family plans” are often used. The popularity of text messaging among teenagers has raised interest in telecommunication companies, as the rather unexpected chances for profit proved to be successful. The prices for SMS-services may be lowered due to competition among companies, but despite the popularity of sending text messages, prices have not decreased very much since introduction of the service.

7.4.2 Use and function of SMS

Survey and content analysis in the studies of Schlobinski et al. (2001) and Döring (2002a; 2002b) show that interpersonal text messaging is used for the following social-communicative functions:

- Human relations (e.g. salutations, congratulations, invitations, jokes, etc.)
- Coordinating activities during a day (e.g. through reports on activities or the whereabouts of a person or activity)

- Coordination of practical assistance (through sending a shopping list or times for being picked up).
- Planning of meeting socially
- Maintenance of contact

First and foremost, text messaging is used for private communication between partners, friends, and family and loved ones. Dyadic communication or even informal group communication takes place between family and groups of friends. Döring (2002a) reports that text messaging is seldom used in formal contexts, such as within business organizations.

Accessibility, using SMS as a means for interaction in order to be in constant contact with other people, is notably important for users between 14 and 19 years of age. Mobile phone users who communicate with each other routinely use SMS. Eldridge and Grinter (2001; Grinter and Eldridge 2001) reported that typically, such person to person messaging is used to say hello or prompt someone for something or arrange a meeting or tell the receiver something. Such messages usually originate from the mobile phone keypad. Besides social-communicative functions, Döring (2002a; 2002b) suggests that the use of SMS also offers a sort of “intrapersonal bonus” (*“Intrapersonale Gratifikationen”*). People send SMS to stave off boredom or anxiety. This was also confirmed by Schlobinski et al. (2001).

Instrumental and expressive functions are often combined in SMS use. Döring gives the example of a message which contains a shopping list as well as a declaration of love (“1x Brot, wurst, 5Sack Äpfel I.L.D.”⁷⁵). This proves that contributions often have multifunctionality, and that a single contribution may serve several purposes in one (cf. Allwood 2000). Döring argues further that using non-face-to-face textual communication makes it easier to reveal emotions, which is exemplified in the flirtations known to be sent via SMS. Schlobinski et al. (2001) found that users regard SMS-flirtation to be easier than over the phone or through a traditional letter. SMS was regarded as less binding. According to a Swedish survey made in July-August 2002, 14% of those who participated (N=814) claimed that they use SMS as a means for sending flirtations messages⁷⁶.

Eldridge and Grinter (2001; Grinter and Eldridge 2001) report that the reasons for teenagers preferring to text one another are that it is quicker, cheaper, easy to use/more convenient than other communicative methods.

Eldridge and Grinter found minimal support for predictive typing technologies. The teenagers’ common use of abbreviations and shorthand

⁷⁵ “1x bread, sausage, 5sacks of apples I.D.L.” (Ich Liebe Dich = I love you)

⁷⁶ http://www.netsurvey.se/offices/Sweden/pdf/Digitala_relationer.pdf

made it barely usable in practice. Users have grown accustomed to the interface and have adapted it to their needs. Predictive typing can interfere with an expert's knowledge with the interface. Most teenagers in this study knew the interface so well that they did not even look at the screen as they typed their messages. Many teenagers have mastered the interface and have optimized their language. The restrictive length of text messages – rather than being a technological disadvantage - serves to expedite correspondence because short abrupt messages are perfectly acceptable, whereas phone conversation requires ritual openings. Texting allows teenagers to forego conversational conventions and makes the communication quicker by reducing the overall time spent on interaction. This was also supported by Döring (2002a; 2002b), who concluded that one may make oneself brief without fear of being perceived as short-spoken. One has to be brief not to go beyond the scope of the limited number of characters per message, because each message sent is charged at a relatively expensive rate. Döring also suggested that, in the long run, one is brief because the text input is so cumbersome. Finally, because most SMS communication is interpersonal communication between people who know each other, one can be brief when relying on pragmatic and shared background knowledge. The character limit of the messages themselves makes this terse and otherwise rude behavior completely acceptable. Text messaging can prevent the other person going “off topic” and making a conversation longer than planned. The character limit forces both sender and respondent to stick to the topic.

The absolute cost for text messaging is cheaper than voice calls (Eldridge and Grinter 2001; Grinter and Eldridge 2001), and the teenagers felt more in control of their costs with text messaging (all text messages are charged the same amount, whereas voice calls vary according to many factors: time of day, length of phone call, kind of operator, etc.). Furthermore, network operators typically charge the same to send a short message to someone in the same room as they do to someone traveling overseas with their mobile phone.

There are also pragmatic reasons for choosing text messaging over some other mode of communication: text messaging is discrete and can be created silently in public spaces or at times when other modes would be inconvenient. Texting supports quiet interactions, making it useful in locations that require discretion; such as in public situations where you do not want every one else to listen in, for example on the bus. Many mobiles have silent ways of alerting their owners to incoming communications, thus, text messaging gives users a way of communicating silently that does not disturb others. Mobile phones are often prohibited in classrooms, but

Texting is silent and can go unnoticed. Texting also allows for communication with people who have difficulty holding conversations. Because short messages are proactively delivered to mobile phones that are typically kept in the user's pocket and can be stored for later reference, SMS is often more convenient than email. The mobility of the mobile phone makes communication possible from anywhere; the communicators do not have to be at any particular location to be able to send and receive messages, as with email or instant messaging. Easier and more convenient also entail that it is possible to communicate even if it is too early or too late to call. Grinter and Eldridge (2001) also report that some teenagers find it easier to text someone who finds phone conversation difficult or when the receiver is someone the sender does not know well. Other examples may be instances when they find it difficult to talk, as when flirting.

To sum up, there are two basic ways of composing and sending text messages to mobile phones: either typing them on a mobile phone or typing them on a keyboard and sending them via the SMS-service provided by some web based service on the Internet. There are hypothetical implications on how messages are composed and when and how often messages are sent due to the effort of typing and the cost of sending affect. Studies have shown that text messaging is mainly used for maintaining relations between people, coordinating social activities. It is used foremost for private communication between partners, friends, and family and loved ones. Studies have shown that teenagers find SMS to be quicker, cheaper and more convenient to use than other modes of communication. We now turn to the linguistic characteristics of SMS communication, as suggested in previous studies.

7.5 Linguistic Characteristics of SMS

7.5.1 Reduced forms in SMS communication

Judging from the factors illustrated by Döring, production and perception conditions in combination with situational parameters (the technical restriction of 160 characters per message, each message sent is charged, text input is so cumbersome, interpersonal communication) both permit and force people to express themselves concisely in mobile text messaging.

It seems likely that we are to find enhanced use of lexical short forms and syntactic reductions that save keystrokes compared to longhand forms.

Döring claims that short forms in SMS are almost always the products of economy function. We have discussed the principle of economy, or principle of least effort above, in connection with Allwood's principle of rationality, which motivates communicators to seek pleasure and avoid pain, along with Clark and Brennan (1991), Dahl (2001), and Zipf (1932; 1949) in connection with an economy principle. During the text input process, keystrokes are saved, thus, also physical motor effort. The economy principle is not absolute, as we have seen, and may be overruled by other principles, for example the social, emotional effect of language play. Besides the economy function, Döring claims that short forms are also the products of a collective identity function. When a group of people shares special social knowledge they may understand and use short forms amongst themselves. The specialized use of short forms is thus an indicator of belonging in a community (Sveningsson 2001) and is a component of the group identity. Döring (2002a; 2002b) argues that it is often postulated that in SMS communication the short forms employed by heavy SMS users are not part of standard language. Mass media seems to love making claims about "the new SMS language" or "the language of teenagers", for example. Döring reports that *der Bild-Zeitung* launched a campaign in November 2000, "The great SMS action against long sentences". The typical SMS-style of phonetic value of syllable-acronyms was established. Since then, lists of acronyms ("SMS-lexica") have been found in newspaper reports and popular accounts. Shortis (2001) remarks that sometimes lists are compiled and published as "SMS guidebooks", such as the Christmas bestseller "WAN2TLK?: ltle bk of txt msgs"⁷⁷. Just as with the popular accounts and compilations of emoticons (also known as smileys) for use in Internet chat rooms or in email communication, one may anticipate that only a small portion of all these are actually used, as we shall see.

In these popular compilations, we find acronyms that we recognize from personal letters and notes, corporate correspondence, and online communication. Besides these, an array of genuinely new acronyms is actually found. Homonymous letter and phonetic value acronyms are also found. Through using a lot of SMS-short forms, the experienced SMS

⁷⁷ The Christmas bestseller (WAN2TLK?:ltle bk of txt msgs (2001) Michael O'Mara Books) criticized by Shortis is also translated into Swedish. However, it is only the full forms of the English abbreviations that are translated. No Swedish equivalents are given, and no references are made to actual use in authentic text messages.

users may send and receive messages that are only understood amongst themselves.

As has been discussed throughout this dissertation, the linguistic characteristics of text messaging are not determined by the medium alone, but to a large extent by the characteristics of the communicators, their interpersonal relationships and their reasons for communicating. As we have seen, Döring reports that text messaging is predominantly used in private interpersonal relationships. Features that are characteristic of spoken language, such as dialectal words, interjections and prosody, are verbalized and spelled out in SMS. The use of short forms may thus help to enhance the feeling of directness, effortlessness and natural spontaneity. Androutsopoulos and Schmidt (2001) report instances of a sort of “child language”, presumably caused by the directness and intimate relationship between communicators. This is not found in Döring’s corpus. Her findings show that text messaging is often used in situations with positive connotations, through informal written language, which allows for:

- Informal spoken language
- Broken native language (German)
- Teenager jargon
- Dialect
- Internet jargon

Both individual preferences and collective norms influence the messages. Despite the dominance of informality in the language of text messaging, the style is dependent of communicative context or situation. Döring points out formal or conceptual style texts in her corpus, as well as contributions in the style of elaborated literature critics. Döring found that the length of messages in the corpus varied greatly. The shortest messages were made up of only two characters (“ok”), and the longest used all the 160 characters allotted. The mean length was 13 words and 78 characters, respectively. The technical limitation of 160 characters was thus seldom fully exploited, and half of the space left unused.

7.5.2 Writing and graphic means

Just as in e-mail communication, there is a much higher tolerance for orthographic deviations from normative written language. Schlobinski et al. (2001) found markedly more errors from cursoriness (orthography, punctuation, syntax, etc.) in SMS messages, compared to traditional correspondence in which these seldom occur. Furthermore, they argue that

a maceration of formal and grammatical rules can be noted. There is an informal communication style by which the mobile user adapts his or her choices to the technical input procedure. Schlobinski et al. show that over 60% deviate from norm-oriented writing. About 43% of the messages were reported to be either typed in all capitals or typed all lower-case. Variations were messages consequently typed in all caps, consequently typed in lower-case, normative sentence initial capitalization, or partial all caps.

Schlobinski et al. (2001) conclude that the medium accounts for these variations (principle of economy from using the keypad) or having similar communicative functions as known from chat rooms, in which typing in all caps is a marker for emphasis or even shouting. Furthermore, Schlobinski et al. observed that punctuation marks are used to varying degrees. Pragmatic use of punctuation and normative patterns of use were observed: no punctuation marks at all or a sort of hybrid regarding all caps and lower-case, as well as normative punctuation. They suggest that use of capital letters in the middle of a word is to draw attention, as well as being a means of structuring the message. Schlobinski et al. argue that in some linguistic communities (*Sprachgemeinschaft*) codified orthographic means for certain needs of expression are insufficient. Instead of space between words, the beginnings of words are marked by a capital letter. The strategy to omit blank spaces between words saves space and enhances readability.

Almost 2/3^{rds} of the users followed normative punctuation in Schlobinski's et al. study, which they pointed out is amazing as different types of mobile phones demand a number of extra keystrokes to type a question mark or exclamation point, which is relatively costly both in physical and cognitive effort, as well as being time consuming.

The use of emoticons as a grapho-stylistic means also stresses the informal character of the communication. Emoticons are used as a meta-communicative means. The communicative function of emoticons is to compensate for the lack of verbal and non-verbal characteristics in spoken language. Schlobinski et al. argue that what is expressed in spoken language through discourse particles, interjections, prosody, facial expressions and gestures, will be expressed in chat using emoticons and other graphical means. As with abbreviations, even though lots of variants of emoticons are compiled in guidebooks or on the Internet, only a few of them are recurrently realized. Schlobinski et al (2001) report Haller's analysis of an SMS-corpus shows that the standard emoticon [:-)] is used 52%, and the "sad-emoticon" [:-(] 14%. Other variants are used more sporadically. Asterisks [*] are claimed to be another graphic means also used in chat communication to show actions or emotions. Schlobinski et al. claim that

repetitive keys as a means to show emphasis appear much more seldom in SMS than in chat, but is found in the SMS-corpus. Androutsopoulos & Schmidt (2001) point out that the “grapheme reiterations” found in SMS created in web-based services tends to be avoided in SMS created on the keypad of a mobile because of the laborious text input.

7.5.3 Means to reduce text in SMS

Syntactical and lexical reductions, as well as graphical techniques are ways to reduce time, effort and keystrokes when producing text messages. This section will investigate syntactical and lexical reductions.

Syntactical short forms

Döring’s (2002a) findings of syntactical reductions confirms the findings of Androutsopoulos and Schmidt (2001):

Table 33. The most common types of syntactical reduction in SMS, shown in German studies of mobile text messaging.

<i>Type of Reduction</i>	<i>Example</i>
<i>Deletion of subject (especially subject pronoun)</i>	[Ich] Komme später Heim...!
<i>Deletion of preposition, article and possessive pronoun</i>	Weißt du was [der] Eintritt kostet
<i>Deletion of copula-, auxiliary- or modal verbs (+XP)</i>	[Bist du] Schon wieder zurück aus [Ø] Urlaub?
<i>Deletion of verb and subject pronoun; Telegram style</i>	[Hast du] Lust, dann komm vorbei?

Androutsopoulos & Schmidt suggest that the most common type of syntactical reduction is the deletion of the subject pronoun. The study of Schlobinski et al. shows the same pattern. The categories of syntactical reduction proposed by Döring, and shown in Table 33 above, are overlapping and not mutually exhaustive; the categories deletion of article and possessive pronoun and deletion of preposition or the compound preposition + article. Döring exemplifies these categories with the same messages. Deletion of verb and subject pronoun naturally seem to give an effect of telegram style, but the latter is given a separate category. The syntactical structure of SMS messages is similar to telegrams in many ways. Ahlsén (1993), for example, studied aphasics with agrammatism, and concluded that they often use a kind of telegraphic style to adapt to and overcome obstacles in their speech. Telegraphic style is characterized by

short phrases with no elaborated grammar and omission of function words and other grammatical morphemes. Ahlsén lists four types of phrase patterns in telegrams:

- Free standing verb and/or adjective phrases
- Free-standing noun phrases
- Noun phrase + verb and/or adjective phrase
- Imperatives

The same kind of pattern is noted in the German studies of SMS messages, noted above, even though the reasons for it are different from that of the aphasics.

Syntactical reduction is well represented in Döring's corpus: 38% of all text messages have at least one word deleted. The most common word type to delete is subject pronoun, and especially the first person singular pronoun. Pronoun ellipsis is also found by Androutsopoulos and Schmidt (2001). Even though pronoun ellipsis is the most common syntactical reduction, the first person singular pronoun (*Ich*=I) turned out to be the most frequent word in the corpus. The second person singular pronoun (*Du*=you) proved to be the second most frequent word, confirming findings in studies of email, chat and instant messaging (Hård af Segerstad 2000b; 2002). Döring explains this with the interpersonal character of the text messages, while Hård af Segerstad additionally relates this as reflecting the activity that the communicators are engaged in (cf. Allwood 1976; 2000).

Lexical short forms

Döring points out that the lack of a proper taxonomy for lexical short forms is problematic. She defines acronyms and abbreviations as:

- *Acronyms* are word forms that are made up from the initial (in some cases the first two or even three) letters in a sequence of words. These will be pronounced letter wise or from the prosodic value.
- *Abbreviations* are written short forms of words, typically built from the initial letter and pronounced in their full form when read out.

Döring divides the category "acronyms" into *letter acronyms* and *phonetic value acronyms*. The category "abbreviations" is divided into *conventional abbreviations* and *unconventional, ad-hoc abbreviations*. Letter acronyms and ad-

hoc abbreviations proved to be most commonly used of the four types. Schlobinski et al. (2001) term the latter type of abbreviations *ad-hoc formations*.

Table 34. Döring's taxonomy of acronyms and abbreviations. Examples are taken from Döring (2002a) and Schlobinski et al. (2001).

<i>Acronyms</i>	<i>Abbreviations</i>
<i>Letter acronyms</i> <i>Examples are:</i> <i>SMS, CD, PS, HDL, I.L.D.</i>	Conventional abbreviations Examples are: inkl., bzw., Nr.
<i>Phonetic value acronyms</i> <i>Examples are:</i> <i>RAM, SIM</i>	Unconventional, ad-hoc abbreviations (or Ad-hoc formations) Examples are: he[heute], i[ich] e [ein]), f [für], ü-nächste [übernächste], zw.durch [zwischen]

Schlobinski et al. argue that Ad-hoc formations are abbreviations that have been conventionalized, though these generally have not been accepted in common language use. They claim that this type is thus specific to SMS communication. Examples are *GuK* (*Gruß und Kuß*), *cu* (derived from English *see you*, and well known from chat communication), or *hdgdl* (*hab dich ganz doll lieb*). Moreover, they point out that there is one abbreviation among the conventionalized, not generally accepted abbreviations which is particularly common, namely: *hdl* (*habe dich lieb*). It occurs in a range of variants: *hdal* (*habe dich auch lieb*), *hdgdl* (*habe dich ganz doll lieb*), *hdsmdl* (*habe dich sebr mega doll lieb*), etc. These abbreviations are neither found in dictionaries nor in net lexica, they are not found in any studies of chat or email communication, and it seems that these are abbreviations that are specific to SMS communication. Schlobinski et al. found that SMS users employ these freely, and the “rules” for how to use of them are based on shared lexical and syntactical knowledge.

Schlobinski et al. show an abundance of abbreviated word forms without changing the semantic content in SMS communication. They argue that the high number is due to the conditions of the mode, namely the limited number of characters per message. Döring (2002a; 2002b) report that out of 1000 words, 30 were acronyms and abbreviations. This is in contrast with Crystal (2001), who overestimate the rate of abbreviation in predicting that an even more abbreviated language than the language of chat groups is evolving in text messaging. According to Crystal this process is due to the challenge of the small screen, the limited character space, as

well as the small keypad. Crystal bases his estimations on purely anecdotal evidence and the accounts of popular SMS guidebooks.

Word initial and word final reductions

Schlobinski et al. argue that assimilations and reductions are influenced by colloquial spoken language. Word final reduction is most common; with few exceptions –e and –t are reduced, foremost in first person singular. This is a feature typical of spoken language which is transferred to written language. Word initial reduction is found to be used predominantly for function words, prepositions, articles, as well as personal pronouns. Initial reduction, almost without exception, involves the dropping of the indefinite article. The omission of an article is sometimes marked with an apostrophe [']. Another type of word initial reduction that is also typically found in spoken language, is found in word forms such as *raus*, *rein* etc. instead of *heraus*, *herein*. Schlobinski et al. list several assimilated forms, perhaps better called contracted forms:

- Preposition and article: *aufm*, *aufn*, *aufs*, and *fürs*
- Pronouns (2nd person singular): *haste* -> *hast du*, *bissu* -> *bist du*
- Particle and verb: *hab'n*, *isn*
- Syllable reduction: infinitive suffix : *müssn* -> *müssen*

They claim that, like in spoken language and in analogy with reductions of English words found in chat, r-dropping is frequent: *supa* -> *super*, *wieda* -> *wieder*, *leida* -> *leider*.

Functions of reduced forms in SMS communication

It is important to point out that in SMS communication, as in chat and email communication, English abbreviations are neither dominant nor are they a threat to the German language. Syntactical reduction, dominated by pronoun ellipsis, makes a relatively small effect on number of keystrokes. Thus, Döring argues, eliminating first person singular pronoun (Ich=I) plus blank space, only saves 4 characters. If multi-press technique is used, 9 keystrokes are saved (444222440). If predictive text software is used, 4 keystrokes are saved (4240). Much more effective economizing is made by lexical reduction. A saving strategy specific to SMS communication is the use of ad-hoc abbreviations. Döring gives an example of abbreviating the

word “vielleicht” (=maybe) into the word initial letter “v”, which saves 8 characters.

The production of complex acronyms such as *ru2cnml8r* (=are you two seeing me later) (exemplified by (Crystal 2001) demands both time, skill and physical as well as cognitive effort to translate, not to mention interpreting it. The same thing holds for the very complex emoticons found in compiled lists and lexica on the Internet and in guidebooks. Hård af Segerstad (2000b) points out that the gains one intends to make by abbreviating words and inserting emoticons – namely, to save time, effort and space as well as to convey nonverbal information – are lost if the abbreviation or emoticon is so intricate that the receiver is unable to decipher it. The whole point of taking the trouble to reduce keystrokes is then lost and nothing won, as one will have to explain anyway. This fact does not exclude that the economy principle is sometimes overruled by other factors, such as language play. When communication becomes too complex, it might be easier to change communication medium, which is sometimes explicitly mentioned in text messaging.

Schlobinski et al. point out further that there is no degradation of language, which conservative language users (*Sprachpuristen*) always seem to believe, but rather writing adequately has become an art. The use of SMS and its influence on the language and written expression of teenagers seems to have been discussed in Finnish media (Kasesniemi and Rautiainen 2002). As SMS communication does not rely on traditional grammar or punctuation required in texts written for school, Finnish teachers have been worried about the negative effects. Teens write messages in all lower case or all capitals. Words are shortened. Inflectional endings characteristic of the Finnish language and spaces between words can be left out. The unique writing style provides opportunities for creativity. Kasesniemi and Rautiainen report that teachers presume that these elements will be transferred from text messages to more formal texts and worry about the consequences. Similar questions have been raised about the influence of e-mail. Schlobinski et al. argue that if one were to believe popular science publications, abbreviations based on English words are the feature most characteristic of SMS communication. Their findings, though, show relatively few English abbreviations, the most common ones being *SMS*, *ok* and *cu*. Of all abbreviations, 88% were based on German and 12% were based on English words. The matter of the supposedly bad influence of text messaging on written language in general will be discussed both at the end of the present study, and in general in the concluding discussion in the final chapter below.

7.6 The Present Study

The present study is the last in a succession of four case studies that investigate how written language is used and adapted to the particular settings of a number of modes of CMC. All case studies have distinctive conditions, and may thus be expected to show distinctive features of language use. The present study investigates the technically asynchronous mode of mobile text messaging. As we have seen, production and perception conditions not only dictate that the messages be brief, but also enable users to be brief without being regarded as impolite. Similar to the case studies reported above, a corpus, in this case of text messages, will be analyzed and findings will be related to the variables of synchronicity, means of expression and situation.

7.6.1 Research questions

To answer the question of how written language is used in SMS is simple enough once one has a corpus to investigate. To answer the questions concerning why it is used this way requires that a number of factors be taken into consideration. Parameters related to Means of expression are; different kinds of text input, type of mobile phone used, use of predictive text entry or not, and computer keyboard when utilizing web based SMS-services. Sending messages either via mobile phone or SMS-service on the net is linked with both physical and cognitive effort as well as monetary cost, the phone account might be charged or in some cases it is free of charge. The latter also is dependent on type of Internet connection. Screen size of the receiving mobile phone is also hypothesized to affect message composition.

The questions related to the variables *Synchronicity* and *Situation*, what SMS is used for and why texting is used instead some other means of expression such as voice calling or meeting in person, require both an analysis of the content of messages and interviews with users. Additional parameters related to *Situation*, such as relationship among communicators, goal of interaction, where and when communication takes place, also have to be considered.

Why is this interesting?

This analysis originates in the idea that an analysis of mobile text messaging will reveal results that are interesting in themselves, namely showing how

written language is used and adapted to setting specific conditions, as well as revealing ways in which humans can adapt their language behavior according to the constraints and enablements of a given situation.

Texting is an important means of communication today, both for private interaction and in business relations. Many people are using it and it forms an important part of everyday life. The supposition that features of text messaging, such as the brevity of messages, the informality of the non-standard language, acceptance of grammatical errors or spelling mistakes, and so on, might be transferred into other domains of written language is something that people sometimes fear. This study investigates language use, and will regard the results as a variety of language use that is optimized according to the conditions of communication. As teenagers are the largest group to use SMS, a future project will endeavor to investigate young people's use of written language in other domains, such as essay writing in school. If language use is transferred from chat rooms and instant messaging into mobile text messaging, it might also be transferred into other domains where written language is used. If that is the case, we are witnessing a change of norms, and also language change, which means that the teachers' presumptions reported by Kasesniemi and Rautiainen (2002) would be proven right.

7.7 Methods for Data Collection and Analyses

7.7.1 A note on the choice of material and its implications on the results

Despite the fact that SMS is extremely popular, it proved difficult to get hold of material. It would have been ideal if the corpus had represented language use evenly spread across gender and all age groups, as well as covering geographical areas and social status. One of the reasons for why people feel it is hard to part with their messages may be that SMS is experienced as being even more private than email. Most messages are sent between friends who know each other well; they communicate about things that they do not want any one else to read, even though the content of most messages is seemingly trivial about everyday matters. Mobile phones are also experienced as being personal gadgets. They are typically carried in a pocket or a bag, most of the time close to the body. They are personalized with ring tones and shells that can be changed. All of these

matters are interesting in their own right, and would be interesting to investigate further. As stated, the corpus does not represent as broad a user group as would have been ideal, and the results may not be generalized to cover language use in all mobile text messaging. The corpus is large enough however - 1,152 messages and a total of 17,024 words - to give interesting insights into how written language may be used in SMS.

7.7.2 Data collection

Three methods for collecting data for a study of text messages were used in this study: a web based questionnaire, user diaries and forwarded methods, and messages from friends and family. Each method has its merits and drawbacks, and the aim was that they should complement each other.

The web based questionnaire

The first method for data collection is from a web site. A web page consisting of a questionnaire was constructed, on which anonymous informants would access and answer questions and share a text message they had sent either by mobile phone or a web site that provides SMS service. The messages were copied character-by-character exactly as typed in the original and submitted. The questions were mainly answered by selecting one of a number of answers on a pull down menu. Although these were kept to a minimum in order to lower the time and effort required by informants, some questions required a written answer, in particular the copying of an SMS. The data from this web-based form were entered automatically into a database. The informants were recruited and invited to join the study through daily newspapers in which the researcher was interviewed, as well as through a link on the researcher's web site. All entries in the web-based data questionnaire were anonymous, and all informants remained completely anonymous to the researcher.

There are pitfalls of copying and of selection. Copying requires retyping on a keyboard a text that is read on the small screen of a mobile phone. Caveats in the case of copying are that it is very easy to make typos or to simply copy the original incorrectly. A misspelling in the original is easily corrected subconsciously when copying, or an extra exclamation mark might be omitted or just as easily added. Not all text messages were composed on a mobile phone, but were composed typing on a keyboard and sent via the SMS service provided by a web site. Copying the latter type of text messages involves copy-and-paste, and makes for less chances

of incorrect copying. Still, the problem of selection remains. The informants selected the messages that they wished the researcher to get or omitted messages, for various possible reasons; the informant might not want the researcher to read private messages and withholds these, or the informant wants the researcher to read certain messages because the content is either neutral or special in a shocking way that the informant wants to display. Also, the anonymity of the web form cannot guarantee the authenticity of the message. There were a few entries that seemed to actually be filled in by someone trying to be annoying; making things up for the fun of it. Those entries were ignored.

User diaries and forwarded messages

As mentioned above, four informants (two male and two female, between 12 and 25 years of age) were recruited to participate in the research project. They were asked to keep a user diary in which they were to record information about each text message that they sent and received during a period of one week, or longer if they so wished. Information about each informant was gathered before embarking on the project. Information such as name, age, sex, mobile phone number, type of mobile phone (different models have different methods for typing messages, affecting ease of effort). When forwarding messages to the researcher's mobile phone, information about the sender's phone number (corresponding to the information entered in the address book of the mobile phone) showed from whom each message was forwarded. The messages were transferred into a database on a computer by connecting the mobile phone to the computer by a serial cable along with the use of software, so as to eliminate further copying and thus possible typos by the researcher. It was pointed out that the messages that were forwarded, and everything that informants told the researcher in the interviews or wrote in the diaries, was to be dealt with in the strictest confidentiality, and that they would remain anonymous to everyone but the researcher.

In order to minimize effort, each informant was provided with a small note pad consisting of preformatted pages with questions to be filled in. The note pad was small enough to be carried along in a bag during the day in order to be filled in directly after sending or receiving each message. Another method for submission was to save all messages on the mobile phone and go through them at the end of the day and fill in the questions, while memory of each message sent or received during the day was still fresh. One page of short questions for each message sent or received.

According to findings prior to commencement of research, the researcher assumed that the number of messages would range from approximately 6-10 messages a day.

During the period of keeping the diary, the users were asked to forward text messages that they wrote and sent via their mobile phone or via the SMS service of some web site, to the researchers mobile phone. Forwarding already composed messages would minimize potential errors due to copying messages by typing them all over again, the method for input on the web-based form.. It was pointed out that informants were free to either send all messages of their communication, or to select those messages they felt comfortable sharing with the researcher. The problem of selection remains the same as with the web based form, but was minimized as much as possible due to the researcher's personal contact with each informant. It was also pointed out that the messages were to be handled with the strictest confidentiality. Nobody except the researcher was to read the messages, especially not informants' parents. Informants were paid for each SMS that they forwarded the researcher by topping up their prepaid cards or having the compensation paid to a bank account; the sum was meant to cover the cost for each participant's forwarded messages.

In the next stage, informants forwarded text messages to the researcher's mobile phone during a period of time, both messages that they received⁷⁸ and that they sent themselves. The forwarded messages were transferred to be stored in a database, with the use of a serial cable between the phone and a computer. It was pointed out that they naturally had the choice of forwarding each and every message if they so chose, or of selecting messages that they felt comfortable sharing. The possible reasons for not sharing the messages were to be followed up in the interviews.

The third stage consisted of interviews with each informant who kept a diary and forwarded messages. The interviews were semi-structured, consisting of a number of questions to guide the talk and of free associations and informal conversation in which the informants had the chance to comment and add remarks on both the research itself and about their own communication via mobile phones. Each interview lasted about 1-1.5 hours, and was recorded with the informants' consent. The chief reason for recording the interviews was to keep the records as verbal notes.

⁷⁸ To forward messages that they had received, they had to have the original senders' permission.

Messages from friends and family

The third method of data collection was to ask friends and family to part with their messages during a period of time. 16 informants (8 female, 8 male) contributed to the corpus by sending 788 messages. Messages were either sent directly to the researcher's mobile phone and transferred to computer readable format with the serial cable, or copied character-by-character into a text document in a word processor.

All in all, the three methods of data collection resulted in a corpus consisting of 1,152 messages with 17,024 tokens (4,045 types).

Table 35. Number of messages and number of words from each method of collection.

	<i>Web based questionnaire</i>	<i>Informants' forwarded messages</i>	<i>Friends and family</i>
<i>No. messages</i>	112	252	788
<i>No. words</i>	1,457	2,512	13,055

7.7.3 Methods for analyzing the data

The corpus was stored in machine readable format, and separated into two categories according to text input technique: messages composed on the keypad of a mobile phone, and messages composed on computer keyboard when sending messages via a web based SMS service.

The same tool for automatic analysis of transcriptions (TraSA), described in the previous case studies, was used to make automated investigation. The following measures were investigated:

- Total number of messages
- Total number of words
- Vocabulary richness
- Mean length per message (=Mean Length of Contribution)
- Word frequency

The frequency lists were manually checked for occurrence of abbreviations (both established and new), complex punctuation expressions, emoticons, asterisks and ASCII-art.

The corpus of messages was checked manually for occurrences of syntactical and lexical reductions such as deletion of subject pronoun,

deletion of verb phrase (e.g. auxiliary verb, copula verb, subject pronoun, preposition), features that are typically associated with spoken interaction. The example below was taken from the SMS corpus and exemplifies a message in which the verb phrase (in square brackets) was omitted.

Example 56. [Ska vi ses på] Café Japan [klockan] 19?
 [Can we meet at] Café Japan 19 [o'clock]?

7.8 Results

The aim of this study is to analyze SMS messages in order to answer the following main questions:

- How is Swedish written language used in SMS?
- Which syntactical and lexical short forms are used?
- Are these short forms based on Swedish or any other language?

The analysis of messages collected for the present study reveals that Swedish SMS communicators use the same basic types of syntactical and lexical reductions as were found in the studies of Döring (2002a; 2002b) and Schlobinski et al. (2001). Table 36 below shows a comparison of quantitative results from the data gathered for the present investigation and the German corpus in Döring's study. It reveals a mean length of Swedish messages of 14.77 words/message. These are longer (consisting of a larger number of words) than the messages in the German study, which averaged 13 words/message.

Table 36. Comparison between Swedish and German SMS studies.

	<i>Swedish SMS corpus</i>	<i>German SMS corpus</i>
<i>No messages</i>	1,152	1,000
<i>No words</i>	17,024	13,357
<i>Mean length of message</i>	14.77 words/mess.	13 words/mess.
<i>Mean no. characters/mess.</i>	64 char./mess.	78 char./mess.

Results support Döring's finding that the length of messages varied strongly. The shortest message in the present study consisted of only one character: "a", which probably in this case represents the Swedish spoken variant of *ja* [yes]). The longest messages actually used more than the normal technical limit of 160 characters. The two longest used 165 and 163

characters respectively. These were “linked messages”, they were typed on a phone that allows the sender to use 459 characters⁷⁹, but the message is divided into shorter pieces when sending. The other four messages that exceeded the 160-character limit were sent via web-based SMS services. Some web-based SMS services allow for more than 160 characters and some allow for less.

Swedish SMS messages used fewer characters per message than the German: the mean number of characters per message was 64 characters, compared to 78 characters in the German corpus. Thus, Swedish users utilized 40% of the message size (160 characters), while German messages used about 49%. These results suggest that Swedish users abbreviate words more frequently to reduce length of words. Variation might also be due to differences in language typological characteristics.

7.8.1 A comparison of word frequency between modes of communication

As has been discussed above, a comparison of word frequency across modes of communication may reveal interesting insights to what determines language use. The four modes that were analyzed in the present dissertation are all monomodal. Among the most frequent tokens were features characteristic of written means of expression, for example punctuation marks. Email and SMS are asynchronous modes of communication, while web chat and instant messaging (IM) are synchronous. All modes except web chat are one-to-one messaging. In SMS communication and in the particular IM-system that was investigated in the present dissertation, the communicators knew each other prior to communicating.

Table 37 below illustrates a comparison of the most frequent words in four modes of CMC. All modes, except email, had the words *du* [you] and *jag* [I] among the most frequent. As was discussed, the email messages investigated in the present dissertation were sent to unknown authorities, a fact that explains the lack of second person pronoun. IM and SMS share nine out of the ten most frequent words, possibly indicating a similar sort of relationship between communicators and that the modes were used for similar activities. IM and SMS share eight out of ten with the web chat, and five out of ten with the email corpus.

⁷⁹ An example of such a phone is Nokia 3330, which has a limit of 459 characters when sending SMS messages.

Table 37. Comparison of most frequent words across four modes of computer-mediated communication.

	<i>Email</i>	<i>Web chat</i>	<i>IM</i>	<i>SMS</i>
1	.	du	.	.
2	i	?	!	!
3	på	jag	,	?
4	att	,	?	,
5	och	är	du	du
6	jag	.	det	på
7	är	det	på	i
8	som	...	är	jag
9	till	som	och	det
10	det	!	jag	är

7.8.2 Linguistic features in the Swedish SMS corpus

A comparison of types of lexical and syntactical reductions reveal that most types that were presented by Döring (2002a) and Schlobinski et al. (2001) also appear in the Swedish corpus. Linguistic features that were found to be used frequently are listed in Table 38; these will be exemplified and discussed below.

Table 38. Linguistic features of SMS.

<i>Category</i>	<i>Feature</i>
<i>Punctuation</i>	Omitting punctuation Unconventional punctuation Omitting blank space
<i>Spelling</i>	Mispredictions Spoken-like spelling Split compounds Consonant writing Conventional abbreviations Unconventional abbreviations Either all capitals or all lower-case Exchange long words for shorter
<i>Grammar</i>	Omission of Subject pronoun Omission of VP (copula, auxiliary, or modal verb + preposition) Omission of Article, Preposition, Possessive pronoun
<i>Graphical (non-alphabetical) means</i>	Emoticons Asterisks Symbol replacing word

7.8.3 Punctuation

Omitting punctuation

By omitting punctuation a user saves the time and effort it takes to type those characters (period, comma, etc). Omitting punctuation also saves keystrokes, which could be important when the message size is restricted by technical limitations. Results show that the strategy of saving space and effort by omitting punctuation may also be used in short messages that do not threaten to exceed the restricted number of characters. Example 57 below gives an illustration of such a case.

Example 57. Ge mig ditt nummer har det inte längre så
ringer ja ikväll ska spela fotboll
[Give me your number dont have it anymore
then I call tonight will play football]

Unconventional punctuation

Unconventional punctuation such as excessive use of exclamation points or question marks was not as frequent in the SMS corpus, compared to web chat, for example. This was probably due to production conditions: it is comparably awkward to type on the tiny keypad of a mobile phone. Additionally, in order to insert a symbol other than the alphabetical ones, many phone models either require the sender to press the particular key several times or to leave the message window and choose to insert symbol from a menu in the phone's system. Never the less, messages displaying unconventional use of punctuation occur in the SMS data. Example 58 below was typed on the keypad of a mobile phone.

Example 58. FAAAN HUR KAN JAG MISSA DIN FÖDELSEDAG !
GRATTIS!!!! kramar x
*[DAAAAAMN HOW COULD I MISS YOUR BIRTHDAY !
CONGRATULATIONS!!!! hugs x]*

It seems likely that Example 59 was typed on a computer keyboard and sent via a web-based SMS service.

Example 59. Jeg vet inte..e d så lr??? Aja..jag ska iaf
ta en trip till sthlm snart..see ya
darling..
*[I don't know.. is that right??? Alright..I
take a trip to sthlm [Stockholm) soon
anyway.. see ya darling..]*

Omitting space between words

Similar to the strategy to omit punctuation, by omitting space between words the user saves keystrokes as well as time and effort. In some cases it was not necessary to save space, but omitting it renders an interesting tone to the message. Example 60 below was typed on the keypad of a mobile phone.

Example 60. KOMINIAFFÄREN
[COMEINTOTHESTORE]

Example 61 was typed on a computer keyboard and sent via a web-based SMS service. The latter utilized all 160 characters, and seems to be a good example of linguistic awareness: by capitalizing each word readability was increased. Moreover, given its content, it seems to render a sort of pleading touch to the message. This technique to produce text would be very difficult and time consuming if typed on a mobile phone, and no equivalents were found that were produced on a keypad.

Example 61. UtanBaraÖkatOchJagÄlskarDigMer&MerFastänDuÄ
rSurPåMig.MenVillDuInteHaMigSå..SåBlirJagJä
ttelelsen.DuBehöverInteFlyttaUppHitKäraVän.
DuÄrNogBaraRädd.
[ButJustExpandedAndILoveYouMore&MoreEvenTho
ughYouAreMadAtMe.ButIfYouDon'tWantMeThen..T
henI'llBeVerySad.YouDon'tHaveToMoveUppHereM
yDear.You'reProbablyJustAfraid.]

7.8.4 Spelling

Mispredictions and Typos

Spelling in mobile text messaging seemed to be very much dependent on whether the sender used predictive text entry or not. In most cases it was not stated whether or not the phone had this software or whether it was actually employed. Predictive text entry compares the entered sequence of characters with a lexicon stored in the software. It predicts the most likely word or words with that particular sequence, and unless checked, the most frequent word will be the one that appears on the screen. The software does not consider syntactical or semantic context, and is not tailored to suit everyday informal conversation, which results prove to be the most common. Thus, many strange software-driven "typos" may appear in mobile text messaging. Example 62 illustrates a misprediction in which the predictive software judged på [on] to be more frequent than så [so, so that]. This misprediction was probably not too difficult to interpret correctly using pragmatic knowledge. For a discussion of mispredictions, see Section 7.8.4.

Example 62. De e **på** [så] ibland bara.
[Mispredicted output: It just is **on**
sometimes
Intended output: It just is **like that**
sometimes]

Example 63 below is an illustration of a misprediction of a colloquial phrase "Gott mos", literally: good mashed potatoes, meaning "good stuff". The literal meaning of which is "Good mash" and the intended meaning is something like "Good stuff") The predictive software appears to judge the following to be more likely and frequent:

Example 63. Gott **öms**

The word "Öms" does not really mean anything in Swedish, which renders it somewhat untranslatable. Anecdotal evidence from informers suggested that as a result of mispredictions of a common phrase such as the one in Example 63, as a joke people might start using the misprediction literally, instead of the original (cf. Cherny 1999).

Spoken-like, or unconventional spelling

In many cases unconventional spelling, or spelling which imitates the phonetic value of speech, saves keystrokes and time and effort. One does not have to be as explicit and careful in spoken interaction as in writing. Sometimes, though, unconventional spelling results in the same number of keystrokes as the normative spelling, sometimes in even more keystrokes or more effort spent. This proves that the economy principle is not absolute and that what is considered rational behavior is instrumental; it is rational for the purpose it serves. The word *åxå* [också=too], in Example 64, saved two keystrokes compared to normative spelling. At the same time it brings an informal and friendly tone to the message.

Example 64. Tjena! tillsammans e rätt tuff! ska åxå köpa den nån gång....snart!! jo ja har d åxå trevligt spelar golf d går bra! men du får ha d så bra!! kramas!!
[Hi! together⁸⁰ is pretty cool! will also but it some time....soon!! yes I'm having a good time too playing golf it's going fine! but you'll have to take care!! hugging!!]

Colloquial words and dialectal pronunciation may be spelled out in SMS texts. Choice of words and how to render them in writing seems to be strongly dependent on the relationship between communicators. It possibly acts as in-group markers and connects with the interlocutor's shared background knowledge. It might also function as a pragmatic strategy. The message in Example 65 might well be interpreted as a way of reminding the receiver that he or she ought to get back with some information soon.

Example 65. **Höllö höllö** [hallå hallå]! **Bler** [blir=going to be] det **nön** [någon=any] beachvolley **elle** [eller=or]?
[Hullo Hullo[Hello Hello]! Is there gunna [going to be] be any beach volley or what?]

Another type of unconventional spelling, or rather a feature that is unconventional to spell out in writing in the first place, is found when OCM (Own Communication Management) features from spoken language appear in mobile text messaging. Results from web chat and IM also showed these features.

⁸⁰ "Tillsammans" ("Together") is a movie.

Example 66. **Hmm**, kanske det!? Vi hör*s* i **morr'n**
[morgon=morning] :-)
[**Hmm**, may be!? Let's get in touch tmorro
(tomorrow) :-)]

Example 66 was opened with an OCM feature and closed with a conscious marking of a word rendered in its spoken form.

Split compounds

The SMS log showed a few cases where the predictive text entry software seemed to be responsible for splitting compound words. Whether this might exedite a tendency in progress remains to be seen. Splitting compounds in many cases entails a semantic change, but the example below is not as severe.

Example 67. HAJ! JAG HAR KÖPT ENA NY LUR. **JÄTTE ROLIG!**
[HI! I'VE BOUGHT A NEW PHONE. **REALLY**
GREAT!]

The illustration of a split compound provided in example 67 does not really come out in translation, as it is not the norm in English writing to compound nouns in the same manner as Swedish does.

Exchange long words for a shorter

Yet another strategy to save space was to exchange a long word for a shorter, even though the shorter word is not as frequently used as the long one. The word *ej* has a slightly archaic stylistic tone, and is not used as often as the everyday *inte*. Both are equivalents to the English *not*.

Example 68. japp men det syns **ej** [inte=not] vem det är
ifrån. X
[yep*p* but it doesn't show who sent it. X]

If there is a shorter word to be found in another language, which almost always is English, it could just as well be used. It then serves the double function of informality marker or perhaps just exhibits word play in

general. It seems that English words are commonly used in this way in young people's interaction.

Example 69. KATTEN HAR FÅTT **KIDS** [ungar]
[THE CAT HAS GOT KIDS [babies]]

Consonant writing

By omitting vowels, the message appeared to be almost like consonant writing. The intended meaning seems to come through even without them. The examples below did not have to be shortened in order to save space, but appear, rather, to be expressions of language play.

Example 70. Har du **prgmrt** [programmerat] videon?
[Have you **prgrmd** [programmed] the video?]

Especially Example 71 illustrates linguistic awareness by indicating consciousness of the "Swenglish" construction *blada*, which is a shortened form of an English loan word (roller blading), adapted according to Swedish inflectional system .

Example 71. BLADA? **SVNGLSKA** [SVENGELSKA=swenglish]!
[BLADING? **SWNGLSH!**]

Conventional abbreviations

Conventional, or established, abbreviations were found in the SMS corpus. They are used for the same reasons as they are used in other contexts of written communication: saving time and space. Example 72 illustrates an abbreviation that is established and used even in formal contexts.

Example 72. Ska vi köpa nåt vin **el. dyl.** [eller dylikt]
[Can we buy some wine **etc.** (or something)]

The abbreviation in Example 73 is established, but used in less formal contexts.

Example 73. Lust att hålla mig sällskap på en cigg så ska du få se **ngt** [något], förutom en lerig Fia?
[If you'd like to keep me company smoking i'll show you **something**, apart from a muddy Fia (name)]

Unconventional abbreviations

The previous studies in the present dissertation (email, webchat, instant messaging) all showed use of unconventional abbreviations. This was also the case in the SMS corpus. Conventional methods of abbreviation may be used unconventionally which would require additional explicit information if it had occurred in an autonomous text. The first letter in a word may stand for the whole word, as in Example 74.

Example 74. SKITSNACK! VILKEN **T** [tid]?
[*BULLSHIT! WHAT **T** (TIME)?*]

It seems likely that the message in Example 75 was the answer to previous communication - which might just as well have been conducted via SMS as through any other mode of communication – and thus not difficult to interpret. Other messages show innovative new types of abbreviations based on Swedish words.

Example 75. X! På väg till y. inga pengar kvar på **tfn:en** [telefonen]. ringer i e.m. **CS** [ses]!
[*X! On my way to y. no money left in **the phn** [the phone (pre-pay card)]. will call this p.m. **CU** [see you]!*]

Example 76. Tjena läget då? Ska vi gå på bio **lr** [eller] nåt ikväll?
[*hi how are things? Wanna go for a movie **r** [or] something*]

Example 77. Va **QL**[kul] det ska bli på LÖR!
[*How **FN** [] it'll be on SAT!*]

Abbreviations based on non-Swedish words were found in the SMS material. Words and whole phrases in English may appear in the midst of Swedish conversation. These might be in the form of snippets of song lyrics to popular songs, poems or fixed phrases. Examples of the latter are illustrated in Examples 78 and 79. It seems likely that this usage has been observed and picked up from chat room norms.

Example 78. YES YES, MASTER OF BARBECUE RETURNS... **BTW**
[by the way = förresten]: LYCKADES INTE FÅ
MED EXPEDITEN PÅ SYSTEMET...
*[YES YES, MASTER OF BARBECUE RETURNS... BTW
[by the way = förresten]: DIDN'T MANAGE TO
BRING THE SALES GIRL AT THE LIQUOR STORE
WITH ME...]*

Example 79. Det var ju inte ett kryptisk meddelande...
man skulle kunna tro att du hade X. på
besök:) Lev Väl! Miss **U** [you=du]!
*[As if that wasn't a cryptic message... one
could well believe that you had x.
visiting:) Take care! Miss U [you=du]!]*

Whole message in either all lower-case or capitals

Whole messages were found to be typed either in all lower-case or all capitals. Generally, when typing on a computer keyboard, typing shift plus the letter key requires effort, even though experienced typists probably do it without thinking. A common CMC norm states that typing words or whole messages in capitals is a way of imitating prosodic features of shouting, or emphasis in general. Adhering to the norms of sentence initial capitalization might be easy or difficult, depending on type of mobile phone and how its preferences are set.

Example 80. biljetter hämtade! bion börjar 18.45 när
ska vi ses?
*[tickets collected! the movie starts at
6.45 when can we meet?]*

Older models of mobile phones do not have the option to change shift, but typed everything in capitals. Those messages were not intended to be interpreted as emphasized. Informants did not always indicate which type of phone was used when creating the message, which made it somewhat difficult to interpret intended meaning. Example 81 below does not seem to be an exclamation, and it was probably typed on an older type of phone.

Example 81. E DE TAEKWANDO DU TRÄNAR?
[IS IT TAEKWANDO YOU PRACTICE?]

7.8.5 Grammar

Grammatical reductions were found to be used in mobile text messaging, in order to save time, effort and space. Results suggest that written language in SMS show characteristics of informal spoken interaction. As most messages are sent between people who know each other already, senders were able to rely on the receiver's ability to pragmatic inference when decoding the message. This allowed senders to omit elements which the message could be interpreted without anyway. SMS texting does not have to be as explicit as traditional writing has to.

Subject pronoun

Results support the findings of Döring study, and e.g. subject pronoun was found to be omitted frequently. If the sender's phone number is stored in the receiver's phone book, the sender's name will appear above the message on the screen, so it is obvious who sent the message. As illustrated by Example 82, it is often obvious to the receiver whom the deictic expression *jag* [I] refers to, and it may just as well be left out.

Example 82. [Jag] kan inte ikväll. [Jag] måste jobba.
[Jag] gillar dig i alla fall. KRAM x
[I] can't tonight. [Jag] have to work.
[Jag] like you anyway. HUGS x

The omission of subject pronoun is a feature that is normally characteristic of spoken informal interaction, and is not associated with traditional writing.

Verb phrase

Other grammatical reductions may save time and space. By leaving out copula, auxiliary or modal verb in combination with preposition may save several keystrokes. Several interpretations of the message in 83 are possible (e.g. *blir det på...*[is it going to be at...], *sa vi på...*[did we say at...]), but the receiver who had all the necessary background information most probably had no trouble decoding it.

Example 83. [Ska vi ses på]cafe japan [kl]19?
[Can we meet at] cafe japan [at] 19
[o'clock]?)

Deletion of definite article, possessive pronoun

Example 84. below illustrates grammatical reduction by leaving out the possessive pronoun.

Example 84. [Tack för **[ditt]** stöd! Verkar gå att ha de små i hallen...
[Thanks for **[your]** support! It seems to be ok to have the little ones in the hallway...]

Definite article may be suffixed to nouns in Swedish. Messages in which the suffix was left out were found in the SMS corpus.

Example 85. STRYKJÄRN**[ET]** STÅR LÄNGST OPP VID OLIVEN
[[**THE**] IRON IS AT THE TOP BY THE OLIVE]

Example 86 illustrates a message in which a few prepositions and a possessive pronoun were omitted. This type of message would probably fall into the category Telegram style in Döring's (2002a) study.

Example 86. [**I**] Stan [**i**] Jönköping i så fall. [**Min**]
Lillasyster fyller 20, kalas...
[[**In**] Town [**in**] Jönköping in that case.
[**My**] Baby sister turns 20, party...]

7.8.6 Graphical (non-alphabetic) means

Emoticons

Emoticons are used to enhance alphabetic writing by conveying moods or emotions that are normally expressed with extralinguistic cues such as facial expressions and tone of voice in spoken interaction. Symbols in imitation of facial expressions in a monomodal written means of expression may help to make it easier to interpret text-only communication. The experienced communicator seems to know that some messages might need additional information to disambiguate text-only communication.

Inserting non-alphabetic symbols in mobile text messages may be somewhat troublesome on most types of mobile phones, due to the limited keypad. Some phone models are targeted and marketed for teenagers. As it seems that the mobile phone market assumes teenagers are the stereotypical user of emoticons, some phones have the most common emoticons (i.e. the happy face [:-)], the sad [:-(] and the winking emoticon [;-)]) preformatted. These do not require insertion of three symbols, but the whole emoticon may be selected and inserted as one. The economy principle does not always hold, as has been shown, and people seem willing to spend the time and effort it takes to insert emoticons to enhance their messages. Example 87 below was typed on a mobile phone that did not have any preformatted emoticons.

Example 87. `Yepp, den har jag ! Fixar en kopia på den
:-)
[Yepp, I've got it! Will make a copy of it
:-)]`

No information was given of how Example 88 was created. It seems likely, though, that it was typed on a computer keyboard as it includes both emoticons and asterisks. But, situational variables such as relationship between communicators and the goal that the sender wishes to reach may exert more powerful influence than economy.

Example 88. `*pet i sidan* Hur vågar du ha upptaget när
goa pågen ringer..!? *peta lite till* ;o)
Hör av dej när linjenär öppen! *Cjamiz*
/Förnamn.`

*[*nudge* How dare you keep the line busy
when the nice boy calls..!? *nudge again*
;o) Call me when the line is open again!
Hugz/First name.]*

Asterisks

As has been shown in studies of web chat and instant messaging, asterisks to frame words or phrases often serve the same purpose as emoticons. By adding explicit words, the message is rendered even more unambiguous. Actions described explicitly in words, such as the one illustrated in Example 89 below, may also be marked to resemble additional prosodic features by repetition of letters.

Example 89. Älskar dig *puss o en lååång kram*
*[Love you *kiss and a looong hug*]*

Example 90 illustrates a feature that is most often associated with spoken interaction (see Spoken-like, or unconventional spelling, above). As discussed above, extralinguistic cues are normally not spelled out in writing. By adding asterisks around the typed version of laughter, it is marked explicitly as an action and probably also as an indicator of how it was supposed to be interpreted.

Example 90. ello babe.. va görs?? spanar du på kalle
eller..? :p *haha* [skratt] keep rockin'
//x.
*[ello babe.. what are you doin?? Looking
for kalle..? :p *haha* [laughter] keep
rockin' //x.]*

Symbols replacing words

By replacing a word with a symbol that stands for the word, several keystrokes may be saved. Example 91 utilizes the maximum character limit. The last word was misspelled by leaving out the last letter, had it conformed to normative spelling: [kväl -> kväll]. This misspelling does not seem to pose any difficulties to interpretation. People add and subtract information, and use their experience and pragmatic knowledge to get to the most likely and relevant interpretation (cf. Sperber and Wilson 1986). By using symbols instead of words 7 keystrokes were saved in this

particular message. Additionally, abbreviations were used to save space [gbg->Göteborg, ngt->något (something)].

Example 91. Jo, är i gbg sen 1 [en=a, one] vecka. men har gjort ngt idiotiskt. skaffat ett 2 [två=two] veckors städjobb i hamnen. måste upp 5 [fem=five] på morron & [och=and] är DÖD när jag kommer hem.. Ringer i kväl [Well, been in gbg 1 week. but have done something stupid. got myself a 2 week cleaning job in the harbor. have to get up at 5 in the morning & am DEAD when I get home.. will call tonigh]

In Example 92 an emoticon was employed to stand not just for an emotion, but actually also for the word glad [happy] itself. The message is a fixed expression common in Sweden for happy wishes around midsummer.

*Example 92. :-) [glad] midsommar!
[:-) [happy] midsummer!]*

7.9 Conclusions

This study endeavored to investigate how written language is used and adapted to the conditions of mobile text messaging, or SMS.

SMS is an asynchronous mode of communication, and the means of expression is constrained by being written. Moreover, production and perception settings are conditioned by several factors. First, the tiny keypad with its restricted number of keys makes text input cumbersome. Messages created on computer keyboards, sent via SMS services provided on the web, are easier to produce. However, these were seldom used in this particular corpus. Second, message size was restricted by an upper limit of 160 characters per message. Third, the small screen of a mobile phone restricts which kinds of messages may be read (cf. the chain message, exemplified in Figure 18).

Situational factors related to for example relationship between communicators, aim of communication, impact message composition and

language use. Most SMS messages were sent between people who already knew each other, and who shared an extensive amount of background information. Thus, despite being asynchronous and written, messages could be more inexplicit than traditional writing. In combination with the parameters mentioned above, SMS is also a mobile means of expression. It allows messages literally to be sent from anywhere and any time. This may contrast with the conditions for communication of email, web chat and instant messaging, which all have to be conducted at a computer. This fact seems to influence on what people communicate about.

Due to the above-described parameters, language use in SMS was found to be adapted to the conditions of mobile text messaging in a number of interesting ways. Regarding punctuation, it was found that users save time, effort and keystrokes by omitting punctuation and omitting blank space between words. Unconventional punctuation, such as excessive use of exclamation marks, was found to be used in an expressive manner to convey emotions in resemblance to the informal norms used in chat rooms, for instance.

As for spelling in mobile text messaging, predictive text entry may be responsible for many typos due to guessing other words than those intended by the writer. The sender may also overlook spelling mistakes, or mispredictions as the case might be, because SMS is used in informal contexts. Additionally, the part of the message where the typo occurs may scroll off the tiny screen before the whole message is read, and not be visible to the sender. Reasons to spell words in a manner reminiscent of how they sound when spoken are the informal contexts in which texting is used, and in the relationship between communicators. Spoken-like, or unconventional spelling also refers to the inclusion of spelling features that are normally associated with spoken interaction. Examples were hesitation sounds and laughter.

There seems to be a tendency to split Swedish compound words. Theories as to the reasons for this vary; one of them is the influence that the English norms of writing have. The difficulty that predictive text entry software has in accepting compound words makes it easier to split compounds rather than to manually enter the word in the memory of the phone. Whether all split compounds are due to software mispredictions or the senders' own tendency to split compounds is not always evident. Another way of saving keystrokes was found to be omitting vowels in words, thus resulting in consonant writing (which in some cases was not fully consistent). This strategy seemed to be used more for the joy of language play than effective economy, even though it is economical too.

Conventional, established abbreviations were found to be used much in the same manner as they are used in other written contexts. The function of abbreviations is the same, to save time, effort and space. More interesting was the discovery of unconventional, new and as yet unestablished abbreviations. Many of these seem to be carried over from the standards in other modes of CMC. Abbreviations based on non-Swedish words (predominantly English) were found, but new types of abbreviations based on Swedish words were also found.

Another way of saving time and effort was found in the use of typing messages in either all lower-case or all capitals. This use was probably connected with the technical conditions of the means of expression used. Some phones may be set to start (what the predictive software judges to be) new sentences with a capital, while older models often do not have any other option than to type everything in capitals. Single words or phrases that were typed in capitals in the midst of a message signal emphasis that resembles use observed in chat rooms and instant messaging.

Grammatical reductions were found in the SMS corpus. They were used for practically the same reasons as reductions in spelling and punctuation. By omitting subject pronouns, for instance, the sender relied on the background information that he or she shared with the receiver, rendering inclusion of subject excessive information. Thus, time, effort and space are saved and the message is given an informal tone.

Graphical symbols other than alphabetical characters, such as the use of emoticons and asterisks, were found to convey extralinguistic cues, by analogy with observable use in chat rooms, for instance. These are an enhancement of alphabetical writing, and were used as a means to disambiguate the monomodal means of communication. Use of emoticons and asterisks resembles corresponding use in chat rooms and instant messaging, and was probably carried over or inspired by the norms that users have observed in those modes of communication. Symbols that replace fully spelled out words (e.g. the symbol for a number replaces the word for the number) were found primarily to save space.

7.10 Future research on mobile messaging

Multimedia Message Service and Enhanced Messaging Service

Up to now, the SMS service has been used mostly for simple text messaging. New and enhanced forms of the service are being introduced.

Multimedia Message Service (MMS) is the further evolution of SMS and mobile-to-mobile messaging. MMS can transmit messages containing text, graphics, photographic images, audio and even video clips between mobile devices using WAP as bearer technology, powered by the high-speed transmission technologies such as GPRS (General Packet Radio Service).

Enhanced Messaging Service (EMS) is a function that allows users to send and receive a combination of simple melodies, pictures, sounds, animations, modified text and standard text as an integrated message for display on an EMS compliant handset. EMS supports several simple media in standardized formats: formatable text (alignments, font sizes and bold, italic, etc. text), three picture formats, predefined sounds, and moving pictures.

Although the technology is already available, it is not yet widely used, as it demands certain types of mobile phones and is yet too expensive to make it worthwhile. As with the increase in popularity of SMS, most of one's friends and family have to have the same type of technology to receive one's messages. How useful people will eventually find it is another matter. How widely new technologies are used depends on speed, cost and usability. This study was devoted to how written language is used and adapted to new modes for written communication, and an investigation of MMS or EMS will have to be left for now. To learn more about MMS and EMS, see for example <<http://www.mobilemms.com/>>.

7.11 Chapter Summary

This sub study endeavored to investigate how language was used and adapted to the conditions of mobile text messaging, or SMS. An introduction to SMS opened the chapter, discussing production and

perception conditions of SMS. Previous research and their findings of linguistic features in SMS were reported. Material, methods and research questions of the present study were then accounted for. Results compared word frequency across four modes of CMC in order to establish influence of activity and means of expression on language use. Linguistic features that were found to be characteristic of the messages in the Swedish SMS corpus were presented, along with illustrating examples. Language use in mobile text messaging was found to be adapted to the constraints of production and perception conditions due to the means of expression, as well as situational parameters. The resulting language use showed many features associated with spoken language, omission of subject pronoun, verbalization of OCM (Own Communication Management), hesitation sounds and laughter. Spelling reminiscent of spoken interaction served to save time, effort and space, and to render an informal touch and serve as in-group markers. Abbreviations, both conventional and unconventional, were used for the same reasons as spoken-like spelling. New and yet unestablished abbreviations based both on English and Swedish words were found. It seemed that this adaptation was a transfer from the standards that are observable in chat rooms.

Use and adaptation of written language in mobile text messaging is deemed to be a variant of language use that is creatively and effectively suited to the conditions of SMS.

8 Concluding Discussion

8.1 Introduction

The ultimate purpose of this dissertation is to analyze how written language is used and adapted to the particular circumstances of four specific modes of computer-mediated communication. A flexible view of human behavior lay behind the arguments. One presupposition for this study is that one of the chief assets of human beings is the ability to adapt behavior to suit the constraints and enablements of a given situation. This ability to adapt is also realized in individual linguistic behavior: how and why people use language the way they do under particular circumstances. Moreover, popular claims that language is deteriorating and that it is deteriorating at an even higher speed due to use in CMC are challenged. Rather, language use in CMC is used as evidence of the uniquely human ability to adapt language according to the interdependent variables of Situation (context, activity, relation to communicators), Means of expression (writing vs. speaking, production and perception conditions of the means of expression), and level of Synchronicity.

Written language, in the types of CMC that are investigated here, is in many ways divergent from normative written language. A number of factors have to be taken into consideration before what is normative for electronic communication can be established, or even useful. The stance taken in this work is that rationality is instrumental: rationality correlates with situation, or activity, goals and aims of communication, means for communication, and so on.

The dissertation also challenges the claims put forward by Crystal (2001) that CMC is a genuinely new medium. Even though it is acknowledged that communication in CMC might be conducted in ways that neither spoken nor written interaction can be (e.g. the contributions in

synchronous, written multiparty conversation in a real-time chat are more persistent than the spoken sound wave, as Crystal pointed out), the features of the modes of CMC investigated in this study are not genuinely new. They are just variants adapted to fit the new situation. Furthermore, it is pointed out that CMC cannot be treated as one single mode of communication (cf. “Netspeak”).

A taxonomy of linguistic features that have been found to be characteristic of the modes of CMC investigated in the present study will be presented next. Summary commentary on each of the chapters in the dissertation will follow below that.

8.2 Linguistic features in CMC

Below follows a taxonomy of linguistic features found to be common characteristics of text in the modes of CMC investigated in the present dissertation. Examples of each category will be provided for each CMC mode in the summarizing description of findings in the sub-studies⁸¹.

First, features related to *space, punctuation, spelling and case* constitute strategies that seem mainly to be used for reducing number of keystrokes. Second, common *grammatical features* have turned out to be different strategies to reduce sentence, or message, length. Third, a category labeled *logotypes* consisted of different ways of using ASCII characters. Fourth, the category *lexical features and abbreviations* includes lexicon, code switching, OCM features from spoken language spelled out in writing, as well as different types of abbreviations.

⁸¹ As mentioned above, the term “CMC” refers to the modes of CMC that have been investigated in this dissertation.

Table 39. Linguistic features characteristic of CMC.

<i>Category</i>	<i>Feature</i>
1. Space, case, punctuation & Spelling	
<i>i. Space</i>	<ul style="list-style-type: none"> a. Omitting blank space between words b. Omitting punctuation c. Unconventional punctuation
<i>ii. Case</i>	<ul style="list-style-type: none"> a. All lower-case b. All capitals c. Mix of lower-case and capitals
<i>iii. Spelling & punctuation</i>	<ul style="list-style-type: none"> a. Unconventional, spoken-like spelling b. Typos, or mispredictions c. Repetition of letters and punctuation d. Repetition of words e. Consonant writing f. Split compounds
2. Grammatical features	
<i>i. Reduced sentences</i>	<ul style="list-style-type: none"> a. Subject pronoun b. Verb phrase c. Preposition or possessive pronoun d. Exchange long words for shorter
<i>ii. Word order</i>	<ul style="list-style-type: none"> a. Inspiration from English (word order, prepositions)
3. Logotypes	
<i>i. ASCII characters</i>	<ul style="list-style-type: none"> a. Emoticons b. Asterisks c. Symbol replacing word d. Addressivity marker
4. Lexical features & Abbreviations	
<i>i. Lexical features</i>	<ul style="list-style-type: none"> a. Colloquial lexicon (dialect, expletives) b. Code switching c. OCM features from spoken language
<i>ii. Abbreviations</i>	<ul style="list-style-type: none"> a. Conventional abbreviations b. Unconventional abbreviations

8.3 Summary of the dissertation

8.3.1 Chapter 1: Introduction

In **Chapter 1**, computer-mediated communication was introduced and the research questions were put forward. Reasons for why these were relevant and interesting to pursue were discussed. This dissertation deals with which factors determine how written language is used in CMC. Research questions are:

- How is written Swedish used and adapted in various modes of CMC?
- In what ways do these modes of CMC differ from the norms of traditional written language and why?
 - Which are the written vs. spoken language features in each mode respectively, and why are these used?
 - Which specific “e-style” characteristics are found in different CMC modes?
- Which variables influence these adaptations?

8.3.2 Chapter 2: Background

The theoretical background, in **Chapter 2**, comprised theories of human communication in general, and moved on to consider human communication more specifically, namely written and spoken communication. An account of previous research on even more specified communication was given in the section that followed, which dealt with how various disciplines have grappled with CMC in general and the four modes of CMC that are the focus of attention in the present work in particular.

A taxonomy of the factors that were taken to be important for how people use language in email, web chat, instant messaging and SMS was set up. An overarching principle of human ability to adapt behavior according to the facets of a given communicative situation was assumed: human beings have the ability to adapt their linguistic behavior to suit the circumstances of a communicative situation. This outlook is not very controversial or surprising. This ability was a presupposition for being able to adapt use of language according to *situation*, *means of expression* and *synchronicity*.

Synchronicity was suggested to have the following aspects: production and perception distributed in space and/or time, time pressure of not.

Means of expression was put forth as having the following aspects: production and perception conditions, the persistence of text⁸² (cf. Herring 1999), one-way or two-way interaction.

Situation was taken to have the following aspects: relationship between communicators, activity, aim or goal of interaction

8.3.3 Chapter 3: Methods and Material

The reasoning and choices of research questions were based on the above theoretical background, as were choice of data sets as well as methods to gather and analyze it, which were presented in **Chapter 3**. The selection of the respective data sets was motivated both with respect to previous studies and with respect to availability of material and ethical issues. The general method was to gather the data in electronically readable text format for automatic, quantitative analyses. The tool used for automatic analysis was TraSA (Transcription Statistics with Automation), to investigate various measures in all four modes of CMC. Examples of those measures are number of words in each data set, number of contributions, frequency lists, mean length of contribution. The frequency lists were searched for abbreviations, both established and new, innovative ones. Frequency lists of emoticons, complex punctuation expressions, ASCII-characters, etc. were also compiled.

Specifics of each particular set of data, both concerning methods of data collection and analyses, were given in the chapters that dealt with each mode respectively.

8.3.4 Chapter 4: Email

Chapter 4 was dedicated to the analysis of email in comparison with traditional pen-and-paper letters. The particular corpora of email and traditional letters were taken from publicly accessible archives at the city council of Göteborg city. The fact that both email messages and traditional letters were sent from members of the public to someone at the city council made them special in comparison to other studies of email. First, the email messages were posted on a web site form and not from a normal

⁸² Written contributions are more persistent than spoken, due to the ephemeral nature of soundwaves. Herring (1999) argues that written conversation in real-time chat rooms is "persistent conversation".

email client. Second, the recipient of the messages was some authority official, who in most cases was not known even by name to the sender. This meant that communication was neither like the many-to-many messages found in Listserv discussion groups, nor quite like private one-to-one email. Third, the corpus of traditional letters were those sent from private members of the public to the same kind of recipient as the email messages: to some authority official. The letters were also sent for the same reasons as the email messages, requesting information about the city, complaining about rules and regulations, asking for help, etc. The corpora were thus comparable for the aim of finding out how the means of expression influence language use in this type of written communication.

Email – findings in relation to research questions

Not many specific “e-style” characteristics were found in the email corpus. The greater part of the messages had an overall air of being far more informal than the corresponding traditional letters, being opened with an informal greeting and closed with first name only. Even more ‘e-style-like’ were the messages that had neither opening nor closing, but consisted of a single question or remark. The styles ranged from very informal (nothing but a misspelled question in the subject line) to highly formal (a job application with full formal features). Informal and spoken-like grammatical structures, such as deletion of subject pronoun, were found. The email messages included spelling mistakes, but so did the traditional ones. Some new abbreviations and the occasional emoticon were found in the email corpus, while none were found in the traditional corpus. Non-normative and informal use of punctuation marks was also found, exemplified in the excerpt in Table 40 below.

Table 40. A taxonomy of linguistic features characteristic of email.

<i>Category</i>	<i>Feature</i>	<i>Example</i>
<i>Space, case, punctuation & spelling</i> <i>Space</i>	Omitting punctuation	<i>Var kan jag få tag på Konsument Göteborg</i>
	Unconventional punctuation	<i>Finns det någon som kan skicka lite information om slottskogen???????????????? ????????????????????</i>
<i>Case</i>	All lower-case	<i>vi vill veta mer om göteborgs kalset</i>
<i>Spelling & punctuation</i>	Unconventional, spoken-like spelling	<i>Kramixxar från en Sthlm´s tös..</i>
	Typos, or mispredictions	<i>Jag skall på Rilling Stones i morgon och skulle vilja ha svar på två frågor.</i>
	Split compounds	<i>Jag tycker det är jätte bra</i>
<i>Grammatical features</i> <i>Reduced sentences</i>	Subject pronoun	<i>[Jag]Har tagit del av er IT- startegi på hemsidan.</i>
	Verb phrase	<i>fördelar och nackdelar brensleåtgång? Priser?</i>
<i>Logotypes</i> <i>ASCII characters</i>	Emoticons	<i>8^)</i>
	Symbol replacing word	<i>Presentationen av Göteborg&co styrelse är minst ett år gammal.</i>
<i>Lexical features & abbreviations</i> <i>Lexical features</i>	Colloquial lexicon (dialect, expletives)	<i>Ni götborgare är jävla lantisar. Ni kan sluta va så jävla kaxiga</i>
	<i>Abbreviations</i>	Conventional abbreviations
Unconventional abbreviations		<i>..Skulle underlätta en hel del om det fanns på denna sida så man bara kan komma hit å leta jobb ..</i>

Examples 93 and 94 below are both requests for information. They display several markers of informality: both are almost completely typed in all lower-case, no return address was given, normative punctuation is disregarded, and typos are left uncorrected. From the name of the sender in Example 94, one can deduce that a young student wrote it.

Example 93.

Email message	Translation into English
Ämne: hjälp mig kan jag få fakta om regionen mellan göteborg och oslo tack jag ären vanlig göteborgare	[Subject: help me kan i have facts about the region between göteborg and oslo thanks i am an ordinary gothenburger]

Example 94.

Email message	Translation into English
Ämne: bilar fördelar och nackdelar brensleåtgång? Priser? risker ... olyckor ... antal olycksfall per år i göteborg? skicka snarast ... xskolan	Subject: cars advantages and disadvantages fuel consumption? Prices? risks ... accidents ... number of accidents per year in göteborg? send as soon as possible ... name of school

General characteristics of written language in this type of email messages were that they were comparably short, spanning a wider range of style than traditional letters. The traditional letters to the city council were generally longer than the email messages, and conformed more to the norms of standard written language in form and style. They did not show any new or not yet established abbreviations. All letters opened with the normative heading which indicates the subject of the letter, and all letters were signed with the sender's name and contact information.

Email – Findings in relation to three dimensions

The three dimensions synchronicity, means of expression and situation are interrelated, which is shown in the findings. Both email and traditional letters are technically asynchronous modes of communication. The electronic means of expression makes email less asynchronous as it were, or at least delivery time and response time can be reduced markedly. The fact that the communication is written rather than spoken, electronic rather

than handwritten seems to invite informality even between people who do not know each other. People are probably also influenced by the standards from Internet culture. Written communication does not commit the sender to anything, and does not require the receiver to face up to questions or demands immediately as would have been the case in spoken interaction, be it face-to-face or over the phone. Moreover, tertiary means of expression such as electronic mail place the sender even more “at arm’s length” from the receiver. Electronic text does not give away any clues as handwritten texts do, and the communication demands less in monetary cost, physical effort and time. As for situation, which incorporates relationship between communicators as well as goal of interaction, even though the receiver was a sort of authority to the sender, most email messages were informal in style. It seemed that goal of communication influenced how messages were formulated. Messages in which the sender requested information or help were more carefully and normatively composed than those that gave information or complained. Judging from content of the email messages, it often seemed that the easy access and fast delivery makes for communication that probably would never have been realized if it had to be conducted via traditional letters or in spoken interaction.

8.3.5 Chapter 5: Web chat

Chapter 5 dealt with the analysis of written language in a popular web chat. It reported the methods and findings of a questionnaire answered by 333 students between the ages 13-21. The aim of the questionnaire was partly to find out which web chat seemed to be one of the most popular at the time, and thus which was to be logged for gathering of linguistic material. Another reason for inquiring into the habits of the students’ concerning web chat and email use was to find out whether they used abbreviations and emoticons as observed in previous studies. The logged chat material was searched for occurrences of these. Results show that the most frequent emoticons are basic smiley faces, despite the many fantastic combinations of keyboard characters that the students claimed to use, judging from the questionnaire. Results also indicated that new and not yet established abbreviations were being used. Some of these seemed to be transferred straight from the norms of use in international chat rooms, others were based on Swedish words in analogy with the ones taken over from international chats. As in the case of emoticons, only a few abbreviations were widely used. It was suggested that this had to do with the reasons for including them in the contributions, namely to save time to

keep up with the speed of conversation, and thus save keystrokes. At the same time, the text was made to convey more than alphabetical writing normally does. By including an emoticon, a few keystrokes can convey what would take a number of words to express. The initial intention of saving keystrokes and keeping up with the flow of conversation would be lost if emoticons or abbreviations were too complex. It would simply require too much effort and time both in production and in decoding. The result of this was that only the very basic emoticons and abbreviations were used widely. There seems to be a conventionalization at issue, the basic ones are more readily understood.

It was also suggested that the multifunctionality of saving keystrokes and time as well as expressing an in-group mastery of language norms of the particular activity, played a great part, and also overruled the principle of least effort.

Web chat – Findings in relation to research questions

Written language in web chat is used and adapted to suit the social-recreational situation in a real-time written conversation with multiple unknown participants. It differs from the norms of traditional written language in that it is a form of conversation, which happens to be written down instead of spoken. Contributions are short and do not have to be as explicit as in traditional writing, the syntactical structure is often spoken-like with verbs and pronouns left out. Spelling and normative punctuation rules were not necessarily respected. Specific ‘e-style’ features such as new types of abbreviations, emoticons, asterisks, complex punctuation expressions, ASCII-characters, were used abundantly.

Innovative, new types of abbreviations were observed; some seemed to be taken over from the norms in international chat rooms, *lol* (laughs out loud) and *brb* (be right back). Some were made in analogy with these but based on Swedish words, *es* (Sw. *ses* (see you) and *ql* (Sw. *kul* = Eng. fun). Asterisks, which are used to indicate an action, attitude or emotion, were used to frame both Swedish and English words, **ler** (Eng. **smiles**). Asterisk-framed words or phrases were sometimes abbreviated too, e.g. **/** (for **ler**, see above). An abundance of variants of complex punctuation expressions were used, e.g. *---->* to indicate addressivity, *!!!!!!!* or *??!?!?!?* to express for example emotion, emphasis or even to draw attention. ASCII-characters such as the proverbial cyber rose *-<-<-<@* and emoticons were also found to be used.

Table 41. A taxonomy of linguistic features characteristic of web chat.

<i>Category</i>	<i>Feature</i>	<i>Example</i>
<i>Space, case, punctuation & spelling</i> <i>Space</i>	Omitting blank space between words	<i>Vafanerehär?</i>
	Omitting punctuation	<i>HAR DU DAMP ELLER TA INTE ILLA UPP MEN DU ÄR LITE SEG</i>
	Unconventional punctuation	<i>Nån?????????!!!!!!!!!!!!!!!!!!!!</i>
<i>Case</i>	All lower-case	<i>mm,,visst,,vart bor du då ??</i>
	All capitals	<i>SVARA FÖR HELVETE</i>
	Mix of lower-case and capitals	<i>JAg Avgudar dom!!</i>
<i>Spelling & punctuation</i>	Unconventional, spoken-like spelling	<i>Du kan väll komma och hälsa på vettja!! värst va[d] du smörar tjebnare</i>
	Typos, or mispredictions	<i>nirvana fäääääns</i>
	Repetition of letters	<i>MADDE MADDE MADDE MADDE</i>
	Repetition of words	<i>MADDE MADDE MADDE MADDE MADDE...</i>
	Split compounds	<i>Jag måste önska dig LYCKA TILL MED JOHAN ÄR LITE SVART SJUK PÅ JOHAN ATT HAN FÅR DIG *SUCK*</i>
	<i>Grammatical features</i> <i>Reduced sentences</i>	Subject pronoun
Verb phrase		<i>[är det] nån [som vill chatta]?????????</i>
Preposition or possessive pronoun		<i>[Jag bor i]x-stad hemma hos pappa och mamma för tillfället....du??</i>
<i>Word order</i>	Inspiration from English (word order, prepositions)	<i>Kan vi gå privat???</i>
<i>Logotypes</i> <i>ASCII characters</i>	Emoticons	<i>HEJ DÅ,SKÖT OM ER.....---{- -@ Du retas med mig;)..</i>
	Asterisks	<i>*trist*....</i>
	Symbol replacing word	<i>1 cool brud</i>

Table 41, continued.

<i>Category</i>	<i>Feature</i>	<i>Example</i>
<i>Lexical features & abbreviations</i> <i>Lexical features</i>	Addressivity marker	<i>Raggaren>>>>Lisa</i>
	Colloquial lexicon (dialect, expletives) Code switching	<i>Hej, X gött att tjöta med riktigt folk!!</i> <i>Kungälv!! The big city, vad gör du , pluggar?</i> <i>What the fuck what that all about??</i>
<i>Abbreviations</i>	OCM features from spoken language	<i>Utfors...*hmmmm*...Där känner jag en kille som jobbar.</i>
	Conventional abbreviations	<i>skämde du bort din förra tjeje ? tex med rosor ??</i>
	Unconventional abbreviations	<i>det är ql att resa... måste dra cs</i>

Web chat – Findings in relation to three dimensions

The findings of language use in web chat are related to the three interrelated dimensions of synchronicity, means of expression and situation. Real-time web chat such as the one investigated in this study is synchronous. It has been discussed in previous research and above, that it can never be fully synchronous as spoken face-to-face interaction, because of the time it takes to type contributions and that the receivers have no means of being aware that a contribution is being created before it is displayed in its entirety in the chat window. One might call web chat quasi-synchronous or near synchronous. Written interaction is also monomodal rather than multimodal, which in connection with synchronicity decides how and which signals may be sent.

Characteristics of written interaction in a web chat were evidenced in the content of the messages and the spoken-like structure. The communicative situation in the logged chat was that it was an open, topic-unspecified chat and that most participants were there to socialize and “hang out”. As with email, the tertiary means of expression of electronically transmitted written contributions seemed to invite informality. The lure of a synchronous written interaction with a mass of unknown people seems to be the anonymity that text-only communication enables. The tone was informal, the topics often containing sexual content. In an anonymous text-based

environment people might try out new social roles and language, which they would not be able to try elsewhere. Intriguingly, spelled and chosen nicknames seemed to be a device for drawing attention and inviting conversation in the ever-flowing textual noise. The lasting nature of textual contributions, in comparison to the ephemeral nature of the speech signal, seemed to be a device for cohesion in the multiple simultaneous threads of conversation that constantly scroll up the screen.

8.3.6 Chapter 6: Instant Messaging

Chapter 6 analyzed the messages sent through an instant messaging system in a synchronous awareness tool called WebWho. It was used to create awareness of presence in a large university computer-lab with workstations in many different rooms on several levels in the building. A schematic view of the lab, accessed through any web browser, gave awareness of who was logged on to which workstation. Instant messages could be sent from one workstation to another, or from outside the lab to a computer in the lab. The message window popped up top-most on the receiver's screen. By default, messages were headed with the sender's student ID, but could also be sent anonymously by checking a box in the message window.

The study analyzed how different kinds of (sender) awareness of presence influenced content in instant messaging between students using WebWho. The students were found to use the messaging system to support collaborative work and coordinate social activities, as well as for playful behavior. Cross analysis of sender location (within the same room, from a different room, and outside the lab), sender status (anonymous vs. identified) and message content were made. The messages were compared with corpora of spoken and traditionally written language, as well as with previous studies of other communication media (email and chat).

Instant Messaging – Findings in relation to research questions

A number of variables influence the use and adaptation of written language in the instant messaging service of the awareness tool WebWho, and results suggest the sender's awareness of receiver's presence, knowledge that the message is delivered instantly, as well as relationships between communicators and goal of interaction.

Instant messaging differs from the norms of traditional written language in several respects. Texts show spoken language characteristics such as

OCM features, syntactical reductions, colloquial language, etc. Specific “e-style” characteristics similar to those found in web chat were discovered, for example emoticons, unconventional abbreviations and non-normative use of punctuation, etc.

Table 42. A taxonomy of linguistic features characteristic of Instant Messaging (IM).

<i>Category</i>	<i>Feature</i>	<i>Example</i>
<i>Space, case, punctuation & Spelling</i>	<i>Space</i>	
	Omitting punctuation Unconventional punctuation	<i>nu kraschade det igen</i> <i>Men hörru!!!!!!????????</i> <i>Kan du inte SQL??????????</i>
<i>Case</i>	All lower-case	<i>magnus är här</i>
	All capitals	<i>JA NU RÄCKER DET. DET HÄR</i> <i>ÄR INTE ROLIGT</i>
	Mix of lower-case and capitals	<i>OmedelBums!!!!!!</i>
<i>Spelling & punctuation</i>	Unconventional, spoken-like spelling	<i>Maskinen breve här e ledig</i>
	Typos, or mispredictions	<i>Kan du specifisera dig?</i>
	Repetition of letters	<i>A, men göööör det dååå...grinpelle!!!</i>
	Repetition of words	<i>RÖKA RÖKA RÖKA RÖKA RÖKA</i> <i>RÖKA RÖKA RÖKA RÖKA RÖKA</i> <i>RÖKA RÖKA RÖKA RÖKA RÖKA</i> <i>RÖKA RÖKA RÖKA RÖKA RÖKA</i> <i>RÖKA RÖKA RÖKA RÖKA RÖKA</i>
	Split compounds	<i>Er statistik sida funkar inte</i>
<i>Grammatical features</i>		
	<i>Reduced sentences</i>	
	Subject pronoun	<i>[Jag] kom precis på att jag har beställning öppen är det säkert att stänga den</i>
	Verb phrase	<i>[Ska vi gå på]Lunch eller??</i>

Table 42, continued.

<i>Category</i>	<i>Feature</i>	<i>Example</i>
<i>Logotypes</i> <i>ASCII characters</i>	Emoticons Asterisks Symbol replacing word Addressivity marker	<i>Kom igen nu..... =8]</i> <i>Men hallååååå?! *skak på huvudet*</i> <i>Hälsningar X & Y</i> <i>*lazze => tråååååååkit*</i>
<i>Lexical features & abbreviations</i> <i>Lexical features</i>	Colloquial lexicon (dialect, expletives) Code switching	<i>d va la synd då..</i> <i>svenska akademien kan ta sej i röva... ;o)</i> <i>nopes, förutom att lägg till vara/ta bort behöver ej finnas med..</i>
<i>Abbreviations</i>	OCM features from spoken language Conventional abbreviations Unconventional abbreviations	<i>öööööööööööhhhhh??</i> <i>Missade dina BÅDA mezz!!!</i> <i>Hur står det till där nere i Gbg?</i> <i>e d där ni e..... Ska jag komma upp lr? Har en ledig burk här men den är iofs bokad</i>

Instant Messaging – Findings in relation to three dimensions

Instant messaging in WebWho is synchronous in that the sender had awareness that the receiver is logged in and would receive the message instantly. Whether the receiver was actually located by the computer which he or she was logged in on in that very instant could not be deduced from the interface alone, though. Results show that awareness of both physical presence, i.e. when both when sender and recipient share the same room, and virtual presence, mediated via the awareness tool, affect the messages and that these factors affect the text differently. The material also shows that the purpose of the interaction, the nature of the medium and the written mode affect the content of the messages.

Similar to email and web chat, instant messaging is a tertiary means of expression. Contributions were typed on computer keyboards and read as text on computer screens. Evidence that users investigated and tested the

range of usage for the system (e.g. trying to send messages from within the lab to people outside the lab, or telling long rambling stories to find out whether there was a limit on the message size) show that they tried out the technical limitations of this particular means of communication.

The situational parameters of instant messaging correlate with the means of expression. The system was designed to support collaborative assignments by allowing users who were distributed in different rooms, perhaps even on different levels in the building to communicate instantly. Messages between people who were located within the same room were often playful and non-task related, while communication between distributed users was often coordinating in nature, dealing with work or location related issues. The possibility to send messages either with identified or anonymous sender also affected what the messages were about. The content of messages with anonymous senders often dealt with pranks and mischief, while messages between identified communicators often dealt with task related issues and social coordination, such as deciding where to have lunch.

8.3.7 Chapter 7: Text Messaging via Mobile Phones – SMS

Chapter 7 comprised a study of text messaging via mobile phones, SMS. Data were collected from anonymous contributions to a web-based questionnaire, from known informants who forwarded their messages and kept notes of their communication in a user diary. Additional material was gathered from friends and relatives.

Syntactical and lexical reductions are characteristics of written language in all four modes of CMC that have been investigated in this dissertation. Results show that reductions are particularly common in text messaging via mobile phones. The production and perception conditions of text messaging are the most constrained of the four modes of communication investigated in this dissertation. The foremost particulars in the conditions of text messaging via mobile phones are the tiny keypad with its few and small keys, the small screen and limitations on message size. These conditions elicit the most interesting linguistic consequences.

Table 43. A taxonomy of linguistic features characteristic of SMS.

<i>Category</i>	<i>Feature</i>	<i>Example</i>
<i>Space, case, punctuation & Spelling</i> <i>Space</i>	Omitting blank space between words	<i>BANKOMATJÄVELNFUNKANTESOMVNLIT KOMMERSENARE</i>
	Omitting punctuation	<i>Ge mig ditt nummer har det inte längre så ringer ja ikväll ska spela fotboll</i>
	Unconventional punctuation	<i>Nu har jag blivit moster!!! kram, X</i>
<i>Case</i>	All lower-case	<i>ha det</i>
	All capitals	<i>TACK. OCH JAG HÅLLER MED!</i>
	Mix of lower-case and capitals	<i>UtanBaraÖkatOchJagÄlskarDigMer&MerFastänDuÄrSurPåMig.MenVillDuInteHaMigSå..SåBlirJagJätteledsen [...]</i>
<i>Spelling & punctuation</i>	Unconventional, spoken-like spelling	<i>DET FINNS EN PÅSE ME KLÄR KAN DU TVÄTTA DEM? LIGGER NGNSTANS I RUMMET, MEST STRUMPER Å SÅNT</i>
	Typos, or mispredictions	<i>Har också båda [badat] sommarens första bad. JAG & ANDERAS [ANDREAS] SPELAR KL. 5</i>
	Repetition of letters	<i>Luft heter det vadå då jag har inget att säga till dej emma snackar la inte heller me dej å hon e inte sur på dej såå</i>
	Consonant writing	<i>JA BKIS [BAKIS] PÅ TÅGET. FAST INTE MER ÄN ATT JAG KAN SKRATTA ÅT</i>
	Split compounds	<i>Det gick skit bra med Berra!</i>
<i>Grammatical features</i> <i>Reduced sentences</i>	Subject pronoun	<i>hej! förlåt mitt sega svar, [jag] har varit försjunken i flyttkoma.</i>
	Verb phrase	<i>[Ska vi gå på] LUNCH?</i>
	Exchange long words for shorter	<i>Vet ej [inte=not] ?</i>
	Preposition or possessive pronoun	<i>JAG HOPPAR [ÖVER DET]. FRASA PÅ</i>

Table 43, continued.

<i>Category</i>	<i>Feature</i>	<i>Example</i>
<i>Logotypes</i> <i>ASCII characters</i>	Emoticons	<i>hej min prins! har du drömt söta drömmar om mig i natt ;-) puss.</i>
	Asterisks	<i>Hej!Men VAR ska vi va o fotas?SVARA!!Annars kommer ja att ga vilse!!Sen hittar ja inte er o blir inte fotad!!*snyft* Men, vi ses iaf!! Kram</i>
	Symbol replacing word	<i>Tack för tipset! Köpte du 1 [ett] par? jag har inte råd just nu..</i>
<i>Lexical features</i> <i>& abbreviations</i> <i>Lexical features</i>	Colloquial lexicon (dialect, expletives)	<i>Hej lillan! Ja gör itte ett skit! Durå? Ja vi ska la va mä varann imorn? Puss puss</i>
	Code switching	<i>Jag skulle gärna prata mer.. Men jag orkar knappt hålla upp telefonen.. Ska ju upp te skolan imorn åxo! :(Sov så gott *en godnattkram* Bye.. :o)</i>
	OCM features from spoken language	<i>Hmm, kanske det!? Vi hörs i morr'n</i>
<i>Abbreviations</i>	Conventional abbreviations	<i>VILL DU TRÄNA NGT IMORN F.M.? JA GÅR T GYMMET ÅTMINSTONE.</i>
	Unconventional abbreviations	<i>hallå x, hur e d? vad händer i helgen. ska vi ta en fika? ha d bra! Vi cs. p&k/y men ffs har w8at en bra stund nu kommer du lr?.</i>

SMS – Findings in relation to research questions

Syntactical and lexical reductions are characteristics of the language in SMS. The most common syntactical reduction is deletion of subject pronoun, which gives the messages an informal, “telegraphic” style, much like informal spoken language. The same pattern was reported in studies of text messaging in Germany (Döring 2002a). Several types of abbreviations exemplified lexical reductions. Traditional, established abbreviations were used, as well as new, innovative ones. These new types of abbreviations

show similar patterns to those used in chat rooms, and the use of these in SMS is probably a transfer from chat norms.

The written language in SMS is used and adapted according to the characteristics of the means of expression and its conditions on production (multi tap or predictive text entry on the tiny keypad of a mobile phone), situation (relation between communicators – most messages are sent between friends who already know each other well).

It differs from the norms of traditional written language in that it is reduced and displays spoken language features. Specific “e-style” characteristics similar to those found in web chat and instant messaging were discovered in SMS messages, for example emoticons, unconventional abbreviations and non-normative use of punctuation (exclamation- and question marks, etc.).

SMS – Findings in relation to three dimensions

Mobile text messaging is an asynchronous mode of communication. There are both technical and pragmatic reasons for the occurrence of syntactical and lexical reduction. The technical restrictions on production and perception conditions influence this particular means of expression. Messages were mostly created either with multi-tap technique or predictive text entry on the tiny keypad of a mobile phone; entry takes physical effort and is time consuming. There is also a limit to the size of each message, and the text has to be composed with this in mind. All users have the same type of limitations on message size, causing an advantageous side effect of being able to avoid awkward conversations and get straight to the point. The full message size was seldom used to its full advantage. Mean message length was 14.8 words, or 64 characters; well below the upper limit of 160 characters.

As we have seen, most messages are sent between friends who already know each other well. This relationship between communicators, and the accompanying shared background knowledge, allow communicators to be inexplicit, use community-specific slang and abbreviations.

8.4 Comparisons among the sub studies

This section will compare the findings among the four different studies. Overall quantitative findings suggest that the activities that communicators are involved in are reflected in the most frequent words used. The more synchronous the mode, the more features it shares with spoken face-to-face interaction.

8.4.1 Word frequency, mean length of utterance, vocabulary richness

Table 44 below shows a comparison of the most frequent words across different modes of communication. Asynchronous modes used in this study were email, traditional letters that were typed or written by hand, SMS and writing. Synchronous modes were speech⁸³, web chat and instant messaging. All modes of communication represented in the table below are written, except speech. The most frequent spoken words are rendered in their written equivalent, supplemented by an index number or letters in brackets for disambiguation. The second most frequent spoken word, see Table 44 below, [e₀] stands for [är=is, are] and disambiguates it from [e₁] which indicates a hesitation sound. In the same way, [å₀] in the table below stands for [och=and], which disambiguates it from [å₁] which indicates [att=to]. The word [ja{g}=I] is disambiguated from [ja=yes].

The most frequent token is the same in five out of seven modes, namely the written language feature period [.]. That is not very surprising, after all the modes that have been investigated are, all but speech, written. Interestingly enough, even though web chat is a sort of written conversation, the most frequent token is the word *du* [you], and not the period. This fact reflects the activity in which the communication takes place, which in the case of web chat is the situational influence of social-recreational chatting. It seems that the norms of standard writing are not as important in this situation. It also shows the importance of topic, or perhaps goal of interaction, in the situation: the words *du* [you] and *jag* [I] are what people talk about (cf. Ko 1996; Yates 1996).

⁸³ Data from written and spoken language were based on the spoken language corpus at the Department of Linguistics, Göteborg University, while data from traditional written letters were based on the corpus gathered for the study on email (vs. typed or hand written letters to the city council).

Table 44. Comparison of most frequent words across different modes of communication.

<i>Asynchronous</i>				<i>Synchronous</i>		
<i>Email</i>	<i>Trad. letters</i>	<i>SMS</i>	<i>Writing</i>	<i>Speech</i>	<i>Web chat</i>	<i>IM</i>
.	.	.	.	de{t}	du	.
i	att	!	,	e ₀	?	!
på	och	?	och	å ₀	jag	,
att	i	,	i	ja{g}	,	?
och	,	du	att	att	är	du
jag	som	på	det	så	.	det
är	för	i	en	ja	det	på
som	en	jag	som	som	...	är
till	till	det	på	vi	som	och
det	är	är	är	inte	!	jag
för	av	att	med	på	inte	att
Jag	på	har	för	man	vill	inte
om	det	Jag	av	i	i	i
en	har	inte	jag	då	på	har
har	med	vi	den	du	har	...
'	inte	och	han	en	att)
Göteborg	de)	inte	ju	och	med
med	om	om	till	dom	då	till
Hej	ett	till	--	men	med	/
av	den	en	var	den	Jag	Hej

Word frequency reflects the activities that communicators were involved in. Informal interaction among peers, such as in SMS, web chat and instant messaging show similar patterns to spoken interaction. It is natural to talk about *du* [you] and *jag* [I] in these types of communication, whereas these are not as likely to occur at the same rate in autonomous traditional writing or in email messages to unknown authorities.

The level of synchronicity of the mode of communication is another important aspect of how contributions are formulated. Communication in synchronous modes, such as spoken interaction, real-time chat and instant messaging, does not allow participants to select expressions leisurely or to pack information densely the way asynchronous modes do. To keep up with the flow of conversation in a web chat, participants have to be brief and quick. Communication in asynchronous modes (e.g. email and SMS) allows senders to compose messages carefully if they like, especially in the case of SMS, where there is an upper limit to the message size leading to very dense messages.

An examination of mean length of contribution⁸⁴ across modes suggests once more the influence of synchronicity on communication. The mode that has greatest time pressure is web chat, and the messages in that particular mode were found to be the shortest, comparable only with speech.

Table 45. Comparison of Mean Length of Contribution (MLC) across modes of communication.

<i>Synchronicity</i>	<i>MLC</i>	<i>Mode</i>
<i>Synchronous</i>	9.24	Web Chat
	12.54	Speech
	13.45	IM
<i>Asynchronous</i>	14.77	SMS
	63.71	Email
	433.43	Traditional letters

Mean length of contribution, or mean length of message as would be the case in written communication, refers to the mean number of tokens per message. The web chat had the shortest contributions. It seems likely that this is due to the fact that this mode has the highest time pressure. Length of contributions is also affected by number of participants in the chat – the more participants, the shorter the messages (cf. Cherny 1999). The mode that, at least theoretically, had the least time pressure was traditional letters. Messages in traditional letters were the longest, almost 7 times the length of email messages. The corpus of written language was based on written texts from books, articles in newspapers, and such, making the concept of contribution somewhat problematic if comparing mean length of utterance with the rest of the modes. A letter, be it handwritten or typed, is limited in length compared to a novel, for instance. That is the reason that MLU (or MLC) with respect to writing based on the corpus of written language, was left out.

Even though SMS is asynchronous, message length is constrained by production conditions (typing on keypad, as well as limited message size). Pragmatic, situational reasons also influence message composition; in the case of the SMS corpus the messages were sent between people who were friends who shared considerable background information, allowing them to be inexplicit.

⁸⁴ Mean length of utterance (MLU) is used for analysis of spoken language (cf. Allwood 1999).

Table 46. Comparison of vocabulary richness between synchronous and asynchronous modes of communication.

<i>Synchronicity</i>	<i>Theoretical vocabulary richness</i>	<i>Mode</i>
<i>Synchronous</i>	2,035	IM
	2,128	Speech
<i>Asynchronous</i>	2,231	Web Chat
	2,330	SMS
	2,654	Email
	2,952	Traditional letters
	3,677	Writing

Theoretical vocabulary richness is another measure which is interesting in respect to level of synchronicity, and that was comparable across modes of communication. The formula of theoretical vocabulary calculates the expected number of types in a text of a certain length L , by calculating the mean number of types for all possible ways to select these L word tokens from the original longer text. In principle, it picks a random number of words, for example 100, from a corpus and then counts the number of types within these 100 words. Since the subset of words could be picked in many different ways, the theoretical vocabulary is calculated as the mean of the number of types for all possible ways of choosing the 100 words from the corpus (cf. Grönqvist 2000). Simply put, theoretical vocabulary richness is measured through type/token ratio, or how many different types of tokens were found among the total number of tokens, that is, how varied the vocabulary is. Writing has been shown to have richer vocabulary than spoken language (cf. Wengelin 2002).

The asynchronous modes of written language and traditional letters were richest in vocabulary; these showed a more varied vocabulary. Email was also asynchronous, and scored rather high. This also suggests that communicators in asynchronous modes have time to plan and revise, while in synchronous modes contributions have to be transmitted more rapidly and leave less time for revision. The activity in which the communication takes place and the relationship that exists between participants are also important issues. If the activity is informal, for instance, there might be no demand placed on using normative spelling or grammar. Normative language use might even have quite the opposite effect. Communicators adapt language according to instrumental rational principles in relation to parameters such as synchronicity, means of expression, and situation.

8.4.2 Generally compared findings

As was discussed above, there is a complex set of finely interwoven variables that influence communication. Written language, in the four modes of CMC that were investigated in the present study, was found to be adapted to the particular settings of each respective mode. Language use is adapted according to level of synchronicity, the particular conditions for production and perception in each means of expression, as well as according to the communicative situation and context.

To re-address the questions raised at the beginning of this dissertation: In what ways were the modes of CMC under investigation found to differ from the norms of traditional written language? Which were the written vs. spoken language features in each mode respectively, and why was this so? Synchronous CMC was hypothesized to exhibit more features of spoken language than asynchronous CMC. The study comprised two synchronous modes (web chat and instant messaging) and two asynchronous modes (email and SMS). Results have revealed that synchronicity plays an important part in how language is used. The more synchronous a mode of communication is, the more likely it is that the sender will get feedback by which to adapt further contributions.

Results also reveal that synchronicity alone cannot account for how language is used. Production and perception conditions, which are related to situation and its technical constraints, are shown to have an important influence on message composition. It was hypothesized that unlimited buffer size would result in more edited or careful language with features characteristic of traditional written language, than in modes with limited buffer size. This is another issue which is related to technical conditions of a given mode of communication. The study comprised two modes with unlimited buffer size (email and instant messaging), and two modes with limited buffer size (web chat and SMS).

Linguistic features in CMC

Email, an asynchronous mode with unlimited message size, was hypothesized to display edited and careful written language. To some extent it does, but the messages were much shorter, more informal than traditional letters and conformed less to the norms of standard written language. These results suggest that the informal situation and relative anonymity of an electronic means of expression are more important than the synchronicity aspect and message size.

Table 47. Comparison of linguistic features characteristic of CMC across modes.

<i>Category</i>	<i>Feature</i>	<i>Mode of CMC</i>			
		<i>Email</i>	<i>Web chat</i>	<i>IM</i>	<i>SMS</i>
1. Space, case, punctuation & spelling					
<i>i. Space</i>					
	<i>a. Omitting blank space between words</i>	-	+	+	+
	<i>b. Omitting punctuation</i>	+	+	+	+
	<i>c. Unconventional punctuation</i>	+	+	+	+
<i>ii. Case</i>					
	<i>a. All lower-case</i>	+	+	+	+
	<i>b. All capitals</i>	-	+	+	+
	<i>c. Mix of lower-case and capitals</i>	-	+	+	+
<i>iii. Spelling & punctuation</i>					
	<i>a. Unconventional, spoken-like spelling</i>	+	+	+	+
	<i>b. Typos, or mispredictions</i>	+	+	+	+
	<i>c. Repetition of letters</i>	-	+	+	+
	<i>d. Repetition of words</i>	-	+	+	-
	<i>e. Consonant writing</i>	-	-	-	+
	<i>f. Split compounds</i>	+	+	+	+
2. Grammatical features					
<i>i. Reduced sentences</i>					
	<i>a. Subject pronoun</i>	+	+	+	+
	<i>b. Verb phrase</i>	+	+	+	+
	<i>c. Preposition or possessive pronoun</i>	+	+	-	+
	<i>d. Exchange long words for shorter</i>	-	-	-	+
<i>ii. Word order</i>					
	<i>a. Inspiration from English (word order, prepositions)</i>	-	+	-	-
3. Logotypes					
<i>i. ASCII characters</i>					
	<i>a. Emoticons</i>	+	+	+	+
	<i>b. Asterisks</i>	-	+	+	+
	<i>c. Symbol replacing word</i>	+	+	+	+
	<i>d. Addressivity marker</i>	-	+	+	-
4. Lexical features & Abbreviations					
<i>i. Lexical features</i>					
	<i>a. Colloquial lexicon (dialect, expletives)</i>	+	+	+	+
	<i>b. Code switching</i>	-	+	+	+
<i>ii. Abbreviations</i>					
	<i>c. OCM features from spoken language</i>	-	+	+	+
	<i>a. Conventional abbreviations</i>	+	+	+	+
	<i>b. Unconventional abbreviations</i>	+	+	+	+

Web chat, a synchronous mode with limited message size, was hypothesized to display more features associated with spoken language. The results prove the hypotheses to be true in this case. Web chat contributions are often short, with a syntactic structure similar to informal spoken interaction. They frequently display informal use of emoticons and new forms of abbreviations, which suggest that the activity of social-recreational chat and the synchronous demands of rapid interaction are important variables.

Instant messaging is a synchronous mode of communication with unlimited message size. The synchronous mode was hypothesized to invite spontaneous, spoken-like language use, and the unlimited size would encourage written-like characteristics. The study suggests, rather, that language use in instant messaging is most influenced by awareness of presence accessed through the technical affordances of the mode itself, situational parameters such as location, status (identified vs. anonymous) and relationship between communicators, as well as activity or goal of interaction. Generally, language in this type of instant messaging is shown to be spontaneous and spoken-like; normative spelling and grammar are apparently not a priority.

SMS is an asynchronous mode with limited message size. Asynchronicity was hypothesized to invite editings and formal features, the limited size was supposed to invite spoken-like features. It turned out that both asynchronicity and limited size allow and require careful editing. The resulting language use displays many characteristics of spoken language, though. These results suggest that situational parameters such as relationship between communicators, and goal of interaction, together with parameters associated with means of expression, such as production and perception conditions are more important.

Results show that a number of unconventional, or non-standard written language, linguistic constructions appear frequently in the material that was investigated. Syntactical constructions that are normally found in spoken interaction are found in written CMC, such as deletion of subject pronoun or verb phrase. Lexical features that are normally associated with spoken interaction, such as OCM and ICM features, dialectal words and expressions have been found as well.

The email data do not show iteration of letters or repetition of words or code switching. Email messages from the corpus do not show any OCM or ICM features, which is what could be expected from an asynchronous and monologic mode of communication. The data show only one example of emoticons, and only a very low number of unconventional abbreviations.

A plausible explanation for these patterns is that this type of email communication is the mode which was used for the most formal activity, communicating with some unknown authority in order to request help or information.

Web chat is the only mode in which the addressivity marker (e.g. Raggaren>>>>>Lisa) is used. That this feature only is found in the web chat is easily explained; this is the mode which is most synchronous, with multiple interlocutors competing to make themselves heard in the textual noise.

Instant messaging shows no features that seem to be specific to the particular mode of communication. Rather, it shares all features with web chat, except for the addressivity marker mentioned above.

Features that are specific for language use in SMS are consonant writing, and typos created because predictive text entry predicted a word that was not intended by the sender. This is also the mode in which omission of space between words is most frequently used. This feature is naturally connected with the limited buffer size as well as the difficulty of text input.

8.5 Linguistic Adaptivity theory

8.5.1 Instrumental rationality

Was Biber right? Is genre the most important influencing factor to account for the variations between speech and writing? Which is the more important variable, genre/activity or means of expression? This is still an open question, which has not been pursued in this thesis, since this study does not examine spoken language.

The present study has shown that means of expression is the more important variable because, however we twist and turn, text-based CMC has to be written in order to be transmitted. The production conditions for text entry vary among the modes that have been investigated, and so does the level of synchronicity. The more synchronous the mode, the more time pressure is exercised on the speed of typing and transmission.

On the other hand, activity is the more important variable because of the situations in which the messages are sent: relationships between communicators, and goals of interaction. It does matter whether communicators know each other, and what the nature of their relationship is. It also matters what they communicate about, whether it is to say hello or to complain.

As we have seen, the variables in this study are interdependent and cannot easily be separated. Synchronicity is interlinked both with the means of expression, what the technical conditions for the mode of communication are, as well as with the situation. The relationship between communicators and the particular goal of the message have a definite influence on why people choose to use a certain mode of communication over another.

How language is used is a matter of several reasons in a complex set. One variable may be more important in one setting, and another may be just as important in another. However, it is interesting to see how language use varies with the variables suggested here, and to see how people use and adapt writing to fit the setting in which they find themselves.

This thesis proposed a “linguistic adaptivity theory” – language behavior is relative to the set of variables put forward in this dissertation, and relative to rational behavior which is part of human ability for communication. Rationality correlates with situation or activity, goals and aims of communication, medium for communication, etc. Humans have the ability to be flexible and adapt their rational behavior according to the numerous variables that condition communication.

As part of rational behavior, strategies to minimize time and effort spent in message production, and in some cases also strategies to save space, play a part in explaining the changes of language that occur in CMC. When do we minimize? It is related to goals of the interaction in an activity, and (adequately) adapted to suit them. What is adequate in each particular case is relative to the effects of constraints and enablements brought in by the variables of synchronicity, means of expression and situation. Clark’s principle of least effort is not an absolute principle, and can easily be overruled by other principles and goals that the communicators find more important at the moment. This is in turn related to different levels of intentionality, as discussed by Allwood (1995; 2000). Humans are not absolutely conscious of every move they make; some behavior is not the result of immediately conscious choices.

8.5.2 A genuinely new medium?

This dissertation challenges the claims put forward by Crystal (2001), that CMC is a genuinely new medium. What can the four modes of CMC “add” in comparison to speech and traditional writing that cannot be explained by genre/activity or the differences between means of expression (speech and writing)? Much of what is communicated via email could just as well

have been accomplished with traditional letters. Factors such as ease of production with its low demand on physical and cognitive effort, immediacy and availability, fast transmission at a low cost and relative anonymity of electronically mediated text all invite informality and perhaps also communication which would not have been conducted if it had to rely on pen-and-paper communication. The mediated, electronically transmitted character of tertiary means of expression allows users to easily attach files that complement communication per se, such as text, audio, and video files to email messages. Hyperlinks referring to other information on the web could easily be included. These things could also be attached or included in traditional letters, but not as conveniently.

Synchronous web chat is one of the modes which allows for interaction which is difficult, if not impossible, to accomplish in other modes of communication. In a real-time web chat participants communicate in interactive written discourse with multiple, often anonymous and unknown, participants. The written contributions remain on the screen to be read and reread until they scroll out of sight. Because conversation is persistent (cf. Herring 1999) in this way, it is easier to keep track of and maintain multiple threads of interaction that would be more difficult in spoken interaction.

Instant messaging allows users to tailor their written messages according to their awareness of the receiver as being present and available for communication, much like spoken interaction. It allows for instant written interaction among people who are distributed in space. Instant messages are delivered immediately, topmost on the receiver's screen, allowing for rapid interaction.

Mobile text messaging allows users to communicate from anywhere at anytime. It allows for silent communication that may be delivered almost instantly, and at a much lower cost than voice calls.

Crystal's claim that communication may be conducted in ways that neither spoken nor written interaction allows (e.g. the persistence of written conversation and synchronous, written multiparty conversation in a real-time chat) must be acknowledged; however, the features that these modes of CMC show are not genuinely new - the combination is perhaps new – otherwise they are just variants adapted to fit the situation. The possibility to hide behind the anonymity of text-only communication, and experiment with online personas in social roles and language use that we witness in numerous types of CMC interaction, seems to foster a kind of informality. This would not be impossible in traditional written communication. The differences seem to lie, rather, in the fact that there are new arenas of

interaction: more activities in which interaction is accomplished through text, than were possible before.

The new fora of interaction with their evolving informal stylistic norms have fostered new ways of expressing oneself in writing. We can perhaps go as far as to claim that the practices described here enhance alphabetic writing by attempting to convey more than is normally possible to express in text, the use of emoticons to indicate facial expressions, reactions and attitudes. The technique is not exactly new; similar attempts to express extra information have been used by employing certain patterns in the placement of stamps on envelopes or in handwritten notes among teenagers in school, for example.

8.5.3 Is language deteriorating?

Popular fears are sometimes heard claiming that language is deteriorating and that it is deteriorating at an even higher rate due to being used in CMC. Throughout this dissertation language use in CMC has been seen instead as evidence of the uniquely human ability to adapt language according to the interdependent variables of a given situation (context, activity, relation to communicators), level of synchronicity, means of expression (writing vs. speaking, production and perception conditions of the means of expression). Language uses in CMC are variants of use, just like any other language use. And it is adapted to suit the conditions of the particular setting in which it occurs. One variant does not necessarily exclude other variants, and we do not lose the ability to adapt because we communicate via CMC. On the other hand, in order to have modes to vary between, we have to at least know what are considered to be the norms of written language in different situations. Hopefully, we will continue to be taught in school, among other institutions of learning, to use language in different situations, for varying purposes. Teaching methods have to take into consideration language varieties, and that what is “correct” and “appropriate” in one setting might be the reverse in another. Education professionals also have to be aware of the fact that norms of language use change over time, and today’s norms will eventually become outdated. Language changes, it is just not necessarily something bad.

8.6 The main contributions of the dissertation

This dissertation has investigated use and adaptation of written language in CMC. How Swedish written language is used in these particular modes of CMC has not been studied previously in this respect. The results from the studies in this dissertation show that a number of non-standard features of written language are being used. A taxonomy of linguistic features of the modes of CMC that were investigated was proposed. Furthermore, the results suggest that there is an interesting transfer of norms and language use in CMC – both between modes of CMC and between languages that users observe in international CMC fora. Language use in international CMC seems to inspire Swedish users to create new types of abbreviations based on Swedish words in analogy with abbreviations based on other languages, mainly English. Emoticons and other types of graphical expressions as observed in international chat rooms for example, are frequently used in new modes of communication. Features that are normally associated with spoken interaction are found in language use in informal activities, such as sending SMS messages to friends or social-recreational chat. These modes display OCM features from spoken language, for instance.

This dissertation challenges the popular view that language is deteriorating and that it is deteriorating at a faster rate because of CMC. The studies here show that language is neither deteriorating nor are modern users less able to use it successfully now than in the past. What is good and successful use is related to the concept of instrumental rationality, and a theory of linguistic adaptability is proposed here. This thesis has shown that these still rather new modes of communication have fostered new norms of expression and developed innovative ways of using written language. Language use in CMC has been proven to be adapted to suit the particular conditions according to level of synchronicity, means of expression and situation in the respective communicative settings. Demands from time pressure and limitations in messages size, among several factors, constrain language use in CMC; communicators employ creative devices in an attempt to convey more meaning than normative written language does as a rule. Use and adaptation of language in the modes of CMC that have been investigated in this dissertation is evidence of and exemplifies human ability to adapt in general, and the uniquely human ability to adapt linguistic behavior specifically.

8.7 Implications of results

Results from the email communication investigated in this study show that email invites an informal tone in addressing “one’s superiors”. The immediacy and ease of interaction might lead users to neglect checking their messages for conforming to norms of spelling and such, as well for content before sending. That might just as well happen in traditional letters. A negative implication of this is that people are in many cases judged according to their language use. Rapidly written messages that are full of misspellings and perhaps not yet established abbreviations or emoticons might give the receiver a certain impression of the sender, and the content might not be taken seriously, despite the receiver’s actual knowledge. A positive implication of this particular type of email is that it probably invites people to contact an unknown authority at the city council in the first place, which at the same time may be a negative effect judging from some rather reckless messages. Generalizations of all forms of email communication can not be inferred from the single study of electronically transmitted letters in this particular setting, but has shed light on how a tertiary means of expression might influence language use.

Language use in a recreational real-time web chat and in instant messaging very much reflects the activity in which it is employed. The spoken-like quality of contributions is aptly tailored to the rapid and informal interaction. Whether there will be a transfer and establishing of norms from chat room communication and in instant messaging to other domains of written interaction remains to be seen.

Regarding text messaging, it was concluded that in order to be successful, predictive text input software has to be adapted to the communicative situations, needs and uses of text messaging in reality. The results of this dissertation reveal that language use in SMS is much more like spoken interaction both in terms of structure and lexical choice. This implies that in order to function effectively and purposefully, predictive text input software needs to be based not on frequency lists based on newspaper text, but on language use in SMS primarily and spoken language secondarily.

Those abbreviations that are regarded as being new today, found in the corpora analyzed in this dissertation, may be transferred into written language that is aimed for other purposes than computer-mediated communication. The process of splitting compound words might be enhanced by the difficulty that the predictive text software has in recognizing them. Language is always in a constant state of change, but some changes might be less desirable. Change in norms of spelling and

syntactical structure is one thing, enhanced splitting of compound words is another. Splitting a compound word entails a potential semantic change which perhaps is not always intended by the writer. If the practice becomes widespread enough, the Swedish language might lose some of its characteristics and one means of expressing semantic nuances. The adaptive view proposed in this dissertation would hypothesize that people would find other ways to express what needs to be expressed, but perhaps this type of change is in a way graver than other changes.

8.8 Future Research

There are still things to be said about the modes of interaction that were analyzed in this study. It would still be challenging to go on with further investigations of material from the same modes of communication, perhaps with slightly different production and perception conditions. Split-screen synchronous chat in which both participants are able to view the messages, as they are created, character-by-character, for instance, is an area worth studying. It would also be interesting to pursue studies of other modes of CMC, for example the importance of non-verbal signals (eye contact, for instance) in video conferencing systems, the ethics of SMS-marketing, how MMS and EMS are going to affect communication. Another project that will be pursued is an investigation of how different types of text input influence the writings of children and teenagers. The project will also analyze whether there seems to be a transfer between norms of written language in CMC to writing in other domains, such as school assignments.

Luckily, there seems to be no end of the ever evolving field of communication technologies, and it will be a great and fascinating challenge to go on exploring human communication in future projects.

References

- Ahlsén, E. (1993). "On agrammatism and telegram style - An exemplifying discussion of Swedish data." Nordic Journal of Linguistics **16**(2).
- Allwood, J. (1976). Linguistic communication as action and cooperation : a study in pragmatics. Dept. of linguistics. Göteborg, Univ. of Göteborg.
- Allwood, J. (1995). An Activity Based Approach to Pragmatics. Göteborg, Dept. of Linguistics.
- Allwood, J. (1997). Dialog as Collective Thinking. Brain, Mind and Physics. P. Pylkkänen, P. Pylkkö and A. Hautamäki. Amsterdam, IOS Press.
- Allwood, J. (1998). The Structure of Dialog. Structure of Multimodal Dialogue II. M. M. Taylor, F. Ne'el and D. G. Bouwhuis. Amsterdam, John Benjamins.
- Allwood, J., Ed. (1999). Talspråksfrekvenser - ny och utvidgad upplaga. Göteborg, Gothenburg Papers in Theoretical Linguistics.
- Allwood, J. (2000). An Activity Based Approach to Pragmatics. Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics. H. B. Bunt, B. Amsterdam, John Benjamins: 47-80.
- Allwood, J., Ed. (2000). Talspråksfrekvenser. Gothenburg Papers in Theoretical Linguistics. Gothenburg.
- Allwood, J. (2002). One-word utterances. Göteborg University.
- Allwood, J., J. Nivre, et al. (1993). "On the Semantics and Pragmatics of Linguistic Feedback." Journal of Semantics **9**(1).
- Allwood, J. e. (1986). Annual Report on Feedback, ESF-project EALA. University of Göteborg, Dept of Linguistics: Chapters 3,4,6 & 7.
- Androutopoulos, J. and G. Schmidt (2001). "SMS-Kommunikation: Etnografische Gattungsanalyse am Beispiel einer Kleingruppe." Zeitschrift für Angewandte Linguistik.
- Austin, J. L. (1962). How to do things with words. Cambridge, Massachusetts, Harvard University Press.
- Baron, N. S. (1984). "Computer mediated communication as a force behind language change." Visible Language **18**(2): 118-141.
- Baron, N. S. (2000). From Alphabet to Email. London, Routledge.
- Baym, N. (1995). The emergence of community in computer-mediated communication. Cybersociety: Computer-mediated communication and community. S. Jones. Thousand Oaks, CA, Sage: 138-163.
- Baym, N. K. (1996). "Agreement and Disagreement in a Computer-Mediated Discussion." Research on Language and Social Interaction **29**(4): 315-345.
- Biber, D. (1988). Variations across speech and writing, Cambridge University Press.
- Bloomfield, L. (1933). Language. Chicago, University of Chicago Press.
- Bolinger, D. (1981). "Consonance, Dissonance, and Grammaticality: The Case of Wanna." Language and Communication **1**: 2-3.
- Bolter, J. (1991). The Computer, Hypertext and the History of Writing. Hillsdale, NJ, Lawrence Erlbaum Associates.
- Bolter, J. D. (1991). Writing Space: The Computer, Hypertext, and the History of Writing, Lawrence Erlbaum.

- Brenner, S.-O. (1975). Formal structure of messages and discourse related to personality. *Psychology*. Göteborg, Göteborg University.
- Bybee, J. and J. Scheibman (1999). "The Effect of Usage on Degrees of Constituency: The Reduction of Don't in English." *Linguistics* 4(362): 575-596.
- Chafe, W. (1986). Writing in the perspective of speaking. *Studying writing: Linguistic approaches*. C. Cooper, R & Greenbaum, Sidney. Beverly Hills, London, New Delhi, Sage Publ. Inc. 1.
- Chafe, W. (1988). "Punctuation and the Prosody of Written Language." *Written Communication* 5(4): 395-426.
- Chafe, W. (1994). Functional Sentence Perspective in Written and Spoken Communication (see LLBA 27/1, 9301056).
- Chafe, W. and J. Danielewicz (1987). Properties of spoken and written language. *Comprehending oral and written language*. D. Tannen, Academic Press, Inc.
- Cherny, L. (1999). *Conversation and Community: Chat in a Virtual World*. Stanford, CSLI Publications.
- Cho, N. (Forthcoming). Linguistic Features of Electronic Mail. *Computer-Mediated Conversation*. S. C. Herring. ?, ? : ?
- Churchill, E. and S. Bly (1999). *It's all in the words: supporting work activities with lightweight tools*. GROUP'99, New York, ACM Press.
- Clark, H. H. (1996). *Using Language*, Cambridge University Press.
- Clark, H. H. and S. E. Brennan (1991). Grounding in Communication. *Perspectives on Socially Shared Cognition*. L. B. L. Resnick, John M & Teasley, Stephanie D. Washington, DC, American Psychological Association: 127-149.
- Collot, M. and N. Belmore (1996). Electronic Language: A New Variety of English.
- Cook, V. (1997). *Inside Language*.
- Crystal, D. (2001). *Language and the Internet*. Cambridge, Cambridge University Press.
- Daft, R. L. and R. H. Lengel (1984). "Information richness: A new approach to managerial behavior and organization design." *Research in Organizational Behavior* 6: 191-233.
- Dahl, Ö. (2001). Inflationary processes in language and elsewhere. *Frequency and the emergence of linguistic structure*. J. a. P. H. e. Bybee. Amsterdam, John Benjamins: 471-80.
- Danet, B. (1998). Text as mask: Gender and identity on the Internet. *Cybersociety 2.0*. S. J. . Thousand Oaks, CA, Sage.
- Danet, B. (2001). *Cyberpl@y : Communicating Online*. Oxford/New York, Berg.
- Danet, B., L. Ruedenberg-Wright, et al. (1997). "'HMMM...WHERE'S THAT SMOKE COMING FROM?'" Writing, play and performance on Internet Relay Chat." *Journal of Computer-Mediated Communication [On-line]* 2(4).
- December, J. (1996). "Units of Analysis for Internet Communication." *Journal of Computer-Mediated Communication* 1(4).
- Diamond, J. (1992). *The third chimpanzee - the evolution and future of the human animal*. New York, HarperCollins Publishers, Inc.
- Donath, J. (1995). *Visual Who: Animating the affinities and activities of an electronic community*. ACM Multimedia '95, ACM Press.
- Dourish, P. and V. Bellotti (1992). *Awareness and Coordination in shared Workspaces*. Conference of on Computer-Supported Cooperative Work (CSCW'92), New York, ACM Press.

- Du Bartell, D. (1995). "Discourse Features of Computer-Mediated Communication: 'Spoken-Like' and 'Written-Like'." Anglicana Turkuensia **14**: 231-239.
- Döring, N. (2002a). ""Kurzm. wird gesendet" - Abkürzungen und Akronyme in der SMS-Kommunikation." Muttersprache. Vierteljahresschrift für deutsche Sprache **2**.
- Döring, N. (2002b). "1x Brot, Wurst, 5sack Äpfel I.L.D. - Kommunikative Funktionen Von Kurzmitteilungen (SMS)." Zeitschrift für Medienpsychologie **3**.
- Eldridge, M. and R. Grinter (2001). Studying text messaging in teenagers. CHI 2001 Workshop #1: mobile Communications: Understanding User, Adoption and Design, Philadelphia, USA.
- Erickson, T., D. Smith, et al. (1999). Socially translucent systems: social proxies, persistent conversation, and the design of "Babble". CHI'99, ACM Press.
- Eubanks, P. (2001). "Understanding metaphors for writing: In defense of the conduit metaphor." National Council of Teachers of English.
- Ferrara, K., H. Brunner, et al. (1991). "Interactive Written Discourse as an Emergent Register." Written Communication **8**(1): 8-34.
- Fidelholtz, J. L. (1975). "Word Frequency and Vowel Reduction in English." Papers from the Regional Meetings, Chicago Linguistic Society **11**: 200-213.
- Fowler, C. A. and J. Housum (1987). "Talkers' Signaling of "New" and "Old" Words in Speech and Listeners' Perception and Use of the Distinction." Journal of Memory and Language **26**(5): 489-504.
- Garcia, A. C. and J. B. Jacobs (1999). "The eyes of the beholder: Understanding the turn-taking system in quasi-synchronous computer-mediated communication." Research on Language and Social Interaction **32**(4): 337-367.
- Godefroid, P., J. D. Herbsleb, et al. (2000). Ensuring Privacy in Presence Awareness Systems: An Automated Verification Approach. ACM 2000 Conference on Computer Supported Cooperative Work, Philadelphia, Pennsylvania, USA.
- Gregory, M. L., W. D. Raymond, et al. (1999). The effects of collocational strength and contextual predictability in lexical production.
- Grice, H. P. (1957). "Meaning." Philosophical Review **78**: 377-388.
- Grice, H. P. (1975). Logic and conversation. Syntax and semantics. P. M. Cole, J. L. New York, Seminar Press. **3: Speech Acts**: 41-58.
- Grinter, R. E. and M. A. Eldridge (2001). y do tngrs luv 2 txt msg. ECSCW2001, Bonn, Germany.
- Grönqvist, L. (2000). The TraSA v0.8 User's Manual. A user friendly graphical tool for automatic transcription statistics. Dept. of Linguistics, Göteborg University.
- Hale, C., Ed. (1996). Wired style : principles of English usage in the digital age. San Francisco, Calif., HardWired.
- Halliday, M. A. K. (1985). Spoken and written language. Oxford, Oxford University Press.
- Herring, S. C., Ed. (1996). Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives. Amsterdam, John Benjamins.
- Herring, S. C. (1996). Two Variants of an Electronic Message Schema.
- Herring, S. C. (1998). "Le style du courrier électronique: variabilité et changement." Terminogramme **84-85**(March): 9-16.
- Herring, S. C. (1999). "Interactional coherence in CMC." Journal of Computer-Mediated Communication **4**(4).

- Herring, S. C. (2000). "Gender Differences in CMC: Findings and Implications." Computer Professionals for Social Responsibility Newsletter **18**(1).
- Herring, S. C. (2001). Computer-Mediated Discourse. Handbook of Discourse Analysis. D. Tannen, D. Schiffrin and H. Hamilton. Oxford, Blackwell: ?
- Herring, S. C. (2002). "Computer-Mediated Communication on the Internet." Annual Review of Information Science and Technology.
- Herring, S. C. (Forthcoming). "A Classification Scheme for Computer-Mediated Discourse."
- Hoffman, J. (2002). Wireless evolution insider, GSM Association: 15.
- Horowitz, M. W. and A. Berkowitz (1964). "Structural Advantage of the Mechanism of spoken expression as a factor in differences in spoken and written expression." Percept. Motor Skills **19**: 619-25.
- Hymes, D. (1974). Foundations in Sociolinguistics - An Ethnographic Approach. Philadelphia, University of Pennsylvania Press.
- Hård af Segerstad, Y. (2000a). "A study of a Swedish chat room." M/C Online - A Journal of Media and Culture **3**(4).
- Hård af Segerstad, Y. (2000b). "Electronic Letters to the City Council – factors influencing the composition of email messages." Human IT(2-3/2000): 179-207.
- Hård af Segerstad, Y. (2001). Instant Messaging and Awareness of Presence in WebWho. In: Conference proceedings from, Mobilizel, Windsor, UK., Digital World Research Centre, University of Surrey.
- Hård af Segerstad, Y. (Forthcoming). Swedish Teenagers' Written Conversation in Electronic Chat Environments. WebTalk - writing as conversation. D. Penrod. Mahwah, NJ, Lawrence Erlbaum Associates.
- Hård af Segerstad, Y. and P. Ljungstrand (2002). "Instant Messaging with WebWho." International Journal of Human-Computer Studies, Special Issue on "Awareness and the WWW"(56).
- Jespersen, O. (1922). Language. London, Allen.
- Karlsson, A.-M. (1997). Kallprat på nätet. Om chattares kommunikativa kompetens och samtalsstil. Svenskan i IT-samhället. O. Josephson. **Ord och stil** **28**: 138-163.
- Kasesniemi, E.-L. and P. Rautiainen (2002). Mobile Culture of Children and Teenagers in Finland. Perceptual Contact - Mobile Communication, Private Talk, Public Performance. J. E. Katz and M. Aakhus. Cambridge, Cambridge University Press: 170-192.
- Kiesler, S., J. Siegel, et al. (1984). "Social Psychological Aspects of Computer-Mediated Communication." American Psychologist **39**(10): 1123-1134.
- Ko, K.-K. (1996). "Structural Characteristics of Computer-Mediated Language: A Comparative Analysis of InterChange Discourse." Electronic Journal of Communication/La revue électronique de communication **6**(3).
- Lea, M., T. O'Shea, et al. (1992). 'Flaming' in computer-mediated communication. Observations, explanations, implications. Contexts of Computer-Mediated Communication. M. Lea. Hertfordshire, Harvester Wheatsheaf: 89-112.
- Levelt, W. J. M. (1981). The speaker's linearization problem. Philosophical Transactions of the Royal Society [Series B]. London. **295**: 305-315.
- Levinson, S. C. (1983). Pragmatics. Cambridge, Cambridge University Press.

- Lewis, C. T. and C. Short (1963). A Latin Dictionary, Oxford University Press.
- Linell, P. (2001/in preparation). The Written Language Bias in Linguistics: Its Nature, Origins and Transformations. Linköping, Department of Communication Studies.
- Ljungstrand, P. (1999). WebWho: support for student awareness and coordination. ESCW'99, Copenhagen, Denmark.
- Ljungstrand, P. and Y. Hård af Segerstad (2000). An analysis of WebWho: How does awareness of presence affect written messages? Proceedings of the Awareness and the World Wide Web workshop at Computer-Supported Cooperative Work (CSCW) 2000, Philadelphia, PA, USA.
- Mann, C. and F. Stewart, Eds. (2000). Internet Communication and Qualitative Research: A Handbook for Researching Online. London, Sage Publications Ltd.
- Maynor, N. (1994). The Language of Electronic Mail: Written Speech? Centennial Usage Studies. G. D. Little and M. Montgomery. Tuscaloosa & London, University of Alabama Press. 78: 48-54.
- McCarthy, J. and E. Meidel (1999). ActiveMap: A Visualization Tool for Location Awareness to Support Informal Interactions. HUC '99, Springer-Verlag, Berlin.
- Mitsuoka, M., S. Watanabe, et al. (2001). Instant Messaging with Mobile Phones to Support Awareness. Symposium on Applications and the Internet: SAINT 2001, San Diego, CA, New York: IEEE Press.
- Murray, D. E. (1987). Discourse Structure of "Conversation for Action": A Cross-Media View. The 8th World Congress of Applied Linguistics, Sydney, John Benjamins publ. comp. 1990.
- Nardi, B., S. Whittaker, et al. (2000). Interaction and Outeraction: Instant Messaging in Action. CSCW'2000, ACM Press.
- Nivre, J. and U. Richthoff (1988). Återkoppling och turtagning i två typer av samtal. En jämförande studie av samtal ansikte mot ansikte och telefonsamtal. Göteborg, Dept. of Linguistics.
- Orwell, G. (1949). Nineteen Eighty-Four. London, Martin Secker & Warburg.
- Paolillo, J. (1999). "The virtual speech community: Social network and language variation on IRC." Journal of Computer-Mediated Communication [On-line] 4(4).
- Pedersen, E. R. and T. Sokoler (1997). AROMA: Abstract Representation Of presence supporting Mutual Awareness. CHI'97, ACM Press.
- Pinker, S. (1994). The language instinct. New York, William Morrow and Company.
- Reddy, M. J. (1979). The conduit metaphor. Metaphor and Thought. A. Ortony, Cambridge University Press: 284--324.
- Reid, E. M. (1991). Electropolis: Communication and Community on Internet Relay Chat. Department of History. Melbourne, University of Melbourne.
- Rintel, E. S. (2001). "First Things First: Internet Relay Chat Openings." Journal of Computer-Mediated Communication 6(3).
- Sachs, H., E. A. Schegloff, et al. (1974). "A Simplest Systematics for the Organization of Turn-taking in Conversation." Language 50: 696-735.
- Saussure, F. d. (1916). Course in general linguistics. London, Duckworth.
- Schlobinski, P., N. Fortmann, et al. (2001). Simsen. Eine Pilotstudie zu sprachlichen und

- kommunikativen Aspekten in der SMS-Kommunikation. Hannover, Networx Nr. 22.
- Schönfeldt, J. (2001). Die Gesprächsorganisation in Der Chat-Kommunikation. Chat-Kommunikation. Sprache, Interaktion, Sozialität und Identität in synchroner computervermittelter Kommunikation. M. Beißwenger. Stuttgart.
- Searle, J. (1969). Speech Acts. Cambridge, Cambridge University Press.
- Severinson Eklundh, K. (1994). Electronic Mail as a Medium for Dialogue. Functional Communication Quality. L. van Waes, E. Woudstra and P. van den Hoven. Amsterdam/Atlanta, Rodopi Publishers.
- Severinson Eklundh, K. E. (1987). Dialogue Processes in Computer-Mediated Communication: A Study of Letters in the COM System: Dissertation Abstracts International, A The Humanities and Social Sciences, 1987, 47, 12, June, 4235-A.
- Shortis, T. (2001). The Language of ICT. London, Routledge.
- Sinclair, J. M. and J. Ball (1996). EAGLES Preliminary Recommendations On Text Typology. Birmingham, Corpus Linguistics Group, School of English, Birmingham University.
- Smith, M., C. J.J., et al. (2000). Conversation Trees and Threaded Chats. CSCW'2000, ACM Press.
- Soukup, C. (2000). "Building a theory of multimedia CMC. An analysis, critique and integration of computer-mediated communication theory and research." New media & Society 2(4): 407-425.
- Sperber, D. and D. Wilson (1986). Relevance: Communication and Cognition. Oxford, Blackwell.
- Sproull, L. and S. Kiesler (1991). Connections: New Ways of Working in the Networked Organization. Cambridge, MA., MIT.
- Stenson, P. (1997). Skriva i tjänsten. Stockholm, Timbro.
- Storrer, A. (2001). Sprachliche Besonderheiten getippter Gespräche: Sprecherwechsel und sprachliches Zeigen in der Chat-Kommunikation. Chat-Kommunikation. Sprache, Interaktion, Sozialität und Identität in synchroner computervermittelter Kommunikation. M. Beißwenger. Stuttgart: 3-24.
- Sveningsson, M. (2001). Creating a Sense of Community: Experiences from a Swedish Web Chat. The TEMA Institute, Dept. of Communication Studies. Linköping, Linköping University: 250.
- Sveningsson, M. (2003). Ethics in Internet Ethnography. Virtual Research Ethics: Issues and Controversies. E. A. Buchanan. Hershey, Idea Group Publishing.
- Telia/Temo (2002). SMS-undersökning.
- Tollmar, K., O. Sandor, et al. (1996). Supporting social awareness @work - design and experience. CSCW'96, ACM Press.
- Turkle, S. (1995). Life on the Screen: Identity in the Age of Internet. New York, Simon & Schuster.
- Wakeford, N. and N. Kotamradju (2002). Mobile Devices and the Cultural World of Young People.
- Wallace, P. M. (1999). The psychology of the Internet, Cambridge : Cambridge University Press.
- Walther, J. B. (1996). "Computer-Mediated Communication: Impersonal, Interpersonal and Hyperpersonal Interaction." Communication Research 23(1): 3-43.
- Wengelin, Å. (2002). Text Production in Adults with Reading and Writing Difficulties. Dept. of Linguistics. Göteborg, Göteborg University: 331.

- Werry, C. C. (1996). Linguistic and Interactional Features of Internet Relay Chat. Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives. S. C. Herring. Amsterdam, John Benjamins: 47-63.
- Wilkins, H. (1991). "Computer Talk. Long-Distance Conversations by Computer." Written Communication 8(1): 56-78.
- Williamsson, S. (2001). Svensk telemarknad 2001, Post och Telestyrelsen.
- Wilson, D. and D. Sperber (1987). "An Outline of Relevance Theory." Notes on Linguistics 39: 5-24.
- Wirén, E.-L. (1998). Vad är en allmän handling?, Luleå Tekniska Universitet. 2002.
- Witmer, D. and S. Katzman (1997). "On-line smiles: Does gender make a difference in the use of graphic accents?" Journal of Computer-Mediated Communication [On-line] 2(4).
- Wolfson, N. (1989). Perspectives: Sociolinguistics and TESOL. New York, Newbury House publishers.
- Wolz, U., J. Palme, et al. (1997). Computer-mediated communication in collaborative educational settings, ACM Press.
- Yates, S. J. (1996). Oral and Written Linguistic Aspects of Computer conferencing. Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives. S. C. Herring. Amsterdam, John Benjamins: 29-46.
- Zipf, G. K. (1932). Selected Studies of the Principle of Relative Frequency in Language. Cambridge, Massachusetts, Harvard University Press.
- Zipf, G. K. (1949). Human Behavior and the Principle of Least Effort. Cambridge, Massachusetts, Addison-Wesley Press, Inc.