

Pirkko Raudaskoski

**THE USE OF COMMUNICATIVE RESOURCES IN
LANGUAGE TECHNOLOGY ENVIRONMENTS
A conversation analytic approach to semiosis
at computer media**

Academic dissertation

1999

P i r k k o R a u d a s k o s k i

**THE USE OF COMMUNICATIVE RESOURCES IN
LANGUAGE TECHNOLOGY ENVIRONMENTS
A conversation analytic approach to semiosis
at computer media**

**Academic dissertation to be publicly discussed, by due permission of the Faculty of
Humanities, at the University of Oulu in the Kaljunen auditorium (Kaljusensali,
KTK112, on Saturday, May 22, 1999, at 12**

Copyright © 1999
Pirkko Raudaskoski

ISBN 951-42-5248-9
Åbo Akademis tryckeri
1999

*To my parents
Anna-Liisa and Heikki
Raudaskoski*

Abstract

The dissertation examines how the varying semiotic possibilities of different language technology environments shape the communicative interactions that take place with or through them. The everyday life of the citizens of information societies is increasingly saturated with encounters with mediated 'others' such as language-using computer systems or printed manuals to help use computer systems. Also, the Internet has enabled the use of inexpensive video conferencing with sound and/or text to communicate through technology. In order to explore how these modern day communicative environments affect the interpretative process of interaction, four case studies were conducted, in which various communicative resources (sound, electronic and paper-based text, graphics, animation, and video picture) were available for participants to interact with or through a computer system. Thus, each setting had different semiotic resources and posed differing constraints on how the interpretative work could proceed. Not only is the aim of the study to uncover the impact of the material surroundings on the ongoing interaction, but also to show that conversation analysis can be used as a method to analyse not only copresent (or telephone) conversations, but also encounters in which the 'other' is linguistically or visually represented on a screen or on a page. In this way, the social semiotic conceptualisation of the effect of the material surroundings on the semiosis can also be empirically validated.

Even though it was originally developed to study talk-in-interaction (between human participants), the methodology used to analyse the semiotically complex encounters is conversation analysis (CA). According to CA, sense making, people's interpretative work (semiosis), takes place in a sequential fashion; indeed, every contribution to linguistic interaction is both context-sensitive and context-renewing. The conversation analytic idea of sequential interpretation is applied also to the Peircean concept of endless semiosis to show that the cognitive idea of interpretant can be 'externalised' at the level of a turn. The externalising of semiosis provides the analyst with a tool that links the meaning-making process to the social semiotic idea that commodities and material formations are also signs.

The dissertation shows that in the research and design of language technologies, the situation of use has to be oriented to as primarily sequential interpretation, in which the visual layout of the technology also plays an important role. Thus, the research demonstrates that CA is an effective method to find out exactly how the communicative resources affect the interpretative work. The participants' formulations of existing linguistic resources provide a means for the analyst to understand the significance of the overall communicative situation for how the activity is accomplished and for how the 'other' is oriented to. The data analysis focuses particularly on directives, or requests to act. The instructions accentuate the idea of sequential interpretation, because the 'other' is directing the participant(s)'s actions, often step-by-step. Repair initiation, a type of request in itself by the 'other' or by the participants, problematises prior interpretations, and makes the potential interpretative uncertainty explicit. The different semiotic constellations of the case studies result in particular possibilities for repair initiation. The research shows that in technology-mediated interactions, both repair and repair-oriented speech acts, such as warnings, have to be distinctive enough to guarantee a successful activity. In general, the division into informative and action-instigating language is not always easy to maintain during the encounter, and this is partly because the guiding texts tend to be written as narratives.

Foreword

In the mid-80s I participated in a vocational training course on the use of data processing and became a trainee in the English Department of Oulu University. This dissertation is a result of the academic journey which began with that position. I stayed on in the English Department, and soon became involved in its research activities. I would like to thank Professor Heikki Nyysönen not only for his contribution as my supervisor, but also for obtaining funding from the Finnish Academy and other sources to finance research projects in the Department. Those, together with Heikki Nyysönen's international orientation, have created important opportunities for myself and other doctoral students to become acquainted with the academic community and research from all over the world. From that wider community my special thanks go to Professor David Silverman for his critique of an early draft of some sections of the dissertation. The manuscript was officially read by Dr. Tove Arendt Rasmussen and Dr. Jack Whalen; their valuable comments will also be kept in mind for future research.

Although this dissertation has been on my agenda ever since I finished my licentiate degree, the work really got started in August 1996 when Professor Nyysönen appointed me to be a researcher in the Finnish Academy project 'The Design and Accessibility in English Texts'. That five month period gave me the possibility to concentrate solely on my doctoral work. The previous project researchers, Tuija Isomursu and Risto Pilto, had collected video data on manual use, which inspired me to do a case study on that topic. In Spring 1997, I received a three months' research grant from the English Department, and, in September 1997, I started a ten month Finnish Academy supported foreign visit to the Communication Department of Aalborg University in Denmark. I am deeply grateful to the English Department and the Finnish Academy for their financial support.

In both Oulu and Aalborg I have been lucky to work with remarkable people, whose support has been of great importance. To mention only some, Leena Kuure's encouraging comments often coincided fortuitously with my times of self doubt in my dissertation project, and Elise Kärkkäinen never finished informing me about my own skills. The helpful secretaries, Liisa Bozkurt and Riitta Rajaniemi, kept me going with their humorous remarks.

One colleague is worthy of a special mention: my dear partner Paul McIlvenny, who has enriched my private and professional life with his love and enthusiasm. He has reassured me to take many a step I thought I would not be able to do, be it descending from a dangerous cliff in Scotland, crossing a stream in Ecuador, or finishing up my doctoral dissertation in Denmark.

I also thank my six siblings and other friends for their care and share.

Aalborg, Spring 1999
Pirkko Raudaskoski

Contents

ABSTRACT	
FOREWORD	
1. INTRODUCTION	9
1.1. Background to the present study	14
1.1.1. Use of communicative resources:	
Interactivity and participation at language technologies	16
1.1.2. Intersubjectivity, intention and the role of the 'other'	18
1.1.3. Situated language use and the 'other'	19
1.1.4. Language technologies as communication environments	22
1.1.5. Instruction giving: language as/for social action	25
1.2. Method of inquiry	26
1.2.1. Qualitative research	29
1.3. The case studies	33
1.3.1. The research material	35
1.4. General outline of the dissertation	36
2. ASPECTS OF AND APPROACHES TO RESEARCHING ENCOUNTERS WITH 'OTHERS'	38
2.1. The 'other' as text/discourse	39
2.1.1. Another look at the discourse context	40
2.1.2. Design features of written communication	42
2.1.3. Reading as social action	43
2.2. The 'other' as a social entity	46
2.2.1. Ethnomethodology/Conversation Analysis and the importance of repair work	48
2.3. The 'other' as a sign	55
2.3.1. Social semiotics from an interactional perspective	61
2.4. My approach: Analysing semiosis as materially-based sequential meaning-making	65
2.4.1. Time / Space / Language	70
2.5. Summary	73
3. INTRODUCTION TO THE CASE STUDIES	75
3.1. Experimental studies	75
3.2. The case studies as encounters with 'others'	77
3.2.1. Creating intimacy	79
3.2.2. Visual and interactional sharing	80
4. TELEPHONE: THE USE OF COMMUNICATIVE RESOURCES IN ENCOUNTERS WITH AN EXPERIMENTAL TELEPHONE DIALOGUE SYSTEM	84
4.1. Channelling the user through spoken dialogue	86
4.2. Repetition and reformulation	88
4.3. Use of pronouns	91
4.3.1. Avoidance of use of pronouns	92
4.4. The user and sequential meaning making	93
4.4.1. Contesting/yielding to the 'other'	96
4.5. Summary	98
5. TUTORIAL: (CAUSES OF) MISUNDERSTANDINGS IN TEXT-USER INTERACTION	102
5.1. Introduction	102
5.2. The electronic text of the case study	104
5.3. Adjusting interpretation: repair work	105
5.4. The semiotic systems of the setting	106
5.4.1. Visuality and materiality	110
5.5. Data analysis	111
5.5.1. <i>Click-L</i>	112

5.5.2. Potential hazards created by the materiality of the text	123
5.5.3 To do or not to do	124
5.5.4. <i>Click-L</i> revisited	128
5.5.5. Encountering an 'other'	129
5.5.6. Summary	131
5.6. Discussion	132
6. TASK: BETWEEN READING AND ENCOUNTERING IN HUMAN TEXT INTERACTION	134
6.1. Introduction	134
6.2. Background	139
6.2.1. Affordances and constraints	139
6.2.2. Text as a requester (instigator of user's actions)	140
6.2.3. Request as part of interaction	142
6.3. The case study TASK	145
6.3.1. Analysis of data	145
6.4. Discussion	165
6.4.1. Shape of repair work	167
6.4.2. Reading vs. encountering	168
6.5. Conclusion	169
7. TEACHING: THE USE OF COMMUNICATIVE RESOURCES IN INTERNET VIDEO CONFERENCING	175
7.1. Introduction	175
7.2. Background	177
7.2.1. The CU-SeeMe video conferencing system	180
7.3. Interactionist approach	181
7.4. The data	182
7.4.1. CU-SeeMe as MeSee(what)UC	183
7.5. Data analysis	185
7.5.1. General observations	185
7.5.2. Orienting to the incoming and outgoing video pictures	187
7.5.3. Typing on the video picture: the synchronous/asynchronous division in the making	201
7.6. Conclusion	216
8. DISCUSSION	220
8.1. Introduction	221
8.2. The interactionally salient features of the four activity systems	225
8.2.1. TELEPHONE	226
8.2.2. TUTORIAL	228
8.2.3. TASK	230
8.2.4. TEACHING	231
8.2.5 Situated interpretation	232
8.3. Knowledge <i>that</i> vs. knowledge <i>how</i>	235
8.4. Results concerning uncertainty and repair work	237
8.4.1. Human-'other' interaction	240
8.4.2. Human-human interaction	242
8.4.3. Summary	244
8.5. The interactionally constituted 'other' and assumptions of intentionality and sharing as prerequisites of meaning negotiation	245
8.5.1. Creating mutuality	247
8.5.2. Aspects of 'other' (as a sign) foregrounded/made relevant in CMC interaction	249
8.6. Summary	257
9. CONCLUSION	262
REFERENCES	269
APPENDICES	

1. INTRODUCTION

In the industrialised world, people's everyday life tends to be saturated with encounters; not only with other people, but with texts and objects as well, and in both official and private circumstances (Watson 1997). Meeting face-to-face is increasingly replaced (or avoided) by technology-mediated human-human, or human-computer interaction. Public computer systems form the backbone of the Information Society. Biocca calls the expansion of computer technologies into the daily realm 'social integration': "Social integration means that the interface is being integrated into everyday activity at work, home, and on the street" (Biocca 1997, 12). For the Information Society to be a successful project, there is a need for not just increasing bandwidth, but increasing knowledge of the built systems as communicative (language based) environments.

Undoubtedly, the increasing sophistication of interactive programs (i.e. responding to the user's action or expecting a response from the user) and more readily usable interfaces have paved the way for the expansion of new technology. In computerised communication spaces, the notions of 'text', 'object' and 'other' (whether 'real' or 'virtual') often become amalgamated. For example, a tool to help a user's actions may deploy spoken or written language; automatic telephone services, instructional and professional computer programs, and printed manuals can steer the user's real-world activities. Though produced in the past, the language can appear to the user in an interactive fashion, often giving directives to the user-reader about their task-at-hand.

The dream of building computers to understand and participate in everyday conversation has not been realised. However, the Internet, for instance, offers a medium of communication between people in which one-way information, and asynchronous (not shared temporally) or synchronous (shared temporally, but not spatially) interactions take place all over the computerised world via interfaces that use text, pictures and even sound and video. These language technologies offer arenas for encounters that cannot be categorised as either symbolic or instrumental, thus blurring the division between language and tool use (cf. Engeström & Middleton 1996).

Some researchers (e.g. Talbott 1995) see a danger in the increasing interactivity, and therefore apparent humanness, of information technologies: people either take the computers around them to have human features, or people are talked about in computational terms. In spite of this warning, information technologies are being developed to increase the possibilities for the user to be a part of the ongoing activity within the domain in question. For instance, multimedia packages merging text, sound, and (video) picture offer a seemingly interactive communication environment, designed by teams of 'senders', which are intended to serve a large number of local users. Also, everyday utility programs, such as word processors, incorporate interactive features in the interface which means that they can guide the user more effectively (cf. the animated 'wizards' of Microsoft). User friendliness, i.e. designing for the users' communicative expectations, is also important for so-called computer dialogue systems such as general information systems usable via telephone or via a terminal (Cole et al. 1995). The present work will offer some insights, based on detailed data analysis, on which aspects of 'humanness' (vis-à-vis language use) can be endangered in the increasing use of language technologies. The problem is not the use of language as such, but some aspects of language use that the interaction process can 'dehumanise'.

When language artefacts are built, the aim of the design is hardly to produce programs that cannot be separated from humans. However, the 'technology' of turn-taking to build social events is such an integral part of human relations that it is hard to keep the two aspects of sense making separate. Thus, though it is not possible to build conversing machines, but only simulacra of interaction (Button & Sharrock 1995), an increasing number of encounters in the present society will be simulated on one part. Therefore it is important to research just how these social events are constructed, and also whether the human relations (cf. Streeck 1980) aspect is oriented to in the encounters. Even if the workings of artefacts are often a centre of attention at the cost of the social event (Whalen 1995, 162), the features of the material environment that become foregrounded and consequential for the interaction might be different from the designer's foci of attention. Therefore, the social aspect of the encounter is also essential for a better understanding of how the material artefact affects the interaction.

In the data analysis of this dissertation, encounters in language technology environments are examined as situations in which both stand-alone and Internet mediated artefacts (and the accompanying instructive texts) are used. Thus, the terms 'language technology environment' and 'computer media' are understood to cover all the relevant artefacts that are utilised. Language is approached as an integral part of socially organised human practice, hence the concentration on language use, rather than on language as an abstract (grammatical) system. Also communication is not regarded only to be about exchanging messages (efficiently), i.e. that "something is being made known to somebody" (Blakar 1992, 236). This is

because communication as use of language cannot be separated from making sense of the ongoing communicative situation. In the sense-making work, language plays a crucial role, but not as a closed system: it is interpreted in a sequence (within a sequence of other activities, as well), and it appears in a certain material communication environment in a certain material manifestation (voice, text). The language technology environment consists of the human participants, the artefacts and their mediated language, the spatial arrangement of artefacts and people, the linguistic and other experience that the human participants make relevant through talk and action in the situation, and the history of the ongoing interaction. Goodwin's 'interactive activities' can attend to all these elements (Goodwin 1996); 'context' would be too loose a term to describe the inseparable communication environment and its human actors in their changing participation frameworks (e.g. Goodwin & Goodwin 1992). The organisation of the interaction of the human participants at various language technologies is a complicated phenomenon which can give access for the researcher to the participants' moment-by-moment understandings of the technology (and its language) in question. The activity the participants are involved in provides the interpretive framework for talk and other action (Goodwin 1995a). There are two basic forms of organisation of directing somebody in their activity: either the language technology actively 'communicates' with the user-reader, or the user-reader selects which piece of text he or she should follow. Whether the language that contributes to the ongoing activity appears to the user 'interactively' or whether the user 'activates' it by choosing a textual sign as the relevant next in the activity, the question arises how this human-'other' interaction is organised as a language-using event.

In the case studies, conversation analysis is used as a method in the minute examinations of language use, participation and uncertainty at language technology. Below, the idea of sequential meaning making advocated by conversation analysis will also be paralleled with the Peircean idea of endless semiosis. Conversation analysis can also be used to analyse the impact of the material surroundings on the semiosis or meaning making theorised in social semiotics. Though ethnomethodology is interested in 'ethno methods', or the way people do other things such as conversation, the attraction of social semiotics is that it is interested not just in local methods but in the wider social and political side of human interpretation, and in commodities and material formations as signs. When encounters with artefacts such as manuals and computer programs are researched, the scope of interest is necessarily wider than "*individual-to-individual* relationships" that e.g. Bordieu criticises ethnomethodology of concentrating on (Bordieu 1977, 81); the language is often produced by groups of individuals and is meant to be 'consumed' by anybody. So, the potential recipients are anonymous for the author(s), and vice versa (e.g. company products seldom credit the individual(s) that designed them). But the texts are used in specific settings by

individuals. How do the user-readers deal with this Bakhtinian 'voice' that – through words on the paper, screen, or telephone – guide their practical activities? What does the change from a mere informative role to a successful instigator of action demand from mediating language? And does the turn to indexicality of instruction giving and taking change the text (or its carrier) towards an 'other', and if so, at what level of interaction? To answer these questions, interaction between human participants still is crucial as it offers a reliable source for investigating ongoing interpretations. In pair work situations, for example, constant observation of and accountability to the spatio-temporally copresent other human participant provides the researcher with the participants' moment-for-moment understandings of the instructions and other signs encountered.

Participation "is always based on situated negotiation and renegotiation of meaning in the world" (Lave & Wenger 1991, 51). In (technology) mediated communication environments, meaning making could also be regarded as a discovery procedure (Boden 1990, 200) that, in team work, for example, is negotiated with other human participants. Meaning making always requires interactional work that unfolds in time and space. Sequentiality of action comprises the temporal aspect of interaction, and the importance of visual in the semiosis involves the spatial orientation in human practice. Goodwin defines participation as "the embodied organization of the primordial categories for participants that constitute states of talk (types of speakers, hearers, characters within stories, etc.)" (Goodwin 1998, 1). The different kinds of sign systems that the participants instantiate in different semiotic media are called semiotic fields by Goodwin (in press). This dissertation will explore how the interpretation, meaning making, of text, voice and (video) picture in a (spatially and/or temporally) distanced communication environment varies when these linguistic and other signs are incorporated in an activity. Through detailed examination of participation and the use of different semiotic fields, the researcher can get access not only to how the human participants make sense out of their surroundings, but also cautious claims can be made about why the interpretation took certain trajectories.

The present research concerns both synchronous and asynchronous encounters, synchronous via video conferencing facilities, and asynchronous with computer programs and manuals. In the case studies, different situated activity systems are investigated to find out how the mediated instructions are interpreted during the course of action. Instructions can both inform (about action) and instigate action, thus reflecting the old division into knowledge *that* and knowledge *how* (Ryle 1975 [1949]). Written texts (e.g. in book format) could be considered as good information storages, and (spatio-temporally present) humans with their transient, embodied language practices as good advice givers. Because some of the 'talk' presented by predesigned computer programs is synchronously produced asynchronous text (i.e. the text is appearing as if in shared interaction, but in fact the text is a product of the

past), one of the interests of the present study lies in whether this interactive language as a technologically mediated entity is oriented to more as an 'other' rather than just as text: how is the 'other' constituted during the activity at hand.

The term computer-mediated communication (CMC) nowadays refers mostly to interactions on the Internet (Herring 1996). All parties are human, and the messages passing between people are produced, for instance, using an electronic mail program interface (normally asynchronous, one-to-one use), asynchronous many-to-many computer conferencing (Davis & Brewer 1997, Sorensen 1997), split screen programs, for instance Talk (synchronous, one-to-one use), or chat programs, (most of the time many-to-many because of the shared public access of the participants to everybody's contributions).¹ CMC, understood broadly, also covers aspects of human-computer interaction (HCI), in the sense that the designers can be understood to 'communicate' with the user. The prepackaged formulae of language are an important resource for people to get their work done. Talbott points out that: "even the "computer-human interface" people—who have contributed so much to our understanding of machine design—have failed to probe adequately the implications of the fact that we're really dealing with a human-human interface. Those were software engineers who designed that obstructive program you struggled with last week" (1995, 32–33). Thus, Talbott considers HCI to be a specific form of CMC. This dissertation is going to deal with Talbott's notion, however, not from the designers' but from the users' point of view: how do they understand the language (as text or 'other') that they are encountering on the interface. Hopefully the results will contribute to better understanding of the CMC and HCI practice and to recognising what consequences mediation has for situated activity systems (e.g. Goodwin in press) such as acting upon instructions.

Interaction is getting more popular as a research topic in the humanities. Everyday conversation, using literature, film, and videos; institutional encounters; and other spheres of people's lives are seen as worth of studying in authentic circumstances, because second-hand information (e.g. interviews or questionnaires) or mental models are representations which have lost the "Intersubjective Time" (Bowers 1991, 554). This dissertation sets out to examine interactions in language technology environments, making use of the knowledge and research methods that have been developed in the humanities in order to account for the *in situ* character of the technology-mediated communicative process in all its material and discoursal complexity.

¹ Murray (1988) calls the latter two 'e-messages' (electronic messages).

1.1. Background to the present study

Whether a person or persons at a computer are involved in communicating with other people over the Internet, or using a stand-alone system (for instance, a word processor) to accomplish something, they encounter a two dimensional screen with text, icons, and sometimes video and sound. The keyboard and mouse are part of this 'perceptual cyberspace' (Strate et al. 1997, 17). The frame of the screen creates the frame of reference for the users; it is their interpretive space. Texts, icons, or other signs appearing on the screen, and even the voice or sound heard from the loudspeakers, might have been prepackaged or recorded two years or two minutes before: the time lag, as well as the distance in space between production and consumption can vary. However, unlike with TV or cinema, the 'viewer' is actively interacting with the signs on the two dimensional screen, not just interpreting them *in situ*, in the time frame of the user or participant at the computer interface².

A common claim is that technologically mediated communication (and perception of the 'other') is distant (e.g. Boden & Molotch 1994) — an *a priori* allegation that needs to be researched, and also the somewhat contradictory claim made by some researchers that the mediated nature of typed computer-mediated communication adds to the feeling of a joint zone. Don Langham writes in relation to multi-user domains:

CMC systems are anything but transparent. At every stage of communicating, from encoding words on a keyboard to decoding hard-to-read text on a screen, the mediated nature of CMC continuously calls attention to itself—especially so when users attempt a "real time" conversation or "chat" via their keyboards. Yet, despite this difference (indeed, perhaps because of it), the CMC environment fosters a sense of place, a sense of a conversational space alien to most people's experience with telephony. Whereas telephony calls attention to the fact that the parties do not share the same space, CMC encourages the perception that their interaction occurs in a shared location.

(Langham 1994, 7)

Langham's quotation stresses the material surroundings of technology-mediated communication: contrary to the claims about the distancing nature of CMC, Langham regards the interplay between the material and the semiotic as enhancing the feeling of sharedness. The two dimensional screen is a mutual interactional space. The spatial and/or temporal distance, and the different material manifestations of communicative signs that the interface brings to the encounter do contribute to the interaction, but to get to the gist of

² Global distribution of software from companies like Microsoft means that the signs interpreted throughout the world are not locally produced. In a way, they provide an electronic example of a world-wide textually mediated social organisation (Smith 1984).

the impact on language use, detailed data analysis of actual encounters is needed.

Written texts and video pictures are two-dimensional material artefacts. However, even these signs are arranged on a three dimensional object. In the case of art, the artefacts can be divided into demand artefacts and offer artefacts (Kress & van Leeuwen 1996, 254): a sculpture can be viewed from different perspectives, and the transactional space (Kendon, 1992, 330), the spatial arrangement of the participants, can differ and change from demand to offer. In the case of language technology mediated communication which involves, for example, written text and video pictures, there is also an oscillation between demand and offer, as the textual object provides an environment for passive monitoring, but also (at times) requires participation – the user's contribution. The manifestations and implications of demand and offer artefacts as sites of interaction will also be addressed in the present work.

In linguistics, the materiality of language and the situations of language use have not been included in the analysis of language structures. Local material circumstances and language have traditionally been kept apart, even in research concerned with language use (e.g. pragmatics, discourse analysis, and sociolinguistics): language has been considered as a system on its own; often the social aspect of language use has meant spreading the focus of attention from one speaker to all the interactants' use of linguistic items as the sole contributors to meaning. In traditional sociolinguistics, 'social' has not meant 'local', because general linguistic trends of groups, rather than (perhaps shared) interpretive procedures of individuals have been the centre of attention. Even the discourse and conversation analytic fields, despite their differences, have concentrated primarily on language as a self-sufficient system, though gestures might have been accepted as a (secondary) part of meaning making in conversation. However, the ethnomethodologically and anthropologically informed conversation analytic research tradition has taken into account the material surroundings, from the gestures and facial expressions of the interlocutors to the constraints (and resources) that the tangible environment provides for the participants' actions and activities (C. Goodwin 1979, 1981, 1986a; Suchman 1987).³ There are other approaches, such as semiotics⁴, which consider there being a connection between language and all other signs (for instance Lévi-Strauss' linguistic approach to culture) or between language and society (Thibault 1991, Hodge & Kress 1988). The latter interconnection is of interest for the present thesis. But the impact of the

³ Schegloff's note on how "the way in which orientation to co-participants and interactional structure matter to discourse and its formation, will vary in different speech exchange systems with different turn-taking systems" (1981, 73), could be also regarded as acknowledging, though very abstractly, the material surroundings.

⁴ Nöth (1995) gives an exhaustive description of the history and present approaches of semiotics. Section 2.3 will provide a summary relevant for this thesis.

material conditions on the interpretive process has not been researched in semiotics, though there is a large body of investigation on the 'meaning' of architecture, layout, arts, and other nonlinguistic signs as separate, usually aesthetic fields. It is not easy to include the material surroundings in the dynamic meaning making process of interaction, and the bulk of research has concentrated on how the surroundings are understood as separate, continuous entities, not as part of human interaction, in which they or some aspect of them might fall in and out of focus. In the humanistic and social science fields, context of communication has become an acknowledged constituent to be taken into account in any research into human activities. Some computer-mediated communication studies also admit that context "is not simply a stable factor that conditions communication behaviour; instead, it is dynamic in both its nature and the character of its relationship with behaviour" (Fulk et al. 1992, 7). However, apart from the conversation analytic tradition, rigorous analyses of context-as-dynamic have been lacking from the enterprises stressing the importance of context.

In the present work the importance of the material social situation is acknowledged for the unfolding interaction and for how the 'other', the representational participant is oriented to. Sequences of actually occurring interaction are researched, in which the stable and dynamic aspects of language, text, and artefacts are incorporated in the spatio-temporal flow of action and understood from within that flux, rather than given *a priori* meanings.

1.1.1. *Use of communicative resources: Interactivity and participation at language technologies*

The cycle production-distribution-consumption of text is oftentimes connected with literature or other printed media which are material language objects usually read in isolation; they are not used in connection with practical activities. The read language is not produced, but consumed, and is not that of speech-in-action. However, in computer surroundings, it is possible to research the interplay between production and social interaction (cf. Thibault 1991); how language products are made use of and understood in the everyday surroundings, and how the predesigned meanings of these language products can change or be contested *in situ*.

Participation, using linguistic and other actions to interact with the 'other', and interactivity are intertwined. Laurel regards engagement in the activity as a crucial definition of interactivity in computer environments: "You either feel yourself to be participating in the ongoing action of the representation or you don't" (Laurel 1991, 20–21). Interactivity is a very powerful mechanism in human-human encounters, and the usability of artefacts is assumed to be improved with increased interactivity: the system

can guide the user more efficiently, instead of the user going through trial and error cycles.

The talking computer was an early research aim in artificial intelligence and related fields:

One of the most important and challenging areas of research in artificial intelligence is the design and construction of natural language dialogue systems. This task is one of the hardest artificial intelligence has to offer, since the overall goal is the implementation of a fully capable and fluent conversational partner.

(Levine 1990, 964)

However, computer systems lack almost all human conversational capacity. "What I can convey to you with absolute fidelity – although it is fidelity to no-content, nothing – is only the empty proposition of logic or equation of mathematics" (Talbot 1995, 307). But when the empty logic produces strings of words on the screen, the result is an appearance of interactive behaviour. Although interactivity in information technology is still far from human, it seems that it is exactly the ability to respond to the users' commands or actions, and to do this with language, that might make the computer more usable for novices. Thus, usability and interactivity tend to be considered as one and the same design aim. For instance the European Union Telematics Applications Language Engineering Programme is clearly geared towards developing voice and text based interactive dialogue systems to enhance the information retrieval of European citizens (see <http://www2.echo.lu/langeng/en/lehome.html>).

The potential of the system does not make it interactive as such, but it is the human participant that reads the meanings into whatever the computer or other media does or displays, and this work resembles what people do when encountering each other:

The significance of any action and the adequacy of its interpretation is judged indirectly, by responses to actions taken, and by an interpretation's usefulness in understanding subsequent actions. It is just this highly contingent process that we call interaction.

(Suchman 1987, 119)

Thus, the way to research how the 'other' is encountered in technology-mediated communication is to look at the users' interpretations of the 'other's' turns – be they linguistic or other actions: what the interlocutor is saying or doing is considered as being said or done on the basis of what was said or done so far. The alleged effect of incorporating interactivity into computer systems has been that computers are treated as partners rather than as machines. However, rather than jumping into conclusions like "The personal computer is in the process of becoming the interpersonal computer" (Leary 1990, 230), the question to ask is: how does people's interacting with "derived

images instead of *immediate experience*" (Talbot 1995, 358) shape the overall organisation of the interaction?

1.1.2. Intersubjectivity, intention and the role of the 'other'

Sharing, being mutually aware of what is going on in interaction, is usually called intersubjectivity. Schiffrin (1990) gives a good overview of the problems and issues that this concept brings in into linguistics. The debate is mostly about how much intersubjectivity has to or can be assumed for interacting people. This again depends on what type of intersubjectivity is at issue and what is presumed about the interactants: people can share goals, procedures and/or background knowledge; the communication can be about intentions or information. If intersubjectivity concerns 1) shared goals (intentions), 2) procedures, and 3) background knowledge, then though sharing intentions or goals and background knowledge can be expressed through language-in-use, only procedures are to do directly with the rules or norms of language.

Gumperz (1995) problematises sharing vis-à-vis contextualisation cues: linguistic code as a system is not enough to understand contextualisation cues; instead, experiences of language use need to be similar because "whereas grammatical knowledge is common to all speakers of a certain language or dialect, sharing of contextualization conventions cannot be assumed, since it presupposes shared communicative experiences" (Gumperz 1995, 105). Thus, background knowledge (experience) of procedures (language-in-use) is needed, making the division into the three subareas of intersubjectivity blurred, especially if actor and recipient intents are to be inferred on the basis of the indexical signs that the contextualisation cues are realised in. So, should intersubjectivity be assumed as given between the interlocutors, or should it be regarded as an emergent phenomenon (e.g. Graumann 1995)? Schiffrin (1990) supports the latter point of view. If communication is not considered to be primarily about finding out about the speaker's intentions (i.e. achieving a shared goal), but about the speaker's or actor's intent which is realised as information, and the recipient's intent which is realised as interpretation, then Schiffrin's position is also supported by the approach taken in the present work. It also resonates with Lave and Wenger's idea about intentionality as a flow of reflective moments which "is organized around trajectories of participation" (Lave & Wenger 1991, 54). According to Schegloff, the interpretive procedure of (talk-in-)interaction provides a "routine grounding for intersubjectivity" (Schegloff 1992, 1295), any breakdowns in which can be fixed by repair work.

If intersubjectivity between copresent interlocutors is based on the mechanism of turn-taking (and of repair), what then can be assumed as 'inter' between a user and language technology? In all the case studies of the present dissertation, the recipients' intent (interpretation) does not always match with

the intent of the 'other'. The users of computer media seem to orient to the preprogrammed entity as someone or something with limited abilities. Therefore, the human participants regard as shared only what they hear on the phone or see on the screen, both in regard to background knowledge and communicative procedures. The encounters are thus approached with interactive caution. But with traditional printed media, the user-reader has more interpretive freedom, as he or she can be in charge of the pace and direction of the 'interaction', which challenges the empowering effect of the user-friendly computer interface (cf. Laurel 1990).

Many preprogrammed entities that users encounter have specific uses. For instance, the overall 'intention' of automatic bank tellers and educational software is fairly obvious. It has to be kept in mind, however, that the 'other' should not be regarded as a stable entity which can be given a label on the basis of the function, such as a tutor. It is most probable that there is oscillation in how the 'other' is perceived, what he, she or it is constituted to be, and at what levels.

1.1.3. Situated language use and the 'other'

Interaction generally assumes an 'other', somebody (or something) to interact with and interactivity presumes active participation in the interaction; "self" and 'other' are two classes of participants in interactive social organizations – in particular those which characterize the sequential organization of conversation, specifically its turn-taking system" (Schegloff et al. 1977, 361-362). Thus, for example, question-answer and request-acceptance/rejection adjacency pairs are normally constructed by two participants rather than one. In these interactional sequences or structures, the linguistic item used to construct the 'other' can be for instance a pronoun; 'you', 'he/she', 'it', or 'them'.

There are computer systems that fulfil the requirements of interactivity and participation to some extent, with the resulting hopes and fears of them being considered human-like. Human-computer interaction is an interesting hybrid between the cultural practice of reading newspapers, letters, books and other textual formats that are mediating the message from a known or an unknown nonpresent producer, and use of a machine, which could be considered either as a mediator of a message or as the creator of it. However, according to Nass and Steuer, "new media technologies increasingly blur the distinction between source and channel" (1993, 523), and therefore "scholars must view the use of these technologies as seamlessly connected to the rest of our social, rather than our instrumental, lives" (1993, 523). Nass et al. (1994) have shown that users who claim they do not regard computers as human, however evaluate programs according to human standards (e.g. 'criticizers are smarter than praisers'). They come to the conclusion that the humanness of

computers is not a characteristic of the artefact, but, rather, only detectable in how the users align to the system. It could be assumed that when interactivity increases, so does the computer's humanness, as usually conversation or any other form of simultaneous interaction requires a concrete other participant (for a discussion of the machine as a dialogue partner, see e.g. McIlvenny 1990). Nass et al. investigate the topic by questionnaires the users fill in after a session with the computer. The present work, however, addresses the issue by inspecting how the users in actual encounters orient to the machine and its language.

If the material surroundings of the interactive process are consequential or a factor in how the communication is understood and shaped, then the manner of the presence of the other participant surely is of importance, as well. The mediated nature of language in technology environments indicates that there is either a synchronous or asynchronous or 'one way' interaction going on. Levinson (1988) introduces a categorisation of participation roles according to the presence or absence of the 'sender' (participatory or nonparticipatory), and the mediated role of the message. He makes a distinction between the production and reception roles of interlocutors, which are more complicated than the traditional 'speaker' and 'hearer' categories. According to Levinson's categorisation then, the potential for the user to understand the computer as a partner rather than a mediator indicates that the production roles of author (e.g. ordinary speaker) and relayer (participates and transmits what is said, but has no motive and does not decide the form of what is being said) would be mixed up at the reception end: the computer might be perceived as an interlocutor rather than as an intermediary.

The single quotes around the term 'other' emphasise that the 'other' is not necessarily a known 'other', but a virtual somebody or something that is constructed in the situation and by which oneself becomes a subject, the 'one'. Volosinov (1973, 86) describes how language is always recipient designed:

Orientation of the word toward the addressee has an extremely high significance. In point of fact, *word is a two-sided act*. It is determined equally by *whose word* it is and *for whom* it is meant. As word, it is precisely *the product of the reciprocal relationship between speaker and listener, addresser and addressee*. Each and every word expresses the 'one' in relation to the 'other'. I give myself verbal shape from another's point of view, ultimately, from the point of view of the community to which I belong. A word is a bridge thrown between myself and another. If one end of the bridge depends on me, then the other depends on my addressee.

In the present dissertation, the formation of the bridge is examined from the point of view of the addressee, as well: how he or she constitutes the 'other' of the addresser, how he or she interprets the sender(s) in the material surroundings of the emerging situation. For instance, in an experiment in which a hypothetical telephone answering machine was tested (see Chapter 4), the users managed the on-line conversation very smoothly, both in *kairos*

and in *chronos* times (see Section 2.2) – they took the telephone interaction forward though the system (S) often parsed the user's (U) turns incorrectly. The users, however, oriented to the system as an able 'other' whose turns were given a local meaning that fitted the user's agenda (see Appendix 1-1 for transcription conventions.):

1(1)

- 1 S: **'So you want to leave a message for Eve - is that right?'**
- 2 U: **a message for eve that's correct**
- 3 S: **'What is it that you want to do then?'**
- 4 U: **can you say that judy is at home**

In Extract 1(1), the system fails to parse the user's turn in line 2 as an affirmative, and hence produces a relevant next (line 3) as an answer to a 'no'. However, the user does not entertain the possibility that her verbose 'yes' was taken to be a negative answer and interprets the request for a restart as a continuation question.

The following extract is from two users at a Learning Word 5.0 tutoring program (cf. Chapter 5):

1(2)

- 1 B: **so (.) press y (2) [((inaudible))]**
- 2 A: **[this] this one speaks American too**

A is referring to the computer/program as *this one*, which is ambiguous in human/inhuman respect, but which nevertheless objectifies the 'other' into a non-copresent entity. A less interactive computer program, a word processor, is used with the help of a manual in the following extract (see Chapter 6):

1(3)

- 1 A: **m:itā?/wh:at?**
- 2 B: **"Do not use (the) toolbar to open a new file"=**
- 3 A: **=oh "because you need to use the temp^o.lates .option^o"**
- 4 **((gaze up to screen, moves back, left index finger next to text))**

This short extract comes from a complex interpretive environment in which two users are doing something with the help of a manual. They have just completed a subtask after which the warning they quote aloud (lines 2 and 3) is written. The users first take the quoted piece of text as an instruction rather than a warning, as if the manual was 'repairing' them. Rather than being an example of the users' orientation to the manual as a copresent 'other', however, the extract reveals the organisation of interpretive work in manual use. The manual is oriented to as a 'participant' whose 'turn' instigates a fourth position (self-)repair from the user: the first position is filled by the manual's previous instruction, the second by the user's interpretation of it (i.e.

a practical action at the computer), the third by the user's reading of the next instruction; A's *m:itä?* 'wh:at?' in line 1 fills the fourth position.

The computer is the locus of the 'other' in desktop video conferencing, as well. This time, however, not only text, but also video and sound can be the mediating vehicles. And, with one's own video picture on the screen, oneself becomes another 'other' in this new communicative situation. This can be detected in recordings of video conferencing (Chapter 7) in which one's own picture is commented on:

1(4)

- | | | |
|----|-----------|---|
| 1 | S-All: | hah hah hah ha |
| 2 | | (1) |
| 3 | F-Harri: | hm ((through nose)) (((smiling, turns to Manna)) |
| 4 | | °(nuilla)° on ihan hauskaa / °(Those guys)° have <u>fun</u> |
| 5 | | ((turns to Meerit)) siel[lä] /the[re] |
| 6 | F-Meerit: | [°(kyl-lä)°]/[°(ye-es)°] |
| 7 | F-Harri: | (.) me istut(aan [vaan] täällä,)/ <u>We</u> are[(just) sit(ting) here,] |
| 8 | | [((hands between the thighs, cramped position; |
| 9 | | turns head, nodding, from Meerit to the screen))] |
| 10 | F-All: | [((laughter))] |

Harri's turn in line 7 makes a comment about what the Finnish video picture looks like. His note is prompted by the two video pictures that are next to each other on the computer screen: the Swedish one looks happy and they can hear laughter from Sweden. (In fact, this example could be used to illustrate what feminist theorists call 'Interrupting Othering' (Kitzinger & Wilkinson 1996, 16). One's own video picture makes it possible to see how the 'others' see oneself, and thus "to relativize and problematize 'our' own perspective: it can be uncomfortable, unsettling or painful, but it is an essential beginning if the process of Othering is to be interrupted" (Kitzinger & Wilkinson 1996, 17).)

1.1.4. Language technologies as communication environments

Language technology communication environments provide for interaction between humans which is not face-to-face. It can take place via a telephone, via written language, via a video link, in other words via a mediating piece of technology. In the present world, computers tend to be an integral part either in the production of language (e.g. word processors) or providing a communication channel (e.g. desktop video conferencing). The environments in which technology-mediated communication takes place can be varied, but so far the surroundings, apart from telephone use, have been mostly institutional and especially educational.

The detailed analyses of the users' linguistic and other turns will show how the differing stages of interactivity and differing interactive resources

(text, voice, video picture) result in differing orientations to the 'other' interactant. Mediated communication distorts the availability of the 'other' as a fully fledged communicative partner, but the unavoidable distance to the 'other' (either in space or in time or in both) also provides new resources for the meaning making process. In face-to-face interaction, intersubjectivity or sharedness is a contingent phenomenon; in user-'other' communication, the sense has to be constructed one-sidedly, rather than coconstructed, making the human participant solely responsible for the emerging meaning. Nass et al. (1994) suggest that human-computer interaction is social, and that "concern with the inability to create a photorealistic, full-motion video, or other high-bandwidth representation may be highly overrated" (Nass et al. 1994, 77). In the present study, four language technology environments with differing sophistication levels will be explored.

Because of the multimodal nature of new technology, it is better to talk about 'encounters', rather than 'conversations' with them, even if some language products, such as 'intelligent' telephony systems, might provide an interface that functions on the basis of talk only. Goffman reserves the term 'encounter' or 'focused gathering' for "when persons are in one another's immediate physical presence" (Goffman 1972, 17) with a "visual and cognitive focus of attention" (ibid.). Typically, computers and manuals are used in close contact so that the user-reader is able to see the signs and use the gadgets. And when the computer screen is conveying a video picture from another site, the medium transforms the spatially distant 'others' into a representation on a physically proximate screen.

In pairwork and group situations, two encounters are taking place: one with the human others, one with the language artefact that could be a special 'other'. In the case studies of the present work, the participants had not used the system or a specific feature of it before, which meant that the encounters were not routinised. Therefore, the data also provides access to the crucial 'first encounters' with systems, interactions which most certainly (would) affect the users' willingness to 'meet' the system again (cf. e.g. Svenning & Ruchinskas (1984) which concentrates on organisational teleconferencing from the point of view of user acceptance). Pairwork is an efficient way of eliciting users' interpretations of new technology, as documented in Frohlich et al. (1994, 392).

What do people actually do when they encounter a computer interface, especially if they are not used to working with computers? How can the activities in using a computer be studied? In the traditional human-computer interaction research, certain types of psychological experiments have been carried out, e.g. the reaction time to prompts on the screen have been measured (e.g. Baecker & Buxton, 1987). Artificial intelligence research has the user model approach which is interested in what the user's world model is in the situation, i.e. what her conception of the task at hand is (e.g. Perrault & Allen, 1980). This type of modelling can be used to help guide the user to do

the right thing. This is a special interest of dialogue system designers whose aim is to develop machines that users can talk to in order to help the user achieve some goal or task. Scripts of unsuccessful or 'failing' dialogues are studied to find out how breakdowns can be avoided, where the ideal is usually a dialogue that runs along certain expectations – to every detail. Less interest has been paid to the actual circumstances of the interaction with the computer, for example. Nor is the fact that in human-human interaction, misunderstandings occur and are routinely dealt with. In the present work, various semiotic modes of interaction will be researched, and one of the aims of the dissertation is to explore how the users coconstruct the interactive situation such that the written or spoken language produced by the other human participant(s), the program and the manual is incorporated into the unfolding meaning making together with visual semiotics. Whether the language or other signs interpreted *in situ* are asynchronous or synchronous, the user has to make them work synchronously in the activity s/he is involved in. Thus, in the words of the advocate of integrational semiology, Roy Harris:

Time is more basic than space where communication is concerned. If A is three thousand miles away from B, it may take days or even weeks for A's letter to arrive. But if A and B communicate by telephone, the distance does not matter. In this sense, space can be treated as a function of time for communicational purposes (as metaphors like 'global village' implicitly recognize). This is not to deny the role of such factors as location, shared visual orientation, physical proximity, etc., in articulating particular communication situations. But when all such factors are held constant, time still structures the communication that takes place.

(Harris 1996, 90-91)

Harris is another communication theorist who has interesting points to make about what communication might be really about, but unfortunately does not analyse real interactions. In the present study, detailed analyses of encounters with various kinds of language artefacts will show in what manner the time of the interaction, i.e. the sequential meaning making, and the stableness of written language (not as meaning but as an object that exists without disappearing like speech) on a computer screen or in a manual are interwoven in deciphering what is going on. Similar work on the nature of language in text-based action is done by Smith and Whalen, for example: "The concept of information, so widely used in an entirely unproblematic fashion in current writing on the "information society" etc., proposes a model of some original item of knowledge or data which simply travels, in stable form, from the original caller to its appropriate response" (Smith & Whalen 1995, 10). It should be noted that many forms of mediated communication (e.g. user manuals; computer programs) could be regarded as belonging to the mass media. However, in this dissertation the distinction between media, mass

media, and (new) technologies is not important as such⁵. What matters is how the 'encounters' are analysed as examples of interpretation: "Meaning is in the situated individual not content" (Anderson & Meyer 1988, 192).

The process of computer-mediated communication as an exchange of messages has been investigated before, but usually as divorced from the actual production and interpretation of 'turns'. For instance, Severinson Eklundh (1986) discusses the consequences of computer-mediated interaction to the structure of dialogues. Her data consists of stored computer conference scripts and interviews on the basis of which some interesting observations about the nature of mediated discussions could be done. However, similarly to the investigations of conversations in which only language is looked at (as a rough transcription, for example), the situated accomplishment of written interaction cannot be grasped.

1.1.5. Instruction giving: Language as/for social action

As in the data analysis a fair amount of attention will be paid on how mediated instructions are interpreted in language technology environments, an overview of instructions as a language format is in place.

M. H. Goodwin (1990) gives a good summary of research on directives in face-to-face interaction. She convincingly explains the lack of interest in linguistic research on directives as a form of social action, and in social sciences on directive/response sequences as a site in which social differentiation can be detected. When somebody else is asked or requested to do something, the speech action can in general be called a directive – the format of which in talk can vary because what matters is that the receiver understood an utterance as a directive.

In mediated instruction giving and receiving, the subtlety of face-to-face interpersonal communication is replaced with communication in which the 'sender' is not present. However, it is the nonpresence of the originator of the messages that actually reinforces the division into 'sender' and 'receiver': neither can monitor the other(s) during their 'turns-at-talk', and therefore the production of the mediated instruction becomes recipient designed (Goodwin 1979) for an imagined, anonymous audience. The receiver, though, has to make the instruction work in a specific setting of activity, and this activity is – like talk-in-interaction or any other human sense-making – sequential in nature. M.H. Goodwin criticises the linguistic research on directives as language use for concentrating on isolated utterances that have been abstracted away from real world activities. Her own research (1995) deals

⁵ For example, in a recent collection called *Semiotics of the media: state of the art, projects, and perspectives* (Nöth 1997), the following areas are represented: pictorial and graphic semiotics; film and acting; television, video and radio; computers, electronic networks, hypertext and cyberspace; museum.

with task activities, i.e. places directives as produced in a particular domain of action.

Even less research has been done on the situated interpretation of mediated directives, i.e. instructions that are given to users/readers by a 'textual object' (see Section 2.4) that has been produced before the encounter and which are activated (cf. Smith & Whalen 1995), either by the receiver or the 'sender' during an ongoing activity. Though the communicative situation differs drastically from that between face-to-face interlocutors, there are similarities, too. For instance, it is the user-reader's interpretation and ensuing action that gives an utterance or other sign a status of directive. Hence, the format of what is understood as a direction to act can only be detected in examining true situations of action.

In some educational and experimental settings instructions are an assumed and conventional way of accomplishing action. Also in the present dissertation, the directives given in the various language technology environments were expected and even sought for by the human participants. The technologies used, or some aspects of them, were unfamiliar for the users, and therefore instructions were needed.

The term 'instruction' rather than 'directive' is used as a general term in the present work, because the former also implies explanation of action rather than instigation for immediate action. It turns out in the case studies, that this division is sometimes difficult for the receiver to perceive. An instruction or a directive can also deal with past actions, i.e. repair what the user-reader has just done. The empirical studies will highlight the difficulty for designing effective asynchronous repair messages.

The material manifestation of mediated language use is normally written, though spoken systems are under development and a hypothetical spoken telephone system is also examined in the present dissertation. Special attention will be given to the effect of instructions as spoken or written language, together with the other semiotic fields that become relevant in the situated interpretation.

1.2. Method of inquiry

My concentration on technology-mediated communicative environments means that not only are fields such as linguistics and discourse analysis relevant, but computational linguistics, artificial intelligence and human-computer interaction studies, and even sociology, are also applicable. Yet, when the focus of attention is the situated use of language, the research traditions in the fields of artificial intelligence, computational linguistics, and to some extent human-computer interaction are not very helpful, simply because not much has been done with the interactionist approach (Agre 1988a). Instead language is treated as something out of which representations

of cognitive processes can be teased out. The role of language *per se* becomes thus central, and the other contextual resources are ignored.

In this study, the main aim is to develop an awareness about how various representational techniques (e.g. text, sound, video) used in past and present computer (related) language technologies are understood in interaction, and how this understanding is produced in the intertwining of activities and actions with talk and language. The material surroundings are seen to affect the interpretation, which is a sequential phenomenon in itself, and are available as resources (and troubles). The interpretive work is done situatedly, i.e. it is unfolding in time and space. Interaction requires constant monitoring of the 'other', be that (realised through) a computer system, a manual or a copresent conversational participant. For instance, in Extract 1(3) above, repeated here, A starts with a comment about the manual in front of her (*m:itä?*, 'wh:at', line 1), which is followed by a coorchestrated reading of the problematic directive (lines 2 and 3, *Do not use (the) toolbar to open a new file because you need to use the temp°.lates .option°*).

1(5)

- 1 A: *m:itä?*/wh:at?
- 2 B: "Do not use (the) toolbar to open a new file"=
- 3 A: =oh "because you need to use the temp°.lates .option°"
- 4 ((gaze up to screen, moves back, left index finger next to text))

In line 4, A's gaze goes back to the screen, and she 'monitors' the text at the same time by placing a finger to the point on the page they were reading. Though manuals have been a target of fairly intensive research, much of the research on manuals or other 'necessary' texts has concentrated on the features of how information is presented in the texts or on the linguistic properties that might make the text accessible or effective (e.g. Jordan, 1994, Rogers & Brown 1993, Shubert et al. 1995, Teklinski 1993). In the present work, the point of departure for researching the successfulness of manuals as instruction givers is to examine their use as textual objects (for use), rather than them as (inherently coherent) texts. Engeström and Middleton regard information technologies as "a crucial demonstration of these two aspects of artifacts – the semiotic and the instrumental forming layers of mediation in new and complex ways" (Engeström & Middleton 1996, 4). New language technologies offer arenas for encounters that do not necessarily belong to one or other category but fluctuate between the two, depending on the success of the textual object in breaking the unavoidable 'textual time' (see Section 2.1) for the service of interactional time (and space). Also Engeström and Middleton note that – from the point of view of activity theory informed work place research – divisions "between instrumental and communicative actions are not very useful" (ibid.).

Social semiotics relies heavily on Halliday's formulation of systemic-functional linguistics (e.g. Halliday & Martin 1981), in which the idea of

language as carrying stable meanings is rejected, and the potential that language offers as a resource for human communication is the starting point of analysis. A principle goal of systemic-functional and also social semiotic research is to give an explanation to the production of text as an utterance (in the Bakhtinian sense) in which the social situation in all its complexity has affected which selections of language were used: Halliday has been able to point out some copatterings in lexico-grammatical use, depending on the function; language varies according to use. In social semiotics, the context is approached via language, and the research has started with the grammatical forms which are possible to be selected to incorporate different aspects of the topic (ideational and textual) and the communicative situation (interpersonal). The premise is that the form does matter; the chosen tense, modality, nominalisations and other lexico-grammatical aspects are motivated by the producer's socio-dynamic situation and the estimated environment of use of the product. Detailed analyses of literary texts and the motivations that might have affected the outcome, for instance, have been made on the basis of this model (Halliday 1973, Thibault 1991). However, the consumption, the situated interpretation, of texts has not been studied extensively, though according to Fairclough and Wodak, social semiotics pays more attention to "productive and interpretive practices associated with types of texts as well as texts *per se*" (Fairclough & Wodak 1997, 264). They also point out the importance of visual semiotics, or "the multi-semiotic character of most texts in contemporary society" (ibid.), which advances social semiotics from text-only based critical linguistics.

In literary criticism and reading research, the focus has moved from text/author-based to reader/context-of-reading-based investigations (see e.g. Leppänen 1995). Recent electronic formats of fictional texts normally require the reader's active participation and manipulation of the progress or, even, negotiation of the content or event of the narrative (Aarseth 1997). Literary research cannot explain the success of actions-in-the-world a user of a 'necessary' text (for instance, a manual) is occupied with, but it can help explicate problems the traditional format of these texts exhibits when 'activated' (Smith & Whalen 1995, Watson 1997).

In my research, ideas are borrowed from social semiotics concerning the interplay between the immediate (material) and social. Lemke describes the indexicality of meaning in the following way:

Just as the meanings of words change in their verbal contexts, in the phrases and sentences that contain them, and the meanings of sentences change in the contexts of paragraphs and larger units of textual organisation, so do all of these also change their meanings when they are juxtaposed with a picture or a graph, or when they are said by (or to) one person rather than another, or when they are said in one social situation rather than another.

(Lemke 1995, 8)

However, in the analysis of meaning making practices, conversation analytic (CA) ideas of sequence, turn-taking and repair are more productive than the notion of the meaning potential of language as a system (of choices) by Halliday and his followers. The reason is that the dynamics of production as (socially and situationally informed) selection tend to operate at a sentence-in-discourse level and from the point of view of the producer instead of taking into account the dynamics of the interactional situation in which at least two (present or otherwise represented) participants are involved in the sense making procedure: "For conversation analysis, with its principal focus on interaction, the turn by turn, sequential organization of talk has provided an important resource for the analytic depiction of context and the 'indexical properties of practical action'" (Heath 1997, 186). Thus, not only is it important to realise that words change their meaning according to the material and social context, but that the only way of rigorously finding out the indexicality of meaning is to examine what sense and how is made in actual language and other practices. Unlike conversation analysts, Halliday rejects the importance of naturally occurring language with all its hesitations and reformulations as research data; instead, Halliday claims that "transcribing spoken discourse in this way gives a false account of what it is really like" (Halliday 1994a, 63). Thus, for Halliday, the situational aspect is limited to the 'final draft' of what the speaker is saying, leaving the interactional contingencies of the situation out of his theory.

In contrast, the methods of CA are strictly empirical. The amount of data examined in a very detailed fashion tends to stay small and will not permit the researcher to make any strong claims about the general nature of human-computer encounters with mediated, thus virtual 'others', and possible repairs in that interactional work; instead, elaborated accounts (description of the physical setting, transcription of the speech, movements and eye gaze of the participants) of real interactional situations will show the local management of communication (troubles) using the resources available (e.g. the other copresent participant, texts, icons and sounds).

So, conversation analysis, with its strict empirical methodology suits best the purpose of the present study, in which the unfolding social activities are researched. As CA work requires very detailed transcripts (e.g. Ochs 1979, Jefferson 1983) and very seldom uses statistical methods, CA is clearly a qualitative research approach.

1.2.1. Qualitative research

The basic idea of quantitative research is that — on the basis of a large enough sample of the potential research material — the researcher knows how to label the different variables in the sample, after which the data can be analysed statistically. This means that when interactions are studied, (the participants')

interpretation of each of the coded instances is not important but rather how collections of different instances correlate with each other: "quantitative studies emphasize the measurement and analysis of casual relationships between variables, not processes" (Denzin & Lincoln 1994, 4). The truth is out there, and like any other scientist, the human communication investigator can, by rigorous methods, bring the essence of interaction to the fore. The interpretive mechanism with which the raw data is turned into classifications of different sorts is usually detectable, but the raw data is not. Thus, in research reports, the categories and their quantities are the only data that can normally be turned to if somebody wishes to use the same research material. Quantitative research has an inherent appreciation for recurring items. "But repetition or frequency is a poor guide to the communicative significance or meaning of a particular item" (Ball and Smith 1992, 28). The danger of *a priori* theorising in quantitative research holds true for what Talbott sees to be the danger of 'scientific' definitions: "Abstract thinking is, in the extreme, counting: we count instances, but do not try to say instances of *what*" (Talbott 1995, 309). In linguistically oriented computer-mediated communication studies, the *what* can be for instance lexicon. Yates (1996) reports of a quantitative contrastive (writing, speech) corpus study, in which the use of personal reference (1st, 2nd and 3rd person pronouns) was one of the interests. He can claim differences in the use of these pronouns and give hypotheses of why there are differing uses of the three pronouns. However, the figures do not give any idea of what the occasioned uses of the pronouns were (there are no examples in the article), what preceded and what followed the use, etc. Extract 1(4) above, repeated here, will serve as an example of the importance of sequence in understanding the nature of language use.

1(6)

- | | | |
|----|-----------|--|
| 1 | S-All: | hah hah hah ha |
| 2 | | (1) |
| 3 | F-Harri: | hm ((through nose)) (((smiling, turns to Manna)) |
| 4 | | °(nuilla)° on ihan hauskaa /°(Those guys)° have fun |
| 5 | | ((turns to Meerit)) siel[lä] /the[re] |
| 6 | F-Meerit: | [°(kyl-lä)°]/[°(ye-es)°] |
| 7 | F-Harri: | (.) me istut(aan [vaan] täällä,]/ We are[(just) sit(ting) here,] |
| 8 | | [((hands between the thighs, cramped position; |
| 9 | | turns head, nodding, from Meerit to the screen))] |
| 10 | F-All: | [((laughter))] |

In line 4 Harri uses *nuilla* ('those (guys)') and in line 7 *me* ('we') to refer to the people in the two video conference sites. However, what would be missing from counting the occurrences is that, for instance, *those (guys)* is accompanied with *siellä* (*there*) and *we* with *täällä* (*here*). Also, *we* is 'replayed' in lines 8 to 9 when Harri mimics the way he sees the Finns in the video picture. In statistical analyses, the process of communication is easily missed, and

therefore the otherwise impressive statistics do not tell much about the interaction, not even in the case of written texts. Coming to a conclusion is deductive: the premise is that person pronouns are used to refer to oneself, the other or a third party, and that their usage tells us something about the nature of the encounter. The conclusion can then be drawn from the statistics: how much the pronouns are used; are there significant difference in usage between different textual environments.

Qualitative research is inductive in nature: there are no *a priori* 'truths' about, say, the effect of person pronouns, but it is up to the researcher to find what they might be doing in the data. This is what happens in conversation analysis: the researcher does not wish to label the data with predesigned categories, but approaches the data with an open mind, and detects instances that may support or challenge the results gained by CA so far; or the researcher finds out about new regularities in human interaction. These regularities then tell us about the practical methods of doing things – ethnomethodology and CA are very close in their theoretical premises. In his paper on the reliability and validity of CA, Peräkylä emphasises the reliability of (audio and video) tape recordings as raw data, pointing out that the researcher “needs to pay attention both to the technical quality and inclusiveness of tape recordings and to the interplay of spoken language with other modalities of communication and social action”. Guarantees of the validity of the researchers’ analytic claims “in all conversation analytic research include the analysis of the next speaker’s interpretation of the preceding action, and deviant case analysis” (Peräkylä 1997, 216). The deviant case analysis refers to the fact that there are no exceptions in the data: deviant cases have to be explained by the analysis, not hidden away.

One of the basic assumptions in CA is the importance of sequence: turns-at-talk are understood in sequence, the ongoing one being at the same time an analysis of the previous one and a contribution to the topic. D. Boden points out that “it may be argued that each and every ethnomethodological principle hinges on an insistence of a *member's* (as opposed to analyst's) perspective on action, a perspective or position that is necessarily local and locally practical in that it guides actions that are at once temporal and sequential” (Boden 1990, 189). The member's viewpoint was one of the most important ideas in the early CA: “We have proceeded under the assumption (an assumption borne out by our research) that in so far as the materials we worked with exhibited orderliness, they did so not only to us, indeed not in the first place to us, but for the coparticipants who had produced them” (Schegloff & Sacks 1973, 290). Thus, in conversation analysis, the analysis is primarily *emic* – concentrating on the participants’ analyses of the situation (cf. Silverman 1993, 147). Extract 1(1) above comes from a special interactional situation: unlike the user(s), the organiser of the experiment had available information about how the system (S) analysed the user's (U) turns, and what the system’s turns were designed to mean.

CA is clearly interested in language and therefore communication as social action. What is rigorous in CA is that nothing in the data is irrelevant *a priori*. Heritage formulates the programme of CA as follows:

The basic orientation of conversation analytic studies may be summarised in terms of four fundamental assumptions: (1) interaction is structurally organised; (2) contributions to interaction are both context-shaped and context-renewing; (3) these two properties inhere in the details of interaction so that no order of detail in conversational interaction can be dismissed *a priori* as disorderly, accidental or interactionally irrelevant; and (4) the study of social interaction in its details is best approached through the analysis of naturally occurring data.

(Heritage 1989, 22)

Thus, CA is clearly a qualitative method. Everything in the recordings of data is potentially relevant for the researcher because it might have been relevant for the participants. The importance of video data has become evident especially in the detailed analyses of Charles Goodwin's (e.g. 1979) and Marjorie Harness Goodwin's (e.g. 1980) studies. For instance, they have shown the importance of gaze which can be used as a resource by coconversationists to locate the addressee; even the meaning of the utterance can be dependent on gaze. They also consider other nonvocal actions that might be of importance in the interpretation process. By only using audio material of a face-to-face situation researchers have missed a lot of crucial details for a better understanding of the unfolding meaning making situation. This type of analysis clearly differs from those discourse studies where only the product, the transcribed or audio data is of interest, i.e. it is only the linguistic and maybe prosodic cues that are regarded to be of importance. Therefore, if somebody is repeating a word, that might be considered as hesitation, for example, when it might be the other speaker's inattention or something else in her nonvocal behaviour that caused the repetition. Thus, what is evidently contextual and external, something in the other person's behaviour, can be misinterpreted as internal, e.g. the speaker's not finding the right word.

The detailed analysis of situated action – both verbal and nonverbal – is the cornerstone on which to build further analyses. The linguistic items of conversations and all interaction will be more clearly related to the other activities that are going on, and the relevance of the actual situation, not just the conversational context, is heavily stressed. The highly interactive nature of language in interaction is shown in C. Goodwin's research (1979) on how the actual production of sentences in face-to-face communication depends on the context. Thus recipient design is not something done before one's turn, but is actually an ongoing process in the talk. In the analysis it has to be borne in mind that though researchers can return to the data as many times as they want, this is not possible in the emergent situation. Therefore interpretation work is different (and has different consequences) in the actual conversational context than it is for the analyst.

The transcripts that are used as working representations of the spoken and visual recordings, are very detailed, which means that the illustration of the research material can be used by others, and they can even go back to the original recording if they so wish. Detailed transcripts also emphasise the aim of *no a priori* categorisations: they "overcome the tendency of transcribers to 'tidy up' the 'messy' features of natural conversation" (Silverman 1993, 117). Thus, CA allows reinterpretations of the data, and many CA practitioners go back to the same data set time after time.

Semiotics has been an inherently qualitative methodology, as well. This might be due to the fact that semiotics has been mainly interested in developing a concise model of all the signs that humans interpret in their environment. In addition to the structuralist movement, an increasing number of semioticians turn their interest to how signs are interpreted or communicated. Semiotics has been mainly used by theorists of literature and cultural studies, fields which have not been quantitatively oriented in the first place.

P. Flynn, who regards conversation analysis as part of ethnomethodology, gives an account of why the two fields of (socio)semiotics and ethnomethodology/CA might be coming together:

The ethnomethodological and semiotic movements today show signs of convergences. The semiotics movement, with its relatively high theoretical syncretism, has begun to look to ethnomethodology as a source of theory and empirical research on the pragmatics of contextual sign communication.
(Flynn 1991, 238)

This dissertation aims at contributing to the conjunction of the two fields or their subfields. It is always problematic to combine differing research paradigms, but in this case, the detailed study of semiosis, meaning making, as a process requires sophisticated methods, which conversation analysis/ethnomethodology can provide.

1.3. The case studies

This section gives an outline of four case studies conducted by the author. The settings and the research materials collected are described, with a short summary of the aims or results of the case studies.

Out of the four different cases of interaction in technology-mediated communication environments that will be studied in this dissertation, one links two groups of students in an on-line video conference (TEACHING). The rest (TELEPHONE, TUTORIAL, TASK) are environments in which the users encounter a predesigned computer program or a printed manual (though the participants in the three cases were university secretaries, lecturers, and students, the experimental nature of the tasks highlight their status as a 'user' (or 'user-reader'), rather than anything else). As a common

word processor (Word 2.0) and its manual were used in TASK, it represents an almost everyday communication environment. In TUTORIAL, the Learning Word program of MS-DOS Word 5.0 was used, and though the package came with the program free, the number of people taking a computer aided lesson must have been much smaller than the total number of Word 5.0 users. The video conferences in TEACHING were arranged via a freeware program (CU-SeeMe), but its use was fairly limited at the time (1995). And TELEPHONE was a completely experimental system which was designed for the purpose of the case study only. It might be closest to the work of artificial intelligence, cognitive science and human-computer interaction research, all of which are interested in 'intelligent' user interfaces.

The first study examines the use of a hypothetical telephony answer machine (case study TELEPHONE). The language technology encountered engages the user in 'doing by saying' in 'talk time'. Detailed analysis of three phone calls shows the success (or failure) of the system in guiding (and repairing) the user and the methods of the user in managing the interaction. The Learning Word 5.0 program (case study TUTORIAL) concentrates on instructions and repair work as well, now in a context of a visual, 'frozen time' interaction. It manifests the importance of the material communication environment for the unfolding tutorial and the contingencies in (text) interpretation. The last two case studies on pairwork situation using Word-for-Windows 2.0 (case study TASK) and a set of video conferences (case study TEACHING) extend the scope of research to situations employing a richer array of modalities (manual in TASK, video picture in TUTORIAL) of a semiotically more complex communication environment, and what the consequences are for the shape of instructional interaction.

In the data, the users of new technology are doing something with the help of the system: in TASK and TELEPHONE, the users were trying to accomplish an assignment, i.e. to be a participant in an experimental study. In TUTORIAL, though also a prearranged case, the participants were trying to learn how to use the program Word 5.0. And in TEACHING, the computer system was a mediator through which (via voice, video picture and text), students were in an authentic learning situation. In a way, all of the case studies exhibited learning *how* (Ryle 1975 (1949)): the users had not encountered the 'delayed message/messenger' before, i.e. either the system or the function of the program was new to them. Each type of language technology provided different material and communicative resources for the (reader-)user to manage the task at hand. Misinterpretations and their repairs occurred, and their trajectories and scopes depended on the language technology in question.

1.3.1. The research material

The material for the TELEPHONE study consists of audio recordings and scripts from the communicator's sessions with the computer. These were transcribed to give an account of what went on in the phone calls. The programming for the system was done by myself to test some conversation analytical findings on repair by implementing them in a hypothetical telephony answering machine. The study shows what communicative resources are deployed in encounters with telephone dialogue systems, also giving pointers to the design of human-computer speech systems.

The research material for pair work at a tutorial program (TUTORIAL) is a videotaped session of an encounter with an assumedly self-explicating computer program. The program that the participants were using was Learning Word 5, a tutoring program for the Word 5.0 text processing program. The analysis was done in a detailed fashion in order to discover how the visual-interactive language technology shaped the users' interpretations.

The third case study (TASK) material comes from a project called 'The Principles of Accessibility and Design of English Texts', undertaken in the Department of English, University of Oulu between 1994 and 1996 (see Nyyssönen & Kuure 1995 & 1997). Pairs of students were asked to undertake a task to produce Mailing Labels. As this feature of Word-for-Windows 2.0 program was thought to be unknown to the students and so potentially problematic for them, a manual was placed in between the participants such that they could use it in case of problems. One task session was analysed in detail to see how the fixed text of the manual and the (partly interactively produced) text of the program were incorporated in the ongoing semiosis.

The final case study, TEACHING, concerns recent developments in desktop video communications technology on the Internet. Data was gathered on the use of CU-SeeMe video conferencing from a one month teaching experiment which was undertaken in an English Department of a Finnish University (nicknamed Suomi) in Spring 1995 (see McIlvenny 1995). The Internet was used to give a university course in two places at the same time; one group of students was in Finland and the other in Sweden. The teacher and the students used Internet communication and data sharing to exchange ideas and papers. The video conference data provided interesting material on how typing on the video picture was used as a synchronous way of communicating instead of sound, or how the video picture was an extra communication space when other, voice based, interaction was going on. As the situation was totally natural, it gave a chance to also research how the participants understood the interactive, time and space bound situation in general, and how they related to the two rooms that were present: the one they were in in Finland, and the other one conveyed to them via sound, text

and picture that the three computer screens in the room made available to them.

Thus, the data consists of a set of textual and visual encounters with technologically-mediated 'others': first, a hypothetical answering machine encountered by secretaries (an example in Extract 1(1) above), a Word 5.0 tutoring package encountered by a pair of university lecturers (Extract 1(2)), the Mailing Labels feature of Word-for-Windows 2.0 used by students of English with the help of a manual (Extract 1(3)), and Finnish students of English encountered by Swedish students of English by video conference on the Internet (Extract 1(4)). The focus of my study is in the methods the human participants in each case used to make sense out of their encounter with each textual object: how the sense of participation, identity and intentionality of the 'other' is constructed.

The dependence on language to manage the encounter varies in the four case studies. For example, the first one was self sustained: the telephone conversations were recorded but not videoed, so the study was very much language oriented. In the second experiment, the users were relying on each other and the feedback from the computer/'other' in deciphering the learning (to use) situation, and in the third case not only the computer and the other participant, but also the manual/'other' was a resource in an experimental task. The fourth case study examined the problems revealed in a technology-mediated 'virtual' university course which was conducted between two classes at two sites via video conference, with the teacher at only one site in Sweden.

1.4. General outline of the dissertation

The first part of the dissertation gives an overview of fields and approaches relevant for the study of language-in-use from the point of view of how an 'other' is understood or mediated. Where texts are concerned (2.1), there are subfields of linguistics that deal with language as a tool rather than as an abstract system. Those fields have been given a general name of discourse studies. Another domain also concerned with human action and agency is sociology; attention will be given to the sociology of time and how this relates to the 'other' as a social entity (2.2). Ethnomethodology and conversation analysis (2.2.1), give a good basis for understanding participation and encounters in general. Both are interested in everyday communication through language as action. One important approach to the 'other' can be found in the field of semiotics, in which the human being can also be understood as another sign or sign system (2.3). Semiotics is also concerned with visual phenomena, and thus is an important field for studying computer mediated communication in which perceptible media are prominent. Materiality is concentrated on in this section; visibility and materiality are

closely connected. Social semiotics – combined with conversation analysis – gives grounds for a comprehensive perspective on interpretation in human linguistic and other interpretive actions by stressing the sequential nature of the situated semiosis of the 'other'.

All the hypotheses of the theoretical background of the dissertation are then explored in a series of case studies that were conducted between 1990 and 1996. The following studies, already introduced above, on human communication in different new technology environments were conducted:

- TELEPHONE: a hypothetical telephone answering system used by volunteers in an experiment, Chapter 4;
- TUTORIAL: a pair using a real Learning Word 5.0 package; an experiment, Chapter 5;
- TASK: a real Word-for-Windows 6.0 program (Mailing Labels) used by two students; an experiment, Chapter 6;
- TEACHING: a real desktop video conference program (CU-SeeMe) used in a natural educational setting, Chapter 7.

These case studies then give information essential to the understanding of the problematics of human interactions with (the language of) the 'other'. Each case study environment is different semiotically and materially. TELEPHONE with its 'talk time' nonvisual language technology provides a starting point for exploring what linguistic resources are utilised in encounters with dialogue systems. The chapters dealing with cases TUTORIAL, TASK and TEACHING will introduce theoretical issues and background relevant for researching the semiotically increasingly complex communicative situations. They can therefore also be read independently from the whole dissertation. A comprehensive discussion of the results of the eight first chapters is going to take place in Chapter 8. The concluding Chapter 9 will provide general conclusions of the whole work, stating some basic principles recommendable for the language technologies of Information Society, and giving pointers for further research.

2. ASPECTS OF AND APPROACHES TO ENCOUNTERS WITH 'OTHERS'

The present work will explore how communicative resources are used in various language technology environments. The entity encountered will be called an 'other' to emphasise the communicative aspect and, at the same time, to keep in mind the difference from everyday encounters with other people.

Otherness has been a popular research theme in many humanistic fields, especially in Cultural Studies. The term often refers to alienation from the perspective of somebody feeling different, or to rejection from the perspective of somebody feeling the other is different. In the present dissertation, however, the 'other' is understood primarily and simply as 'somebody (or something) else', and feelings of estrangement are not assumed, even if they can be associated with the 'other': the interest lies in how the human interlocutor constructs the technology-mediated 'other', but negative (or positive) feelings can be only part of the depiction. For instance, in Example 1(1), repeated here, the user refers to the system as *you* (line 4), thus orienting to the system as a dialogic person rather than as a machine.

2(1)

- 1 S: 'So you want to leave a message for Eve - is that right?'
- 2 U: a message for eve that's correct
- 3 S: 'What is it that you want to do then?'
- 4 U: can you say that judy is at home

New technologies have brought many taken-for-granted issues to the fore in linguistics and other human sciences. The difficulty of building conversing machines forced researchers to go back to the premises of interaction and see what is programmable, or what features of interaction can be otherwise supported. In the systems, the written mode has been prevalent, but with video conferencing, visuality in the sense of space as shared or represented, has become another issue to reevaluate. Parkes and Thrift, discussing live broadcasts on television, asked "are the media which take messages and images across the world able to act as mediators in the realization of place or simply as reminders of the here-and-now?" (1978, 129). Nowadays people can also engage in on-line video conferencing in which interactions take place and

the 'viewers' are active participants instead of passive watchers of a programme along with a mass audience of millions of other people. Parkes and Thrift cannot answer the question they pose; the present work addresses this problem in the last case study where the question is posed whether the 'other' is oriented to as sharing the time and space of the other end of the video-link in educational video conferencing.

In the following three sections, three relevant approaches to the theme 'other' within the humanistic field will be illustrated. The conversation analytical approach already introduced above will reenter under the topic social entity, but, because the idea of situated interpretation is considered as essential for the present work, the text/discourse and (social) semiotic approaches will be discussed from this aspect of meaning making, as well. All of the approaches regard language as essential to interaction and therefore to 'accessing' or being aware of the 'other'. Their relevance to the study of technology-mediated communication will also be clarified. The chapter ends with a separate section about my approach which pays attention to the visual, textual, and sequential aspects of encounters in language technology environments.

2.1. The 'other' as text/discourse

Texts are a primary locus of research in the Humanities. Depending on the field, 'text' can refer to different genres of oral or written language, film, theatre, photography, archaeological sites, historical documents – artefacts produced by humans. There is a common denominator to all of these types: they are traces of a past event, even if the recorded performance of a speaker and all the other documents are received in the present. Thus, even in Example 2(1) above, the system's (S) turns appear on a computer screen on the basis of a program that 'reacts' to what the user says in an unplanned (though predesigned) manner – the repair initiations were programmed to appear randomly. Though they are read aloud over the phone to imitate a voice based system, what the system can 'say' during the discourse is limited beforehand.

Discourse is a research interest in various fields; linguistics tends to produce predominantly structuralist accounts of discourse (e.g. work on 'discourse grammar'); in artificial intelligence the work concentrates on the representable plans of discourse agents; in psychology, cognitive science typically deals with schemata, and social psychology studies power and repertoires of discourse. Basically all the areas of linguistics which deal with language meaning, and not just structure, have to take the communicative function of language into account at some level. There seem to be two competing schools of thought within the discourse oriented approaches: 1) those who seek deep or internal discourse structure(s), and 2) those who want

to understand language in use – discourse as a social phenomenon. This division also reflects the view of language as either 1) a general phenomenon of which written or otherwise stored archives exist to study (language as a past event, with the focus on the sender(s)'s message), or 2) language as action by somebody to do various things in the world (language understood as a present event, with the focus on the effects and interaction). The present study belongs to the latter category, though some of the ‘others’ are represented by products rather than processes.

Within linguistics, Austin’s idea of speech as acts (1962) became a very popular account of language as a communicative event. Austin, whose work has been taken forward by Searle (1969) and Grice (1975), wanted to expose the rules which speakers and hearers seem to rely on in speech events, now concerned as acts rather than referring structures. Coming from language philosophy, in which language is traditionally seen as propositional (as being true or not true) speech act theory envisaged, for instance, indirect speech acts, i.e. speech acts that have a ‘literal’ meaning but which then are used to do something else (e.g. *Can you pass the salt?*).

As shown by Streeck’s critique (1980), speech act theory is based on speakers’ intentions (realised as illocutionary acts). In mediated discourse such as instruction giving, there should be no problems if speech act theory were adequate to explain what is going on: the receivers merely unpack the intended meaning. However, the difficulty people have in acting upon mediated instructions shows how little a role the ‘speaker intention’ plays in interaction. Nor is it possible to have a limited set of automatic rules which are applied in using speech acts. Though Winograd and Flores (1986) regard speech acts theory as a useful basis for building cooperative systems, they also emphasise that programs cannot infer speech acts from language because computers are just input-output devices which have no access to the details of the speech act’s production or reception. Thus, Labov and Fanshel’s work on shared knowledge of the topic as consequential to the interaction – A-events are known to one speaker, B-events to another, and AB-events to both (Labov & Fanshel 1977, 62) – is only applicable to understanding the actions of the human user of a computer media, even if the design of the program would assume some shared knowledge.

2.1.1. Another look at the discourse context

Language is one of human kind's most important characteristics. It is certainly the most useful tool people possess. Attempting to study the forms the tool may take, apart from any uses which it has, may lead to a great deal of success initially, for undoubtedly a great deal can and should be said about its forms. But tools exist to be used. And what they are used for is no less important than what they are!"

(Wardhaugh 1976, 211).

As evidenced by the widespread use of speech act theory, Wardhaugh's plea from over 20 years ago for researching language from the point of view of its use has been accepted within mainstream linguistics. The present day debate concerns not so much use of language but how seriously the context of its use is taken into consideration in language research, not as something adding to the meaning or making it more precise but as an intricate part of meaning. This anticipates an interdisciplinary approach in order to grasp "practical relationships between representation and materiality, the still ambiguous duality of meaning and matter of time and space" (Friedland & Boden 1994, 32–33). Conversations have been researched as a special form of language use; as situated, dynamic processes. However, there is no such thing as nonsituational language use: if we write we do it there and then, the finished text is then read by somebody else in a certain time and space. B's turn in Extract 2(2) can serve as an example of the complex relationship Friedland and Boden are referring to.

2(2)

- 1 B: **ooksä aikaisemmin tehny tämmösiä tarroja?** ((soft voice))/
- 2 have you made these sort of stickers before?
- 3 A: **e-n**, ((gaze: instruction sheet))/no-o,

A and B are participating in an experimental task and reading the instructions which are printed on a sheet of paper. *Tämmösiä tarroja* ('these sort of stickers') in line 1 is thus referring to the goal they are about to set out to achieve, something that is not there yet physically but which has been verbalised on the instruction sheet and which should be brought about with the resources available. Though *these sort of stickers* could refer to the sheet of exemplary stickers with which A and B had also been provided, B's bodily posture and gaze, together with the unstressed *this sort* make it clear that she is not talking about the specific stickers in front of them. B is inquiring about an A-event (Labov & Fanshel 1977), making the contrast with the present situation and possible past experience prominent by stressing *aikaisemmin* ('before'). In this example, the materiality also covers the layout of the instruction sheet, where it can be found in the environment, and anything else in the perceivable surroundings that might effect what was said and how it was said or interpreted by the other participant, both of whom are part of each other's visible surroundings.

In written communication, the roles of the 'sender' and 'receiver' of a message are distinct as the acts of producing and reading text are normally separated in time and space. Therefore, the 'speaker' is absent when the 'hearer' gets the message. But, in the same way as the context sensitive phenomenon of talk-in-interaction does not just convey messages from one mind to another through the abstract system of language, written communication is also a situated activity, split into two activity systems of writing and reading. The basic material requirements and differences in

written and spoken communication can be highlighted by using Hockett's (1977) inventory of the design features of human communication as a point of reference.

2.1.2. Design features of written communication

Hockett (1977, 124) compares animal and human communication through a set of so-called design features (DF) which list some essential characteristics of human communication. He concentrates on speech, the linearity of which tends to disintegrate, especially (but not only) compared with hypertext. In the following, some of the design features of speech will be discussed vis-à-vis written language:

- The Vocal-Auditory Channel: speech is produced by vocal chords and received by ears.

This rough reformulation of Hockett's description shows that signed languages were not included in his classification at all. Signed languages are visual, as is writing, but — in addition to the occasioned, sequentially organised use of language in human activities — writing lacks what is involved in any spoken/signed language, namely so-called paralinguistic or nonverbal elements. Of course these can be included in written language, but they lack the intricacy of spoken, face-to-face interaction. For instance, ':-)' is nowadays a widely used 'smiley' in e-mail discussions (Werry 1996): does it convey that a person smiles at the end of the sentence, where it occurs, or does it 'cover' the whole of the sentence? With printed texts, it is only the voice quality that can be conveyed over a stretch of talk (e.g. capitals for loudness).

- Broadcast Transmission and Directional Reception: sound moves in all directions from its source, even through obstacles.

The recipient of a written message has to actively seek to receive the message, i.e. read it; it is not possible to 'overread' a written message as it is to overhear a spoken one. The forcefulness of spoken language is used in modern packaging: to make sure that the consumer is made aware of possible dangers of some products, a microchip warns the client of the potential hazards by vocalising them. Thus, though (hearing) users cannot be forced to read a label, they cannot avoid listening to a warning.

- Specialisation: speaking developed for the communicative function of language only; does not require much of physical stress; a speaker can undertake other tasks simultaneously.

In contrast, writing and reading are much higher developed skills than speaking and hearing. It is typical for writing to edit and reformulate the text; unlike in speech, where the editing process is receivable by the hearer, only the end product is normally perceivable in writing. Both the producer and the recipient of writing are much more restricted than speakers and hearers to simultaneously engage in other activities.

- Rapid Fading: spoken signals are transient and leave the channel open for the next message.

The fading of the message has traditionally been the most important difference compared to writing, as written documents make it possible to store knowledge. But, the durable character of writing can also be an organisational factor of human interaction with a textual object that is advising him or her in a practical activity.

Crystal (1995) also lists differences between speech and writing. Out of his inventory, the consequences of asynchronicity are important for the present study:

Lack of visual contact means that participants cannot rely on context to make their meaning clear; nor is there any immediate feedback. Most writing therefore avoids the use of deictic expressions, which are likely to be ambiguous. Writers must also anticipate the effects of the time-lag between production and reception, and the problems posed by having their language read and interpreted by many recipients in diverse settings.

(Crystal 1995, 291)

Issues of shared knowledge are researched in cognitive linguistics which is interested in the relationship between mind and language. Hutchins in his explorations into situated cognition in a navigation environment points out how writing as a material form of language can cause problems: "Representing the bearings symbolically also introduces new possibilities for error" (Hutchins 1995, 227). Written language (e.g. *Light2*) can – because of the visual layout – be mistaken for something else (e.g. *LightZ*). However, the objects of the signs (physical lights) might be easily separated. Sometimes, then, spoken language (e.g. "*light two*", "*light z*") causes less potential for misunderstanding, because there is no ambiguity in the spoken version unlike in the symbolic representation.

2.1.3. Reading as social action

There are two basic approaches of investigating the 'other' as text or discourse: 1) How does the reader construct the meaning of text and 'other',

and 2) How does the text display the self/other distinction. The latter is especially interesting when the text is telling the reader to do something.

Leppänen (1995) gives an overview of the developments in text analysis, more specifically in literary criticism and reading research. The synopsis of approaches to interpretation and reading (Leppänen 1995, 166) shows how the research in both fields has moved from text/author-based to reader/context-of-reading-based investigations. Along with the change of focus, the analyses of meaning making have moved from textual structure (e.g. semantic, pragmatic) to the reader's capacities (e.g. knowledge, schemata), or to the society and culture as reflected in, or coconstructed/challenged by the (author)text/reader (e.g. power relations, conventions of writing/reading). The view of meanings being made only in the encounter of a text and its reader is maybe at its most extreme in Barthes' claim that the temporality of texts has changed from before (the writer) and after (the text) into "there is no other time than that of the enunciation and every text is eternally written *here and now*" (Barthes 1977, 145). Barthes' claims are reflected in later literary studies in which the role of the reader as an 'understander' has changed to be a 'participant':

We no longer adopt the premise that knowledge of the text is absolute, not at the level of simple perception of print on paper, not at the more abstract level of global meaning, which is often disguised rather than made explicit by the surface structure of sentences. The meaning of a text may also be 'relative to observers' because the reader is participant. And the reason why he is a participant is that texts have their own uncertainty principle – missing information that the reader must supply from his own context of knowledge. And the sorts of questions he asks influence the sorts of answers he is likely to obtain.

(Campbell 1990, 36)

The reader being an active maker of meaning is a common disclaimer nowadays. However, it does not go into what being a participant in an activity means to the meaning making; how do people supply knowledge, and why do they ask the questions they ask? These are the research problems examined by ethnomethodologists and conversation analysts (see 2.2.1). From being seen as unproblematic referring signs, the function of language and linguistic items has turned out to be an obscure process in which the mind of the interpreter is not only a language processor but – even for cognitive scientists – something relying on the environment: "The illusion of meaning in the message is a hard-won and cultural accomplishment" (Hutchins 1995, 239). According to conversation analysis, sense-making, or people's interpretative work (semiosis), takes place in a sequential fashion; indeed, every contribution to an ongoing activity is both context sensitive and context-renewing. In human face-to-face interaction, intersubjectivity, or sharing an understanding of what is happening, is a contingent phenomenon. However, in user-textual 'other' communication, the situated interpretation

has to be constructed asymmetrically rather than coconstructed, which makes the (present) human participant(s) solely responsible for the meaning making.

When the 'other' is primarily represented by language what sort of references are made to this 'other' linguistically? The normal way of deciphering the self/other distinction has been to call the self 'I' and the other 'you'. In human-text interaction, the writer, if thought of as represented by the text, is not accessible as 'you' (it is hard to imagine a situation in which the text is referred to as 'you')⁶. But the reader of a text can be referred to as 'you' in the text — this can be also done implicitly via orders, directions etc. Thus, texts create interactional asymmetry between the reader and the text, providing the possibility of the writer talking to the reader but not vice versa.

In fiction, the 'I' of the text normally does not refer to the writer, nor does 'you' refer to the reader. The reader of fiction enters into 'textual time' (Smith 1990). In so-called utility texts (definition given in e.g. in Pilto & Rapakko 1995), however, the reader is talked to directly, the text is about the *exophoric* (situational), not so much about *endophoric* (textual) relations (Halliday & Hasan, 1976): the reader stays in the 'real time' (cf. Smith and Whalen 1995). Aarseth (1997) deals with the notion of cybertext which is not bound to the material artefact (i.e. a traditional book can have cybertextual features). However, computerised fiction and other cybertexts can allow the user-reader to negotiate what will be happening next, bringing thus the encounter closer to 'the user's time'.

With some applications, the text on a screen can manage to give the appearance of totally understanding the user. Weizenbaum's ELIZA is a perfect example of how little 'knowledge' a program needs to be taken as a real conversational partner: randomly given reactions to the user's input worked very well as a 'psychologist'. Talbott claims that even Weizenbaum's secretary, who knew that ELIZA was just a program, would want to 'confide in her' in private.

Weizenbaum's secretary, quite undistracted by the mechanical contrivances she was dealing with, immersed herself from habit in the meaning of the text addressed to her, and she felt (with good justification) that this text originated in another mind, *one that considered how to respond to just the sorts of comments she was making*. What she was most likely not doing was considering explicitly whether she was speaking with the computer itself, or a programmer, or some other person. She was simply *conversing with words*. Who was behind them didn't matter. The episode may say more about the pervasive and accustomed anonymity of our society than anything else.

(Talbott 1995, 220–221)

Though we do not know the details of what happened between the program and the secretary, and how well Weizenbaum's secretary actually understood

⁶ It might be possible to contact the writer, the 'you', separately from the reading process, but for instance in the case of necessary texts (e.g. manuals) there usually is no singular or even named 'other' as author.

and knew the ELIZA program, Talbott's claim that words can be enough to converse with is interesting as it challenges the general notion of language having a source, a somebody uttering the words. Also, Talbott does not make a straightforward connection between language use and humanness. Maybe Boden and Molotch in saying that "the new technologies radically advance the "impersonality" of modern life" (1994, 257) refer to this aspect of technology-mediated communication. However, for instance Levinson's (1988) classification of participation roles does not entertain the idea of not acknowledging a source for words that the recipient obtains even if the 'sender' is absent.

In the following section, approaches to the 'other' as a social agent, as an unmistakable source of language, are discussed.

2.2. The 'other' as a social entity

Being a social entity usually implies agency: one lives in a society and therefore one is socialised into acting as a member of a culture. The whole of the society works through its agents' interactions with each other. Thus, any encounter can be regarded as two or more social agents meeting. In present day Western societies, claims are on the increase that "production has become increasingly grounded in discursive knowledge" (Lash & Urry 1994, 60), i.e. working life requires intensified communication abilities and not just manual or other solitary skills. The emphasis on language as a tool for social agency has also led to an increase in language oriented research within sociology.

Members of society live and experience in time and space. In contrast to philosophy, in which the essence of time can be challenged, sociology "cannot participate in the arguments over the justification nor the abolition of time in favour of eternity" (Gurvitch 1990, 35). If time is essential for social agency, then depictions of all human activities, including language use, have to take time into account.

The sociology of time can be researched from the individual agent's or the whole society's perspective⁷. The first type tends to coincide with the lived-in flux of psychological time of past, present and future (*kairos*) (which Jacques (1982) relates to the internal time of memory, perception, intention); the latter with the scientific, measurable, 'objective' time of duration (*chronos*). Thus, for instance silences in interaction could be researched with using the scientific method, measuring the lengths of the silences and maybe giving calculations of what type of silences (in duration) could be found; social-time (Lewis & Weigart 1988) method (*kairos* type), might want to examine how silences affect the interaction, are they dispreferred, what work do they do in talk. For instance, in Extract 1(2), repeated here, the silences of less than one second

⁷ Hassard (1990) gives an overview of time research in sociology.

and two seconds in B's turn are not interesting as measures, but because of the work they do in interaction:

2(3)

- 1 B: so (.) press y (2) [((inaudible))]
2 A: [this] this one speaks American too

A and B are colleagues in an experiment, using a tutorial for a word processing program. B is sitting next to A, and gives him directions on the basis of what they both see on the screen. It could be assumed then, that the short pause (.) after *so* in line 1 gives A an opportunity to show his conclusion of the directives. The two second pause has a different status: the speaker is waiting for the physical action to take place. Mere measuring of the lengths of time would not give any idea about these interactional effects or causes.

The times of past, present and future can be personal (self-time) or interactional (see e.g. Lewis & Weigart 1988). Although artefacts as objects are part of physical reality and thus "the difference between present, past and future tends to be obliterated" (Gurvitch 1988, 42), language artefacts in interactional situations are inherently future oriented as the language (for instance of manuals⁸) is there to be resorted to in case (in near future) the words on the pages are needed to get the task done. Human-computer interaction is a curious mixture of a past and present 'other'; past because the text was created before the present encounter, and present because the user and the text are interacting as if sharing the present time.

The increased interactivity of the computer programs might add to the ease of use of the system, but at the same time, interactivity means that the user cannot function in self-time (proceeding according to his or her agenda), but must function in machine induced interactional time (prompted by the computer). Any interaction, be it face-to-face or technology mediated, has the inherent division of present (the situation) and past/future (the immediate past/future of the situation or past/future as a topic). If interactions are tasks within an organisation, as is the case in my data, they also are examples of organisational time, in which self-time is sanctioned. The interactional past and future are very much within the event, and personal agendas are not put forward.

In the next section, interactional approaches to language use are introduced in which time is – at least inherently – acknowledged as a major factor in human meaning making which is seen as primarily sequential in nature.

⁸ In fact all utility texts, informative texts or directions – though produced before the encounter – are future oriented vis-à-vis the moment of their reading.

2.2.1. Ethnomethodology/Conversation Analysis and the importance of repair work

Ethnomethodology and conversation analysis (CA), already touched upon in Chapter 1, have their roots in sociology. They were developed as a unique way to research society or social order as something that is continuously produced by members in everyday settings. The availability of talk as an object of social action through audio recordings resulted in language use becoming a target of study in its own right. The notion that language and other actions mean something according to their sequential position realised in turn-taking has appealed to linguists and other researchers of language use (such as myself) who are increasingly adopting a CA methodology as a reliable way of studying meaning-making. At the same time as conversation analysts have shown sociology the importance of researching situated language use to understand social order, they have given researchers of language meaning and interpretation a sound method to investigate contextual understandings. Originating in sociology, ethnomethodology and conversation analysis treat language users as producing social action. Thus, language is a social rather than mental phenomenon, making encountering 'others' is the primary locus of study.

Ethnomethodology and conversation analysis give a powerful tool for researchers interested in conversational or intersubjective understandings:

We can start with the general observation that understanding is a practical achievement in conversation and that there may be junctures in and items of the conversation where the business at hand is manifestly and specifically to exhibit participants' understanding. A second observation is that repeat utterances are equivocal as demonstrations of understanding, and that unequivocal displays of understanding can be achieved by producing a transformation or paraphrase of some prior utterance.

(Heritage & Watson 1979, 129, quoting H. Sacks' lecture 9)

People usually have to assume — for the sake of smooth communication — that the conversational partners do understand what the other is saying. However, the participants have certain ways of checking their understanding, and repetition or reformulation can serve this purpose. There are more straightforward meaning adjusting devices in use, repair work being one very much researched within this tradition (Emanuel Schegloff's work). Thus, intersubjectivity, sequential understanding, and repair work are interconnected, as explained by Schegloff:

Organizational features of ordinary conversation and other talk-in-interaction provide for the routine display of participants' understandings of one another's conduct and of the field of action, thereby building in a routine grounding for intersubjectivity. This same organization provides interactants the resources for recognizing breakdowns of intersubjectivity and for repairing them.

(Schegloff 1992, 1295)

Through repair work the meaning negotiation becomes visible. Repair work has been divided into the following types, in the order of preference (on the basis of interactional work⁹): 1) self-initiated self-repair, 2) other-initiated self-repair (different conventions), and 3) other-repair (Schegloff et al. 1977). Uncertainty could be regarded as self-initiated (potential) other-repair (Raudaskoski, 1992). There are many ways of doing self-initiated other repair. Schegloff et al. (1977, 364) give one example:

- B: He had dis uh Mistuh W- whatever k- I can't think of his first name, Watts
on, the on thet wrote // that piece,
A: Dan Watts.

Though they do not elaborate on self-initiated other-repair, the modulation of other-correction seems to use the same devices: 1) "downgraded on a 'confidence/uncertainty' scale, e.g. by the affiliation to the correction of uncertainty markers, or by use of various types of question format", 2) jokes (ibid., 378). The example above would fit in the first category, but the repair is self- rather than other-initiated. In addition to questions, hesitant voice quality demonstrates uncertainty and invites repair from the other.

Self-initiated and other-initiated self-repair can be defined as past-action-reversed-at-this-moment; self-initiated other-repair is future oriented in that it gives the 'other' a possibility to correct the speaker, but it is also past oriented because a turn-at-talk is always an analysis of what happened before. For instance, in the extract from the case study TASK, repeated below, the participants are preparing mailing labels and using the manual to get directions for how to do it. After completing a certain phase, A reads the manual and, in line 1, shows bewilderment at what she is reading.

2(4)

- 1 A: m:itä?/wh:at?
2 B: "Do not use (the) toolbar to open a new file"=
3 A: =oh "because you need to use the temp^o.lates .option^o"

⁹ In repair, self-initiation is preferred to other-initiation: the space for both types of initiation are within three turns (the present, at the transition relevance place and in the next turn) (Schegloff et al. 1977). In other-initiation, the trouble source can be made explicit, and there usually is a short pause, indicating a possibility for self-initiation, and thus making it visible that self-initiation (and correction) are preferred. Anything in the conversation can be a repairable.

A's '*m:itā?*' ('*wh:at?*') is a surprise token with a questioning format which can be interpreted as an invitation to clarify the trouble source for A (B's turn shows that this was her interpretation). B's turn in line 2 functions as a repeat for A (she is reading aloud what A had just read in silence), but at the same time she is getting to know what the problematic line is. In line 3, A's *oh* functions as a change-of-state token about her knowledge (Heritage 1984a); she uses the rest of the written sentence (*because you need to use the templates option*) to demonstrate what she just understood, i.e. the repair sequence is over. In this extract, A's *m:itā?* ('*wh:at?*') shows her orientation to the pairwork task as comprising of sequential actions which are done according to the manual's directions. B uses the manual directly by reading aloud the line (repeating what was 'said' to A by the manual), and A's next turn (line 3) shows that B's turn in line 2 was a successful remedy of the difficulty.

Reformulations and repair are both examples of meaning checking and negotiation. This is where conversation analysis differs from many linguistically originated discourse analytical traditions in which meaning is seen as belonging to language rather than to its speakers. Conversation analysis "examines co-construction through a sociologically and linguistically tuned microscope to reveal realms of interactional work that take place even in fractions of a second, involving the coordination of talk, sound, gaze, bodies, and built environments" (Jacoby & Ochs, 1995, p. 174). As can be seen from this quote, not only speakers as producers of linguistic items, but speakers as bodies in material space are seen as participating in the meaning making. Also people and their products are (potential) signs. CA has been interested in the cooperativeness of human action; in the case of technology-mediated communication, however, the human participant alone has to do the work of 'other'-construction, because of the limits of interpretation of the machine (unless the 'machine' really is a person; e.g. Wooffitt et al. (1997) in which a person's voice was synthesised to give an impression of a machine).

The context of face-to-face interaction is much richer than that of a mediated encounter:

Under any media condition, words derive their meanings only from contexts; copresence delivers far more context than any other form of human exchange.

and

The meaning of any detail—including a word—derives from actors' work in using each particular to inform or "index" every other.

(Boden & Molotch 1994, 259)

In new technology environments, copresence is always virtual — the 'other' is not a copresent person, but a representation of some sort. According to Coulter, "any text whatsoever, if it is in the first instance properly to be regarded *as a text*, must, as a condition of its sheerly minimal intelligibility,

contain its own possibilities of contextualisation, set its own limits upon what a relevant context could possibly be for that text, and thus establish a priori the kind of contextual particulars that, for the purposes of the contextualising practice, could illuminate its problematic component(s)" (Coulter 1994, 690). This is what is usually the case with preprogrammed texts on the computer, as with any text: they are designed for their future uses. Coulter wants to keep a distinction between understanding and interpreting a text. According to him, understanding means that the reader is able to make the text intelligible, but interpreting is an activity in which the text is given a significance. If the difference is brought to the investigation of computerised textual environments and especially the analysis of the data of this thesis, 'understanding' could equal the ability of (native and nonnative) speakers of English to decipher the English language, and 'interpreting' the actions that they do on the basis of what they are reading, i.e. the practical outcome of their understanding. Also, a parallel could be drawn between text linguistic research as concentrating on the text itself and its structure as the target of analysis ('understanding'), and interaction oriented research as treating meaning as situated and emerging ('interpreting'), which cannot be determined *a priori*. 'Understanding' (meaning) seems to concentrate on the level of the sign, how the information, the actor's intent, is organised, whereas 'interpretation' is the same as Schifffrin's (see 1.1.2) idea about interpretation as the receiver's intent.

Schegloff's recommendations for examining any discourse give a good summary of the CA approach:

- (1) The discourse should be treated as an achievement; that involves treating the discourse as something 'produced' over time, incrementally accomplished, rather than born naturally whole out the speaker's forehead, the delivery of a cognitive plan.
- (2) The accomplishment or achievement is an interactional one.
- (- - -) (3) The character of this interactional accomplishment is at least in part shaped by the sociosequential organization of participation in conversation, for example by its turn-taking organization, which is not organized to be indifferent to the size of the turns parties take, but whose underlying (though supercessable) organization is designed to minimize turn size.
- (- - -) (4) Because the actual outcome will have been achieved by the parties in real time and as, at each point, a contingent accomplishment, the mechanisms of the achievement and its effort are displayed, or are analyzably hidden in or absent from, various bits of behavior composing and accompanying that discourse, and analyzable with it.

(Schegloff 1981, 73)

The spatiotemporal meaning construction is — in the case of human-computer dialogue — one-sided, i.e. the human user is responsible for constructing the meaning on the basis of the designer(s)'s represented interpretations ('turns'). In the case of synchronous computer-mediated communication, the amount of representation diminishes; it is only the sign that gets mediated. However, even this affects the interaction and

interpretation. New media environments usually contain texts, in both electronic and paper form. In my work, text use is seen as part of the ongoing meaning making; in this respect, the study differs from traditional ethnomethodology/conversation analysis: "While utterances in the course of a conversation are treated in ethnomethodology as inseparable from the ordering the analyst finds in the sequencing of utterances, texts are not. They should be." (Smith and Whalen 1995, 31.) In my work, they are.

Situated activity systems

One important aspect of the context of text use is that meanings are sought for and interpreted in practical activities and with different participant organisations. Charles and Marjorie Goodwin have employed the notion of '(situated) activity system' (e.g. Goodwin & Goodwin 1987) or 'multi-party interactive activities' (Goodwin 1996) to grasp the complexity of collaborative action in which semiotic fields are an important resource.¹⁰ Their research shows how intersubjectivity is a practical problem for people who are doing something together, a problem which has to be resolved *in situ*. When participants try to keep up with what is going on, what their interactional role is in the interaction process, and what can be done next in the activity, Goodwin (1996) can see at least three important interactional resources that become important: 1) sequential organisation, 2) sentential grammar, and 3) participation frameworks. The first one has been dealt with already above: a next turn is produced and can be understood on the basis of the current one. The sentential grammar of language and especially its indexical properties is a resource that both arises from and builds the context. Different participation frameworks make possible different constellations of participant categories (e.g. speaker and hearer). A corollary to the idea of meaning being related to the interactional particularities can be found in Bakhtin's philosophy, too: "the meaning of whatever is observed is shaped by the place from which it is perceived" (Holquist 1990, 21). If the sense making involves acting according to the meaning, not just observing, Bakhtin's point becomes even more relevant; understanding the meaning of a computer manual, for instance, differs if the activity involves making the computer do something or if the manual is bed time reading. In the first case, the manual text has to be activated in a sequential fashion, for instance as 'directions', whereas in the case of the situated activity system of 'pure' reading, the sequence of interpretation is text-internal.

¹⁰ The term 'situated activity system' was first introduced by Goffman to describe "face-to-face interaction with others for the performance of a single joint activity, a somewhat closed, self-compensating, self-terminating circuit of interdependent actions" (1972, 84-85). Goffman's special interest was in social roles, and he saw the usefulness of situated activity system to understand why and how people mix social roles with the expected institutional ones (e.g. joking in an operation theatre).

Language as action can mean two things: using language as a fairly unproblematic (predefined) tool to do things in the world (e.g. speech act theory), or language as constituting the encounter, which means that language is regarded as a flexible way of acting in the social world, and no one-to-one correspondence can map to strings of words onto, say, intentions. The goal of conversation analysis is "the study of the common-sense reasoning skills and abilities through which the ordinary members of a culture produce and recognise intelligible courses of action" (Heritage 1989, 21). It is actions that are pursued, rather than isolated linguistic utterances. The notion of situated activity is in opposition to the (ideally) limited set of speech acts or moves of discourse analytic research. It is important to realise, however, that the interest in naturally occurring action does not make the CA approach behaviouristic:

One of the advantages of focussing on the term *action* rather than the term *behaviour* is that a clear distinction is implied between the conventional laboratory-based, psychological study of responses and motor movements (what has usually been termed behaviour) and the study of *situated sequences of human activity* (what I take to be the starting point for the study of *social action*).
(Canter 1985, 171)

Although action is emphasised, it should be kept in mind that when language is involved in the evolving activity, its grammatical and semantic properties are fully utilised in the interaction as resources which the participants can use to build joint activities and which are given meaning and give meaning on the basis of the sequential and material context. In that sense, the traditional linguistic concepts should not necessarily be thrown away – they need to be approached from a new, interactional perspective¹¹.

‘Participation framework’ is another concept that has roots in Goffman’s writings. His idea that speaker and hearer roles can be categorised more finely¹² has been developed by Levinson (1988) into a more comprehensive taxonomy of production and reception roles. Although both Goffman and Levinson admit the importance of situated role taking, they concentrate on segmenting the constitutive components of different types of encounters. However, such inventories give a stable picture of communication and cannot grasp the flux, the changing focus of situated interaction: "Deictic reference organizes the field of interaction into a foreground upon a background, as figure and ground organize the visual field" (Hanks 1992, 61).

Charles and Marjorie Goodwin (1992) approach activities from the point of view of the participants, and show how the concept of ‘participation

¹¹ In fact, the new ‘interaction and grammar’ approach within linguistics is heavily based on CA research (Ochs et al. 1996).

¹² Goffman introduced the idea of footing in which participation roles are ‘fine tuned’: a speaker can be subdivided into the possible roles of author, animator and principal, and a hearer into those of addressee and receiver (Goffman 1979).

framework' is useful for analysing a single participant's actions. Not only are there different participants in an activity, but those activities are constructed through the practices they are active participants in:

While using activities as interpretive resources participants are simultaneously faced with the task of building these very same activities. This process is accomplished through a complex deployment of inference, action and behavior which is situated within time and space.

(Goodwin & Goodwin 1992, 97)

The emergent and changing participation roles of social agents interacting in space and time involve language and other actions in copresence. Thus intersubjectivity, building the meaning into interaction with an 'other', is also constituted "as a visible, ongoing activity" (Goodwin & Goodwin 1992, 84). Suchman (1996) shows how in a semiotically complex work interaction, multiple interacting participation frameworks contribute to successful communication.

As participation is "a temporally unfolding, interactively sustained embodied course of activity" (Goodwin 1996, 375), speaker or hearer (or any finer classification) roles are not effortlessly filled, but have to be cooperatively achieved and maintained. The result of this tight embeddedness in activity building is that in interaction any use of language has a function; for instance, language use can constitute attention or inattention (cf. Goodwin 1986b), and visible participation changes according to the intensity of mutual orientation (Goodwin 1996).¹³

Perhaps one of the most difficult aspects of situated activity systems to grasp is the consequential nature of the material and visible semiotics of the meaning making. The environment with its artefacts and the participants' bodies yields important semiotic resources to build the ongoing interactional context:

Whether articulated through talk or body movement, the context is produced in and through a social organisation; an organisation which is both context sensitive and context-renewing. In and through this organisation, and their visual and vocal actions and activities, participants systematically shape and

¹³ The concept of attention is relevant in relation to the present work as well, because the participants of a technologically-mediated encounter are, however interactive the system may seem on the surface, the only ones able to comprehend what is going on. Therefore, the human participant sometimes could be compared to an audience rather than a conversational partner. Thus their actions can also be analysed as attention and inattention to the text-as-'other'. Inattentiveness could then be explained by the intertwining of action, time and space in practices: the computer as a concrete medium (its shape, the screen, keyboard and mouse) draws the encounter towards one with a machine, but language and interactivity (though fairly inflexible) towards an encounter with an 'other'.

preserve the context at hand, repairing the indexicality of practical action 'in flight'.

(Heath 1992, 122)

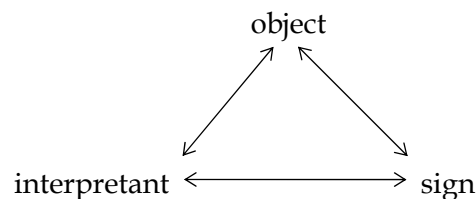
As mentioned above, the intensive cooperation in meaning construction breaks down the participant dichotomy speaker/hearer: "Whether this research leads, in the long term to descriptions of action structure or participant categories is yet a mute point" (Heath 1992, 124). In technologically-mediated encounters, the delicacy of face-to-face interaction as copresent communication is often lost; the indexicality of practical action and repair work tends to be different even in on-line video conferencing. Although it might be tempting to use, for instance, Levinson's (1988) categories to help understand how language technology environments might change the traditional speaker and hearer roles, the conversation analytic way of researching empirical data to find out what categories, why, and when the participants seem to invoke endogenous to the interaction, is though challenging more pertinent. Also, the data can lead the analyst to discover things her hypothesis-bound gaze might never have detected.

The following section turns from how 'other' can be understood as a social entity to how 'other' can be perceived as a sign: (social) semiotics is discussed. The importance of semiotics for this dissertation is that it gives another approach that aims at explaining meaning as a property that exceeds the limits of language and incorporates other humans and human made signs within its realm.

2.3. The 'other' as sign

Signs as mediators of meanings are researched in the field of semiotics. Fiske (1990) introduces three of the most well known ways of understanding how the relation between the sign, its referent and the sign user can be approached: the Peircean triad, Odgen and Richard's elements of meaning, and Saussure's definition of a sign.

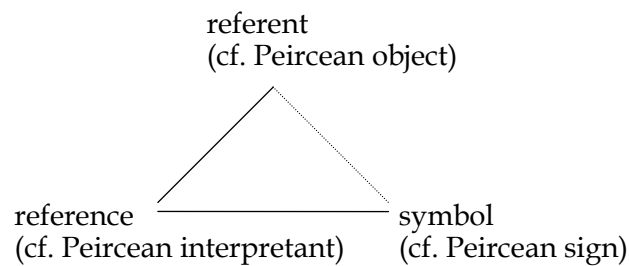
The Peircean triad:



The turn of the century American semiologist Peirce sees the sign as a (never ending) loop, in which the three parts of a sign are interrelated. "A *sign* refers to something other than itself — the *object*, and is understood by somebody: that is, it has an effect in the mind of the user — the *interpretant*" (Fiske 1990,

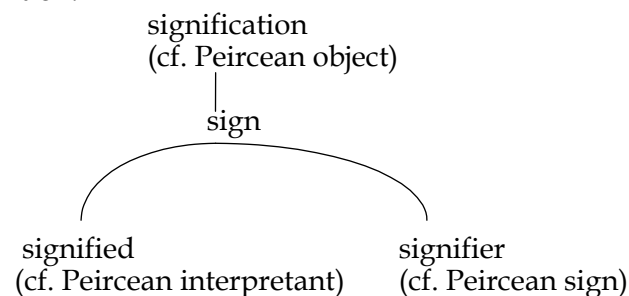
42). When Harri in Example 1(4/6) above says *those guys have fun there*, his utterance is a sign which is making a statement about students at the Swedish end of the video link. Meerit reacts to the utterance with *ye-es*, thus making explicit that she agrees with Harri, or, more precisely, she agrees with what her interpretation of Harri's utterance was¹⁴. The interpretant then can function as another sign (with an object and an interpretant), hinting at learning, development of thinking, and minds with different experiences and thus differing biases and hence differing interpretants.

Odgen and Richard's elements of meaning are:



Odgen and Richard's model resembles that of Peirce's but "the connection between symbol and referent is indirect or imputed" (Fiske 1990, 43). In other words, there is no direct relation between a sign and its referent, but only the mind of the perceiver of a sign can make that link.

Saussure's definition:



Saussure was much less concerned with signs' relations to the world (the object). He concentrated on the signifier which "is the sign's image as we perceive it – the marks on the paper or the sounds in the air; the signified is the mental concept to which it refers." (Fiske 1990, 44).

The closeness of the terms *signified*, *interpretant* and *reference* on the one hand; *signifier*, *sign* and *symbol* on the other hand; and further *signification*, *object* and *referent* means that they are sometimes used interchangeably.

Peirce made the now commonplace distinction between different types of relations between the sign and its object: index, icon and symbol. An index is a "sign with a direct existential connection with its object" (Fiske 1990, 47). A trace on the ground is an indexical sign of an animal, for instance. An icon resembles the sign, but indirectly: "a photograph of my aunt is an icon; a map

¹⁴ "Undoubtedly, many subsequent behavioral responses, verbal answers, images interpreting a caption, and *vice versa* are interpretants" (Eco 1979, 190).

is an icon" (Fiske 1990, 47). And a symbol "is a sign whose connection with its object is a matter of convention, agreement, or rule". Language therefore is symbolic. But all signs are polyfunctional: for instance, 'this' is a symbol, but can be used with a pointing finger which makes it indexical; the meaning of the sign cannot be separated from the object. As a sign can be a longer stretch of language or conversation, language use has been also explained from the point of view of this three-part division. For instance, text is symbolic in that it consists of language, but it is "predominantly indexical when its primary function is conative (appellative)" (Nöth 1995, 46). According to this opinion, commands, instructions and questions would be indexical. Also dramatic texts and realistic literature are mainly indexical (Nöth 1995, 47). Nöth suggests that textual iconicity can be displayed in an image, a diagram, or a metaphor (ibid.). Typography can be used to imitate what a poem is about (cf. e.e. cummings' poetry), and the natural order of narrative is iconic, because "the sequence of the textual signs is a linear icon of the sequence of events depicted in the text" (ibid., 47). Metaphors are iconic as they are figures of speech that cannot be taken 'literally'.

Semiotics is a vast field with differing interests and approaches. One division can be found by regarding the signs (or signifier-signified relations) as material, either in the sense of them being concrete products, implying that signs and language systems are not free from societal and political pressures, or signs (or the signifier-signified relation) as arbitrary, as an independent logical system (e.g. Silverman & Torode 1980 discussion). The sign-as-arbitrary approach is predominantly structuralist which "is decentred in that it sees the codes, roles, perspectives, and structure of language as the ultimate reality" (Manning 1987, 31).

Materiality of signs has been also discussed by Petrilli (1990). According to her, two types of materiality can be detected in signs, viz. extrasign materiality (physical materiality and instrumental materiality) and semiotic materiality (ideological materiality, extraintentional materiality, signifying otherness materiality and elaboration materiality) (Petrilli 1990, 392). Thus, her two types of materiality could be seen as further elaboration on the material (vs. arbitrary) approach to signs and signification.

Extrasign physical materiality refers to the matter the sign is made out of, while instrumental materiality "designates the kind of materiality thanks to which a sign carries out a nonsign function in addition to its sign function" (1990, 372). To Petrilli, only nonverbal signs can have extrasign instrumental materiality, as verbal signs exist for meaning only whereas, for instance, a fur coat can 'mean' wealth, but also has a practical, instrumental function. In technology-mediated communication environments, extrasign physical materiality refers to the substance, colour, shape and other material features of the artefacts, and instrumental materiality to the prestige that some users might feel being surrounded by high tech.

Semiotic materiality proposes that no sign, be it a verbal or nonverbal one, has only one, stable meaning but that the complexities of the communicative situation influence semiosis. Ideological materiality refers to how "the relationship between language, thought, and reality, for both the individual and the collectivity, is mediated and formed within the components of a particular socio-cultural and economic system" (Petrilli 1990, 382) (cf. Volosinov's (1973) argument on the expression organising the experience). Extraintentional materiality refers to "those cases where signs impose themselves upon us independently of our own volition or control: that is, they exist outside the sphere of conscious awareness" (ibid., 391). Signifying otherness materiality refers to how a sign can have various interpretants, different meanings, instead of being semiotically consistent, and elaboration materiality refers to the complexity of signs when they are a product of conscious effort, and also to the heteroglossia of words and language by being 'inherited' from other(s)' uses (cf. Bakhtin 1986).

The relation between signs and meaning has been the main interest of semioticians. This has meant an emphasis on how the object, sign and interpretant are interrelated in the mind of the receiver of the sign. Communication as a sign exchange mechanism, as interaction between people or sign systems, has not been studied very much from semiotic point of view. For instance Wiley, discussing Peirce's ideas, states that "the sign-object-interpretant, semiotic definition of thought, which is quite different from the pragmatic maxim approach, is not particularly open to the conversational model. For one thing, neither sign, object, nor interpretant are, for Peirce, persons. The communicating persons — speaker and listener, utterer and interpreter — are tacked onto the two ends of the semiotic triad, making it a pentad. The internal triads themselves are structures of meaning, not agents or persons. The sign does not talk to the interpretant, nor is it at all clear how the interpretant could talk back or respond to the sign" (Wiley 1994, 24).

However, when Ponzio (1990) talks about man (or self) as a sign, we are made aware that not only can human cognition be considered as a multilayered sign but that we and other people can be treated as material, even extrasign material, as 'others' from our interlocutors' perspectives. For Wiley, the self combines the "I-present-sign, you-future-interpretant, and me-past-object" (1990, 215). Wiley's model is dialogic (there are three types of selves), it is temporal, and semiotic. But, rather than treating interactions between humans as realising pentads (the three dimensional sign, time and semiosis), their practices could be regarded as those of selves with a thinking capacity, but selves that are for the most part of their lives also 'others'. It is very hard to keep these two roles apart in a communicative situation. Thus, the 'other' could, in Wiley's categorisation, fulfil the three roles: you-present-sign, you-future-interpretant, and you-past-object. The dialogic would emerge in interaction: the turn-at-talk (or action) being addressed to the present 'other', and that turn-at-talk being formulated on the basis of the past

interaction with the 'other' (recipient design), with the future 'other' in mind (e.g. uncertainty).

In technology-mediated communication environments, the user, the interpreter, meets artefacts that clearly differ in their extrasign physical materiality from human beings, as they are complex, usually immobile, machines. However, especially with language forms appearing on the machine, their semiotic materiality is more familiar: they use a familiar form of communication, viz. either synchronous or asynchronous writing and reading, sometimes even sound/video; with 'interactive' programs, the borderline between a computer-mediated synchronous 'other' and a computer program becomes weaker. In most linguistic research on computer-mediated-communication, the extrasign physical materiality is not discussed at all, and even semiotic materiality is bypassed to study the interaction as a 'pure' exchange of meanings (cf. Herring 1996). The growing amount of research on the complexity of the technology-mediated communicative environment (e.g. Button 1993, Frohlich et al. 1994, C. Goodwin 1994, 1995b, 1996, M. H. Goodwin 1995, Goodwin & Goodwin 1996, Heath et al. 1995, Hopper, R. 1990, Luff et al. 1990, Suchman 1987) gives a better framework to how the 'other' in these situations is depicted. Petrilli's approach to materiality helps realise the import of the linguistic, visual, and auditory signs that the environments consist of. Semiotic treatments of new technology environments often tend to regard language/texts as semiotic tools and computer/software as physical tools, thus ruling out the semiotic import of the material surroundings (see e.g. Freire, 1995).

Even if the language of the 'other' in these situations were to be researched on its own, there are approaches within semiotics that offer interesting challenges to how the 'other' might be interpreted. Barthes' famous argument about the death of the author challenges any 'other' as delimiting the meaning of the text. That is why the text is an independent entity, and therefore there is no person behind it: "Linguistically, the author is never more than the instance writing, just as *I* is nothing other than the instance saying *I*: language knows a 'subject', not a 'person', and this subject, empty outside of the very enunciation that defines it, suffices to make language 'hold together', suffices, that is to say, to exhaust it" (Barthes 1977, 145). Thus, Barthes uses linguistic analysis to establish his idea of the reader who has to produce the speaker and in general the meaning of the text; he or she has to take responsibility from the author. If the relationship between the author and the reader is disconnected this way, we can ask: how about 'you'? Is 'you' in text referring to the reader or is the relationship split in the same way as that between 'I' and the author, i.e. is 'you' the empty subject or object of the text or can it be regarded as being enunciated by the reader? Certainly there are differences according to genre: the 'I' and 'you' of a novel have different semiotic materiality from those of instructions. In the latter case, Barthes'

claim of the independent status of text is very interesting: the self of the text (who is the 'other' of the reader) is something else than the writer of it.

The computer with its icons and symbols has attracted semiotic analyses. Andersen (1990) gives an account of the typology of interface signs. He classifies the signs according to "the roles they can play in the interface" (Andersen 1990, 198). Andersen does not regard the division into index, icon and symbol as "sensitive to the characteristics of computer-based signs, namely that they can be handled and interacted with" (ibid., 199). His classification stems from the potential of the signs as an interactive device to do something in the system, not from the users' understanding what the sign stands for. Thus, Andersen wants to maintain the division, rejected by Engeström and Middleton "between instrumental and communicative actions" (Engeström & Middleton 1996, 4).

Visuality and interpretation of visual phenomena have been studied by the visually oriented research traditions within semiotics (semiotics of the media, architecture and visual semiotics). However, because semiotics is traditionally a structural enterprise, the visual phenomena are often analysed in order to search for a structure, in the same way as the structure of language has been of interest to linguists. "The third dimension: from reader to user" continues to be at a stage of "a first exploration" (Kress & van Leeuwen 1996, 242). The structural depictions concern the parts of meaning which are frozen in time – semiosis is not considered as a process. In contrast, the constructivist approach to meaning making includes time because if "the connections between the signs and the interpretants are not given but have to be made" (Medway 1996, 478), then this connection making takes place in time and space. In making meanings, people use the context in its spatiotemporal totality as a resource. Silverstein (1992) regards semiotics as the most comprehensive approach to contextualisation: linguistics as a primarily mentalistic field of study misses the cooperative accomplishment of semiosis¹⁵.

Above, the principle ideas of semiosis, or meaning making were introduced. The discussion concentrated on signifying as a process rather than as a structure, as an interactional rather than mental event in which the immediate context of language use and interpretation is crucial. In the following subsection, the socially oriented semiotics is discussed, and coupled with conversation analysis to bring in the sequential aspect of meaning making.

¹⁵ Possibly because some work in ethnomethodology is on cognition as social, Silverstein surprisingly places ethnomethodology in the cognitivist paradigm.

2.3.1. Social semiotics from an interactional perspective

The social semiotic conceptual framework is concerned with the systems of meaning making resources, their patterns of use in texts and social occasions of discourse, and the social practices of the social formations in and through which these textual meanings are made, remade, imposed, contested, and changed from one textual production or social occasion of discourse to another. The focus is on the material and dialectical interrelations of copatterned textual meaning relations and their uses in specific domains of social practice.

(Thibault 1991, 6).

Thibault's definition of social semiotics stresses the active role of the reader/receiver of a message in giving it meaning: texts are really produced in enunciation (cf. Interpretation as the receiver's intent in 1.1.2). For instance, the extract from a telephone dialogue given in 2(1) above, the meaning making resource is a program which is implemented such that there is a communicator reading aloud the program's rather random 'turns'; the social practice is an experiment in which three secretaries are phoning up an experimental telephone answering system, each giving their own meaning to the telephone 'text'. Petrilli's ideas about extrasign and semiotic materiality can easily be seen to support Thibault's argument: they cover the situated meaning making practices that also are of primary interest in the present work. But however much the material approach to semiosis stresses meaning making as a process, the actual analyses tend not to make detailed depictions of actual language use. In this section, Kress (1993) is taken as a starting point to compare social semiotics to critical linguistics, or to view critical discourse analysis from an interactional perspective.

Halliday's (1984) formulations of systemic-functional linguistics together with Hodge and Kress' (1988) social semiotics share the assumption that signs are not arbitrary, contrary to the 'sign fetishism' "which envisages the sign only in relation to other signs and leaves aside the process of communication" (Petrilli 1990, 367). In social semiotics, signs are seen to be motivated and transparent/opaque according to the situation of sign production and reading.

Halliday's systemic-functional linguistics does not make a difference between production and consumption of language, though the emphasis is clearly on the production: on the basis of situational and social constraints, people make choices out of a web of possible linguistic constructions or expressions. In his paper on the social production of the sign, Kress (1993) outlines a theory of the sign to contribute to developing the field of critical linguistics (CL), or critical discourse analysis (CDA). If semiotics is seen as a research field where semiosis – the understanding, reception and production of signs – is scrutinised, then it is easy to agree with Kress that of all this, traditional linguistics has only been interested in one part, namely signifiers. As structural semiotics sees the sign as stable, so does also traditional semantics and even some varieties of pragmatics within linguistics. There is a

search for a deep structure, whether to explain a grammatical entity, the meaning of a word, or the way speech acts are scattered in a conversation. The truly dynamic nature of real life interactions (and not within a predefined set of options to choose from) is easily lost with the idea of fixed signs or structures. Kress' point about the nonarbitrariness of the form-content (signifier-signified) relation relates to the recent research on syntax-for-conversation in the sense that syntax is not seen as 'innocent', but sometimes has a link to what is actually meant by the utterance (see e.g. Ochs et al. 1996).

Jensen (1995) discusses the Peircean Immediate, Dynamic and Final interpretants, echoing the same ideas as put forward in critical discourse studies. Immediate interpretant is closely connected with discursive difference and Firstness, Dynamic with interpretive difference and Secondness, and Final with social difference and Thirdness. All three share the idea of meaning potential. Immediateness is the structure of sign that gives it its inherent vagueness. Dynamicity signifies situated interpreting, where the sign's actual effect on the interpreter comes to fore. Final interpretant is connected with performance and habit-change, i.e. with social readiness to act; Jensen emphasises that the social difference means that semiosis reshapes its object (1995, 14), i.e. 'reality' is reproduced and transformed in semiosis (again, "expression organises experience" (Volosinov 1973, 85)).

Situated sign production and reading

Kress (1993) attacks the linguistic research which does not see the other semiotic systems (e.g. visual) working in the interpretation of (written) verbal semiotic. The same argument is very relevant for conversations which are time and place bound. In the work of such conversation analysts as Goodwin (1979) and Heath (1986) the situatedness of conversationalists and conversations are clearly shown. Belonging to the CA tradition (see 2.2.1), Goodwin and Heath try not to treat any detail in the interactions as irrelevant *a priori*: the context is not seen as something extra, but as an intrinsic part of meaning production, "a series of interrelating semiotic systems", to use Kress' wording (1993, 187).

The dynamic nature of the sign as produced and interpreted by people with cultural, social and local histories (Kress 1993, 174), ties in neatly with the whole problematics of meaning and interpretation and understanding in conversations. When we are reading or conversing, we are doing something different from passive reception of signs. Goffman (1971) talks about 'focused interaction' when people engage in doing something together. The concept could be expanded to people's solitary activities, such as reading or using a computer, or whenever the product that is being 'used' is a semiotic system created by fellow human beings. But clearly face-to-face discussions are different from face-to-screen or face-to-page interactions, not only because of

the possibility of immediate checking of interpretation, but also due to the potential vulnerability to the participants of the interactional situation in which the two (or more) living histories meet.

Both in spoken and written language, the sign producer's social positioning, their relationship to a particular object or event *in situ*, is reflected in the sign (Kress, 1993). For instance, in Example 2(2), one of the participant makes the remark *this one speaks American too*, thus comparing the language on the screen, the 'this one', to himself and the other human participant (both were American). Kress says that "all signs are metaphors, hence code ideological positions in that they realise the social, cultural and therefore political position of their producer" (ibid., 174). However, signs, most of the time, mean different things to the producer and the reader, partly because of the differing positions on the part of the reader. Kress does not describe exactly how the penetration of all the positions mentioned above can be analysed in the sign. He makes a distinction between the production and reception of signs in claiming that producing a sign is a motivated act; receiving a sign can be troublesome according to the differential distribution of power: "All signs are always transparent to the makers of the sign; and all signs are always opaque to some degree to the readers of signs" (ibid., 180).

If we consider a conversational situation to have the same inequality of production and reception, then our turns-at-talk would be analyses of the prior talker's sign or utterance or turn-at-talk, and another opaque piece of language for our hearer to react to according to his or her conditions of reception. We could claim that it is in these negotiations of meaning that the individuals are reproducing their cultural and moral values, their sense of society at large and their relations with the persons in question. Turn-taking systems and their maintenance are practices which produce relations between 'self' and 'other', difference and symmetry/asymmetry; for instance, institutional contexts and roles are reproduced through asymmetric turn rights. This is why conversation analysis is such a powerful method for investigating institutional interactions in which the reproduction of authority becomes analysable.

The relationship between sign production and reception is not unimportant. Traditionally it has been seen as unproblematic, whereas nowadays the common understanding seems to be that meanings are locally negotiated; Kress is clearly of the latter opinion as well. However keen researchers are to give the moment of interpretation more and more weight, there seems to be no consensus about what the situated understanding implies.

When social semioticians want to bring the real world into their analyses of signs, they attempt to expand the context of the situation under scrutiny: instead of concentrating on conversationalists as individual processors of language, we should see them as signs produced by their own life histories and culture; instead of scrutinising the linguistic features of a newspaper text

we should take into account the layout of the page (pictures, boundaries), and the whole context of the political situation and society. This, of course, is an improvement to many of the 'language-in-use' theories and studies which are not willing to see the contexts of interpretations as encompassing other things than language. But what is lacking from this kind of contextualisation is one aspect of actuality which must be crucial for the on-going interpretation: the sequential nature of interaction. Neither Hallidayan nor critical social semiotics pay any attention at all to how utterances are not just interpersonal or ideational, or conveying a special interest. Utterances are produced in a context of talk which might have an effect on what is said in that specific moment. In other words, the interpersonal meaning, how the 'other' is interpreted, is detectable in context, and the sequential context also effects how the meaning is constructed. Understanding the sequential nature of meaning construction also takes weight away from the utterance: it is not only the linguistic entity but its point in interaction that matters.

The coconstruction of meaning gives the researcher materials to analyse how the participants not only orient to the course of action they are building together, but also to each other, to moral order, and to each other's identity (the 'other'). For instance, in pair work experiments (e.g. Chapter 6), the students do the work of getting to know each other's background during the task, they get to know the other as an interpretant (see 2.3.), what her experience of this sort of dynamic situation is:

2(5)

- 1 B: **ooksä aikaisemmin tehny tämmösiä tarroja?** ((soft voice))/
- 2 have you made these sort of stickers before?
- 3 A: **e-n**, ((gaze: instruction sheet))/no-o,

We cannot know the intention, the conversational goal, of B when she posits the question of previous knowledge to A. The directness (no hedging) of the formulation implies that the query is not potentially face threatening: it does not question A's present abilities, but is an information seeking question. If the question had been asked in the middle of their trying to solve a perhaps differing interpretation, then it would be a potential face threat: as in a conflict situation the same question could be interpreted as a challenge of the other's knowledge.

Conversation analysis with its strict demand of demonstration of participant orientation to phenomena as a precondition of them being analysable by an outsider is very helpful here as a means to analyse semiosis. Instead of trying to guess what the participants are thinking in interaction, or taking the turns to be signs referring to a 'truth', the intricate web of meaning making and interpretant building is detectable in the turns-at-talk and other actions.

The sequential nature of semiosis

Ponzio (1985) explains how the notions of semiosis in Peirce and Bakhtin relate to each other. Peirce was interested in how the interpretation of signs is a locally accomplished process: Bakhtin, on the other hand, focused on literature as a special domain in which the writer can disengage from the consequential nature of everyday spoken interaction.

Both reject the idea of sign as stable, with a one-to-one correlation to its interpretant. Meaning is seen as a set of all the interpretants the sign can produce. Contrary to Halliday and his idea of a meaning potential, however, meanings are not a given set from which people consciously or not choose theirs; instead, every occasion of sign use creates the meaning anew.

Conversation analysts look at everyday interactions and how the inference work by the participants is done *in situ* – how the sequential organisation of talk demonstrates the understandings of the conversationalists. Interactions proceed via turns at talk and other meaningful actions, and how people understand the other's actions is displayed in their own in consecutive actions and talk. These turns-at-talk or turns-at-action at the same time take the interaction forward, i.e. they are both analyses of the previous interactant's actions and a new utterance to be analysed. To relate turn-taking to (social) semiotics, we could relate a turn to an interpretant and a sign, thus bringing the cognitive ideas of Peirce's dynamic interpretant outside the mind and making it analysable. Natural interactions offer us endless material to research how semiosis is really about 'these signs here now' and how signifying or the inference work is made visible in talk, if we consider the turns-at-talk in the same way as do CA analysts and semioticians.

Coconstruction of meaning in talk, the endless formation of interpretants, is accomplished by people. Thus the interactants bring their moral character and cultural values, their life histories, their multilayered interpretants to the inference process. Sometimes they leave the situation with new ideas, new interpretants prevailing: people have learnt, they have changed. This change can then be detected from their future discussions.

2.4. My approach: Analysing semiosis as materially based sequential meaning making

The agenda of semiotics has been to explore the system of all human-made signs, and social semiotics has expanded the scope to visual phenomena and interaction (Fairclough and Wodak 1997). Because social semiotics regards both interpretation and material circumstances as essential in researching language and other semiotic formations, it is a very good framework for studying people's encounters with mediated language (in its various material

manifestations) in language technology environments. But social semiotics lacks a rigorous method for studying the impact of the communicative environment on the ongoing interpretative work. Smith and Whalen discuss a similar lack of methodology: “the research we are working with here calls for methods of analysis that can analyze the interrelations of talk and text as sequences of courses of action, where the text is recognized a locally occurring “utterance” produced or activated by a reader or readers as an integral part of such a sequence” (1995, 2). On the basis of analysing relays of emergency calls, they “propose finding the method of reading in the local historical settings in which it enters a course of action” (ibid., 29). The present study will explore situations of reading (and hearing; also speaking and writing) in which the activated text is oriented to, for instance as instruction or as a request to do (or say) something. Therefore, the problematics of research methodology are the same as for Smith and Whalen.

As argued in the section above, the conversation analytic idea of interpretation as sequential can be adopted to the Peircean triangle of semiosis. The scope of domain can be extended from conversations to mediated communications. To quote Smith and Whalen: “The text is no less integral to the local ordering of sequence of action and its component conversational sequences than any utterance is in a conversation” (1995, 31). This way, the Bakhtinian idea that texts are always dialogic (e.g. Bakhtin 1986) is addressed, but also the intricacies of situated action are accommodated. The following diagram will serve to highlight how the Peircean sign, object, interpretant can be related to an ongoing semiosis of talk-in-interaction:

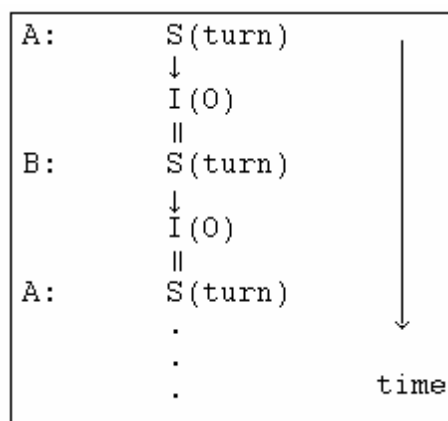


Figure 2-1

A and B discuss something, and in their discussion the turns (i.e. signs, S) encompass not only a linguistic and other signs but also the object (O) of one's own turn and, at the same time, an interpretation (I) of the previous speaker's turn. In this fashion, the highly mentalistic Peircean triad of negotiating the object (layering the interpretant) can be externalised and made available to the analyst.

Witte (1992) applies the Peircean triangle to explain the writer's intentions through text, context and intertext which stand in the same relation as sign, object, and interpretant. Witte regards the reader's role as one in which the sign-context-intertext triad is contextually unpacked: "As in the case of a writer, a reader's internal representation of the text would shape and be shaped by his or her internal representation of the context that the text enters and by the reader's representation of the intertext(s) called into play through a situated act of reading during which the intertext(s) called up would shape and be shaped by the reader's representation of both text and context" (Witte 1992, 288). My claim is that we can find at least glimpses of this internal semiosis in pairwork situations for instance. In other words, the hypothetical representation becomes an analysable turn-at-talk, gesture, or facial expression, the larger interactional context of which can aid in detecting how exactly the shaping of the contribution got to happen. This way, the focus can change from text-internal potential (causes of) misunderstandings to actual problems in the temporally shared social reality, the reasons for which can be detected in the intricacies of the text-user interaction. For texts that guide the real-world activities of the reader, the context of text use becomes an integral part of text meaning: a computer tutorial deals with the user-reader in a computer context about a computer context. In pairwork, the computer context is shared between the users and so is the situated reading of/acting upon the screen. The recordings of this interpretative work give the analyst first-hand information about what textual, visual or other resources are utilised in deciphering (what to do with) the interface; and also which semiotic resources cause misinterpretations of the author's intent. Out of solitary interactions with predesigned texts, those from automatic telephony services offer data in which the human participant's meaning making work becomes available only through talk because dealing with the system is done solely by spoken language.

When texts become part of social activity, Smith and Whalen (1995) note that the term 'text' does not convey the materiality in the same way as does, for instance, 'document'. They prefer 'text', however, because it is "a bridgehead connecting the analysis of texts as constituents of social organization with literary and cultural theory" (Smith & Whalen 1995, 3). My decision is to talk about 'textual objects' when the materiality of the text is underscored. Literary theory will be touched upon in Aarseth's (1997) notion of 'cybertext' and its implications to acting in the world (vs. 'moving ahead' in fiction).

In language technology environments, the human participants' ongoing semiosis of the various signs in the situation — the computer, the screen, the instructions, the manual and the other human participants — is made public via talk and nonverbal signs. The texts that are available in pair work, for instance, are, from the participants' point of view, either stable like the printed ones in paper manuals or appear on the screen as a result of the

participants' previous actions. The apparent dynamic nature of the interactive ones is something different from the stable print of manuals, thus combining the fixed nature of text and the fluid nature of interaction. There is a continuum of (overlapping) participation frameworks: the human encounters permanent texts, 'interactionally' appearing texts, and other participants. When the user-readers are involved in practical activities, it is possible to detect the processual nature of understanding, or semiosis, and to show how, for instance, stable texts like computer manuals are a resource which is interpreted according to the nature and order of the sequential activities in which the participants engage. In all the case studies, textual directions are followed, and users' interpretation checking is shaped differently according to the interactional resources each textual object offers.

Conversation analytical (CA) work on repair deals with uncertainty, if the notion is expanded: the object of action or activity being problematic, self-initiated other-repair as a concept could be stretched to cover the orientation to potential trouble sources as well. In these sort of uncertainties, language and actions are not separated from each other; repair is that of action, be it linguistic or other. The participants are orienting to trouble also as possible, as forward looking, rather than trouble as detected, as backward looking (cf. Raudaskoski 1992, 172); they would willingly offer multiple meanings to their hearer – the dialectic between 'self' and 'other' "also comes to play during the formulation of the expression, thus determining both its content and form" (Ponzio 1990, 262). But in language technology environments, multiple meanings are not sought for because of the nonsensitivity of the mediated 'other' to the situated interpretation. However, the relation between the interpreted (sign) and interpretant (another sign, an analysis of the sign) is not conventional in the sense that the environment in which the relation occurs is different from most communicative events for the participants. Therefore, the situation also is one of learning, to overcome the difficulties of extrasign and semiotic materiality that the mediated 'other' brings into the interpretative situation.

Petrilli's account of materiality (see 2.3) keeps extrasign and semiotic materiality separate. However, language also has visual and audio features, i.e. extrasign physical materiality that does play a role in its interpretation. For instance, computer interfaces often display stable (e.g. solitary, icon-type words) and interactive (e.g. error messages) language. Thus, physical materiality is part of the semiotic system.

Extrasign physical materiality also concerns participants as physical entities in time and space: they act and interpret the (language and other) actions of others spatiotemporally. In language technology environments, the 'other' is virtual in the sense that the originator of the words is not present. In Extract 1(3/5), the manual 'interfered' with the task through printed language; in Extract 1(4/6) above, the time (though lagged) but not the space was shared: the students could see a video picture of the 'others', and of

themselves. For the human participant, the semiotic and physical materiality of language technology is different from that of fellow humans, but in neither encounter should the words be looked at as a sole carrier of meanings. In a computer-mediated communication situation, the extrasign physical materiality encountered still tends to be new (though the Information Society is meant to make them everyday practice), and therefore, though the linguistic items appearing on the screen, for instance, are familiar as such, they have to be learnt anew because of the new interactional situation. By close inspection of interaction at language technology environments it is possible to detect whether and how extrasign physical materiality might play part in the sequential meaning making, effecting interpretant formation.

Work on artificial intelligence and cognitive science introduced schemata and frames to explain situation specific actions and talk (e.g. Minsky 1975). Thibault criticises schema theory on the basis of how the 'other' is seen: "The normative psychologistic basis of this research presumes a culturally nonspecific other with whom the individual is assumed to interact. In discursive practice there can be no such thing as a nonspecific other, for all discourse and all interactants in discourse are situationally and socially specific." (Thibault 1991, 166.) His argument can be seen to challenge the Barthesian way of treating readers of texts as communicating with words only (see 2.1.1), thus stressing the social side of social semiotics when encountering technology-mediated 'others'. It is important to note here that both structurally and socially oriented semiotics give an *a priori* rendition of how the reader understands the 'other'. In the present work, it is an empirical matter to determine how 'self'/'other', participation and uncertainty are negotiated in courses of interaction and artefact use. In the data analysis, various semiotic modes of interaction are researched, with the aim to explore how the users coconstruct the interactive situation such that the written or spoken language produced by the other human participant(s), the program, or the manual, is incorporated into the unfolding meaning making. Through their linguistic and other activities, the participants are continuously exhibiting their understanding of what is going on. Reformulations and repair are explicit examples of meaning checking and negotiation. Whether the language or other signs interpreted *in situ* are asynchronously or synchronously produced, the user has to make them work synchronously in the activity s/he is involved in. The detailed analyses of the users' linguistic and other turns show how the differing semblances of participation and differing interactive resources result in a variety of dissimilar ways of interacting with the 'other'. Mediated communication distorts the availability of the 'other' as a fully fledged communicative partner, but the unavoidable distance to the 'other' (either in space or in time or in both) also provides new resources for the human participants' interpretation of what is going on. The detailed analyses of interactive activities in situated activity systems with various kinds of language artefacts show, for example, which linguistic and

visual resources the human participants resort to when they encounter a computer dialogue (spoken aloud) system, an interactive hypertextual tutorial program, a word processor and its manual, and a (visually, aurally, and textually) mediated human 'other' (and oneself), all of which engage the participant in linguistic or other practical action, and all of which shape different trajectories for checking and repairing the user-readers' interpretations.

2.4.1. Time / Space / Language

As spatiotemporality of language use has been a recurrent theme above (cf. 2.2), the relationship between time, space and language is discussed in more detail in this section.

Though time and space are mentioned in many research fields which focus on language, the concepts are not often elaborated on in the actual method of analysis. For instance, the social semiotician Robert Hodge defines what he calls materialist diachronic analysis of discourse, stressing the primacy of time as a factor; even synchronic phenomena occur in diachronic time (Hodge 1988, 103–104). He is searching for a way of connecting the fleeting moments of time of conversations to the epochs of history. From the structuralist paradigm Hodge stresses the relations and transformations¹⁶ of structures as taking place over time. Agency is important: "Every discursive act, on every level, has agents and objects, and these constitutive elements should be specified as far as possible in any analysis of discursive events or processes" (Hodge 1988, 104). Transformations can affect both the paradigmatic (choices) and syntagmatic (production) levels, both in production and in reception. Hodge seems to treat texts primarily as historical traces, not as products by one or several social agents; the alternative histories that every text or discursive event is composed of resembles Bakhtin's heteroglossia.

Historicity, detecting effects of past interactions in a present one, indicates experience, which in semiotic terms could be called a multilayered sign (see 2.3). Past encounters are the basis of learning, resulting in people finding places not only as spaces for three dimensional activities but also as sites which are meaningful (Canter 1985, 173). Some researchers explain the meaningfulness of locations on the basis of scripts and schemata (Eiser 1985, 199). However, the scripts and schemata approaches often do not concentrate on the lived-in experience of an event, but provide a framework for there being certain phases in certain types of interaction (e.g. what happens in

¹⁶ Transformation, then, refers to the change or development in the meaning of an object; for instance, a psychiatric session with a patient is turned into a written document which is the psychiatrist's depiction of what happened in the original discourse.

restaurants). But within participation framework studies, research has been conducted on how participants in a focused interaction¹⁷ create a space between themselves which can be face-to-face, L-shaped or 'ear-to-ear' (e.g. pairwork at a computer; cf. Extract 2(4)). "A participant wishing to change to a new frame may precede any actual change by small manoeuvres in the direction of the new position that would, if completed, constitute a position suitable for a different kind of arrangement" (Kendon 1985, 241). Thus, Kendon has been able to show, by detailed observation of participants' behaviour, that participation framework is constituted via actions, not just language.

In section 2.2, *kairos* as the lived time of the concrete body in concrete space was introduced, *chronos* referring to the 'scientific' time of measuring and clocks. In the following, Erickson and Shultz's (1982) treatment of the theme of temporal interaction is discussed. It gives *chronos* an interactionally important part, as well. They see *kairos* to mean "*the right time* – the *now* whose time has come" (Erickson & Shultz 1982, 72), in other words, the interactional sequence, the here and now of the event is managed delicately. For instance, there are appropriate ways in an emerging discussion to shift topics. In this interactional work, reciprocity, the immediate past and future history of the interaction has to be taken into account (71); "events are interactively organized through time and in space" (Goodwin & Goodwin 1992, 93).

"*Chronos*, in contrast, refers to duration of time in the ordinary space-time world apparent to the senses." (Erickson & Shultz 1982, 72). This real time is detectable in rhythm (staying in phase), which is measurable from the conversational flow. Rhythm "seems to permit a complementarity of interactional inference that would otherwise be impossible to accomplish during the course of conversation" (ibid., 96). Thus, *chronos* is important for interpretation: the speaker/hearer roles seem to be merged into one rhythmic entity instead of the hearer responding to the speaker's linguistic and other actions: "The synchrony among conversationalists seems not to be a matter of stimulus-response organization at microsecond intervals, but of mutual *entrainment* of all conversational partners within an overall pattern of rhythm" (ibid., 72).

Social interaction takes place in both *kairos* and *chronos* times, and therefore "analytic abstraction, literally a "pulling out" of relationships of sequence from their temporally local contexts, is not appropriate for the construction of theories of social organization of *performance* in interaction, nor for theories of the underlying *social competence* persons must have in order to enact such performance" (Erickson & Shultz 1982, 98). If either timing is not right, interactionally a 'wrong thing' can happen at the right time, or a 'right thing' at the wrong time. "It seems that it is the sequential (*kairos*)

¹⁷ Kendon (1985) calls this a f-formation.

organization of communicative behaviour in real time (*chronos* time) that enables conversationalists to engage in fluent discourse and to regard one another as conversationally competent" (ibid., 76).

The importance of rhythm in face-to-face spoken communication brings forth the question whether the *chronos* of technologically-mediated communication is out of joint, whether the loss of rhythm with the 'other' causes interactional problems. Social agents are aware of the difference between encountering texts and spoken interaction. However, the new media bring textually-mediated synchronous or asynchronous 'others' to them, conveying features of face-to-face encounters. Are the *kairos* and *chronos*, the social competence in encountering a technologically mediated 'other' the same as in face-to-face settings, and if not, how and why might they be different? What is the lived-in time of *kairos*, the appropriateness of each interactional step, like when interacting with a not copresent 'other' who is present as (or represented by) written text. *Chronos* time, or timing, is important when accomplishing something in a technologically-mediated environment (e.g. in real time, leaving a message into an answering machine, learning or completing a task through an assignment): "In interaction – whether natural or simulated as such – timing is everything" (Boden & Molotch 1994, 265). However, in encountering machines, people quickly learn what to expect from the system in terms of timing (in the same way they adapt to the pace of an elderly person or a child). When the material circumstances of the interpretative work are different from the recently much researched face-to-face, copresent interaction, and if semiotic and material processes are interdependent, then technologically-mediated communication cannot be in one-to-one relation to face-to-face encounters. For instance, it is possible for the recipient of a manual's instructions to become more of an "observer contemplating on utterance and trying to make appropriate sense of it" (Goodwin & Goodwin 1992, 83-84), something which is out of the question in copresent interaction. This is why a transcript of a user-system interaction which only depicts the linguistic 'turns' by the user and the system (i.e. as if a face-to-face conversation had taken place) as depicted, for instance, in Luff and Frohlich (1991), is not a good enough representation if the complex semiotic environment and its effect on the interaction is researched.

Tuan (1978) reports on human experience in which space, place and time are overlapping categories: "Here is now and there is then, and then means either a time in the past or in the future" (Tuan 1978, 11-12). According to Tuan, 'there' cannot be 'now'. However, in on-line new technology environments such as video conferencing, 'there' is used in a shared time frame. But, as written texts are normally traces from the past, signs of 'then', problems will occur when these traces are incorporated in the encounter as if they were 'here/now' instead of being marks from 'there/then'. Thus, material processes "form the ground of all possible and actual change in the relations

of semiotic practices" (Lemke 1993, 250). 'There' in the world of new communication technologies no longer implies 'then'.

In the analysis of the case studies in Chapters 4 to 7, the intertwining of language, space and time is of special interest because the encounters with the technology-mediated 'others' provide research material for (trying to do) appropriate or (ending up in) inappropriate *chronos* and *kairos* times, for reading vs. encountering, and for the visual and audio semiotics of space, time and sound. The hypothesis is that the concept of the 'other' (at some level, at least) depends on whether the managing of time, space and language is successful or not.

2.5. Summary

This chapter discussed humanistic approaches to 'other', keeping conversation analysis as a point of comparison. The aim was to comprehend how complicated the nature of communication in language technology environments is: the mediated 'other' (and sometimes 'self') is met in materially and semiotically specific circumstances in which the emerging language (through print, typing, or sound) of the 'other' is produced and interpreted spatiotemporally.

In the following chapters, four case studies conducted in technology-mediated communication environments, are discussed in the light of the present and preceding chapters to find out how the 'other' is constructed *in situ*, making use of the tools for analysis consulted so far. The analyses aim at exploring the so far taken-for-granted complex interactional environment in which the 'other' is not present in body but in language (and picture), the representational nature of which differs in the different environments to be looked at (sound, synchronous typing, asynchronous writing, and printed text). Thus, computer-mediated communication and human-computer interaction are researched as situated sense making, available for inspection through participants' reading, seeing, hearing, speaking, typing, gesturing and other ways of making the sequential semiosis detectable. Through detailed data analysis, the relevance of various semiotic fields (Goodwin, in press) for acting in language technology environments can be reliably discovered.

The four case studies represent differing semiotic constellations, in which the importance and role of time, space and language change from synchronous sound (and video) based contact to asynchronous reading. In the first three cases, the human participants are doing a task with a computer – hence the situations are intense in (the one-sided) meaning construction; the fourth case study is a 'real' computer-mediated communication situation in which the complexities of constructing mediated (group) identities and

meanings arise. Only in the first case study are the program(mer)'s intentions clear, as the design was done by myself.

3. INTRODUCTION TO THE CASE STUDIES

In this chapter, the case studies are discussed as experiments and as naturally occurring data. The first section concentrates on the interactional situation and the problems that may occur when experiments are used as research data. The second section will focus on the questions that the technologically and textually mediated 'other' will bring to the fore, and how the theoretical discussions of the previous chapters are linked to the present data.

3.1. Experimental studies

Except for one all the case studies were experiments, i.e. the participants were asked to do something different from their daily routine: they were 'prearranged interactions' or 'interactional experiments'. However, all the experiments (as did the naturally occurring case study) took place in university surroundings, which meant that the participants were in their natural everyday environment. In each setting, there were always a minimum of two human participants and interpretations were easy to trace (about interactions with a computer interface, cf. Hutchins 1995, 270) and only in the case study TELEPHONE was the human participant working alone¹⁸.

The places in which the data were gathered were the following: 1) TELEPHONE, an experiment: three secretaries in their own offices (everyday working environment), 2) TUTORIAL, an experiment: university lecturers in a computer class (not their everyday surroundings but they knew about the existence of the computer class), 3) TASK, an experiment: students in an office at the university (cf. cases TUTORIAL and TEACHING), and 4) TEACHING, natural (though not an everyday teaching situation): students in an office turned into a virtual classroom (cf. cases TUTORIAL and TASK). It is important to recognise that the surroundings were familiar, because "the consistent patterns of space use can be seen as reflecting the meanings different places have for their users, which in turn relates to what they see as

¹⁸ Although in TELEPHONE there was a human 'other' at the other end of the line, she was, however, only reading aloud the system's messages and typing in the user's turn.

the primary purposes (or functions) of those places" (Canter 1985, 181). Also, the actions they were asked to do (making a phone call, using a word processor (tutorial) program) were everyday tasks, and not, for instance, memory tests. The degree of experimentalness was fairly low as the participants were not asked to play the role of anybody else but were invited to accomplish something as themselves (cf. Ginsburg 1985, 265–6). As university staff and students are generally open to experiments, the tasks were largely 'everyday' behaviour for people in a university context, and as they could hardly be called laboratory tests there was no need to present concerns like the following: "The representativeness of contexts for the elicitation of behavior in laboratories is seldom addressed" (Hutchins 1995, 287).

In the language-mediated communication, *lingua franca* English was spoken in the last case study (TEACHING) between the Swedes and the Finns. In the TASK case, the 'other' mediated by the computer and the manual functioned in English, and the students were Finns fluent in English, and so the situation was one between 'natives' and nonnatives. In the two first cases (TELEPHONE and TUTORIAL), everybody was a 'native speaker' except for the communicator in TELEPHONE. However, the distinction between native or nonnative talk is not considered important in this study, unless it becomes relevant in the interactions studied.

Social agents' access to meanings, texts, and social situations (see Thibault 1991, 237) is uneven. My case studies are examples of what electronically-mediated texts were accessible to staff and students at universities in Finland and Europe in the first half of the 1990's. However, the computer systems used in the case studies are becoming available for larger audiences with the rapid expansion of computers as text processors and gateways to the Internet in average (middle class) households in the West. It is typical for the computerised productive (vs. experimental as in Laurel 1991, 22) textual environments that, unlike in literature (or experimental interfaces, e.g. games), the meanings are meant to be clearly understandable and not ambiguous.

In the three case studies which were experiments, the participants were concentrating on what they were asked to do, which meant that they were not "involved in a range of different activities simultaneously" (Goodwin & Goodwin 1992, 84). However, the experiments were analysed as sequences of interaction, rather than, for instance, measuring reaction times: "the experiment can be analysed as a communicative encounter, as involving shared Time" (Bowers 1991, 560). The lengths of interactions varied from short phone calls (TELEPHONE) to hours of video conferencing (TEACHING), and the time span within which the data was collected was six years. Out of these time depictions, the most important is the 'internal' time of the unfolding situations, how the *chronos* and *kairos* timings work.

Except for the first case study (three telephone conversations), all the cases were videoed. The usefulness of video recordings is made clear in the following quote:

The use of video to record interchanges not only offers the opportunity for analysis of interaction processes, but it also provides a relatively permanent record of what actually occurred and makes it difficult for an investigator to ignore those occurrences in favor of a more simple theoretical interpretation.
(Ginsburg 1985, 263)

3.2. The case studies as encounters with 'others'

In face-to-face conversations between human participants, differing interpretations of what is going on are tolerated; "on a given occasion of formulating, participants may not explicitly orient to (or treat as problematic) the possibility of competing readings of the conversation" (Heritage & Watson 1979, 129). However, if we think of the computerised 'other' as a participant, it is extremely important that the user(s) have a 'correct' understanding of what is going on. This is because the program running the 'other's' interpretations is very limited: "The unprecedented use of a word, the unexpected metaphor, may lay bare for the first time a sublime truth—but it will only cause the information processor to stumble. For the information processor classifies, stores, links, and searches text based solely on surface appearance—the "shape" of the words—and is programmed on the assumption that these shapes can be mapped to a set of preestablished meanings" (Talbot 1995, 190).

In pairwork situations (TUTORIAL, TASK), both human-computer interaction and that between the human participants are done in a similar material situation. This gives a possibility to not only detect meanings (when they are conveyed to the other human participant or negotiated), but also to see how interactions in these environments differ; communication with the manual gives data for comparison. Even those conversations are limited to and by the interaction with the computer, the initiator of turns. Thus, the following statement can only be reserved to two or more humans' interaction: "While precisely organized at one level, everyday talk is remarkable for its looseness in terms of topic, speaker participation, allocation of turns, and forms of speech". (Boden and Molotch 1994, 268). In everyday conversations, turns can be left unaddressed (cf. Schegloff 1988, 131), whereas in the present cases, every turn by the system is meant to be addressed. The problem with the text based interfaces (TUTORIAL, TASK) is 'overaddressing': treating descriptions as first pair parts of an adjacency pair.

The 'other' can be constituted through accounts (cf. Goodwin & Goodwin 1992). In TELEPHONE and TEACHING, spoken interaction technically allows for projections (ibid., 81), i.e. the participants can predict what is coming next and produce their own assessments. However, in text

based communication of TUTORIAL and TASK type, the users cannot display their assessment as a collaborative activity, for instance by finishing off a turn at the same time (because the text appears as chunks, instead of word by word, on the screen). However, they can still give accounts of what is appearing on the screen, but those accounts are usually made to the other human participant in the room. In this work, gestures are also used to literally point at some relevant detail in the 'other's' text. They are able to predict what comes next by, for instance, doing a relevant next keystroke before the prompt comes up for that¹⁹. This resembles projection: you give the answer before the question is asked. At the same time, it displays learning: the user already knows the 'interaction style' of the 'other'.

The case studies will show whether the technologically-mediated 'other' will be understood as demanding the user to use the system in a certain way, or is the 'other' seen as offering the user a space for interacting (cf. Kress & van Leeuwen 1996, 254). Smith and Whalen (1995) concentrate on how descriptions of individuals in telephone calls to the police get transformed to text and back to speech (within a short time space). This work is very routine, whereas I focus on new (for the users, as well) communication environments and tasks with requests (to do, verbally or by acting out the request). Though the requests can be originally produced with a long temporal delay (cf. computer programs or manuals), the text is enunciated and acted upon *in situ*. As in Smith and Whalen's case, it is the electronic (even if spoken aloud) or paper-based language of the text object (computer, manual) that 'drives' the exchange (Smith & Whalen 1995, 12). Although all the predesigned systems of the case studies cannot particularise the user-reader, one of them (Learning Word 5) uses 'emotional' language such as *Great!* to give positive feedback to the user (whose name is also asked at the beginning of the TUTORIAL but only used once at the beginning of it). The systems are not given any identity (cf. Weizenbaum's ELIZA) and they do not use 'I' to refer to the system, though 'you' can be abundantly used to refer to the user. Therefore, the aim is at an 'institutional' or 'faceless' encounter.

However, in all the case studies, the relationship between the user and the 'other' is meant to be dialogic, as both written and spoken texts of the 'other' refer to the reader/hearer as 'you' (see e.g. Silverstein 1976, 36). The apparent dialogic nature of the written and electronic messages becomes a problem because the other resources for sequential turn-taking are missing: "Sequence design is at risk with paper (or even screen) communication, and misreadings are harder to detect and repair" (Boden & Molotch, 1994, 271). Out of the language technology systems studied here, the TELEPHONE could be called a 'dialogue' system, but it is much less sophisticated than the ones

¹⁹ For instance in the e-mail program Pine, it is possible to strike the 'Y' or 'N' keys in a row before the actual inquiries come up on the screen: I hit the 'Q', 'Y', and 'N' keys and see the questions 'Really quit Pine?' 'Save the 60 messages in received?' appear afterwards.

developed within the artificial intelligence community. However, as Nass et al. point out (Nass et al. 1994, 77), fairly simple computer systems tend to be treated as 'others'.

3.2.1. *Creating intimacy*

Other practices that can be monitored in the language technology environments are the opening phase, – are they other, self or shared knowledge oriented (cf. Cheepen 1988, 48 & Laver 1975)? Other-orientedness is necessarily lacking from language technologies (except for simple questions for name, for instance) because other-orientedness cannot be used as a basis for further discussion (due to the computer's limits²⁰). Even though there might be pressures to do phatic communion with the 'other', the users might treat attempts at other-orientedness by the system's 'other' as not wanted:

Within English-speaking cultures (and perhaps all cultures) phatic communion seems to be an almost universal habit indulged in during the opening phase of interactions. Nevertheless there are some situations in our culture where we normally avoid phatic communion during the opening phase as being inappropriate to that particular type of situation. An obvious case is where the interactants have already met that day, or at least within the last six or seven hours, and have already indulged in extended phatic communion in their first meeting. Another case is where the roles of the interactants are already very clearly defined, as in situations such as a university lecture, buying a railway ticket, or talking to a telephone operator. In all these situations, the role structure of the encounter is known to the interactants in advance

(Laver 1975, 218).

Out of the case studies, in the TELEPHONE there was clearly some pressure by the secretaries to 'do the opening phase'. This means, that out of the 1) telephone ring, answer 2) identification/recognition, 3) an exchange of greetings, 4) initial inquiries sequence (cf. Schegloff 1979), the system would do the first and bypass the rest three by answering *Hello – what do you want to do?*. Two out of the three users would do the second phase, either introducing themselves or recognising the communicator. In the TUTORIAL case study, the identification phase (asking for the user's first name) came fairly late after queries about the monitor and the mouse, even the welcoming screen gave the impersonal directive to remove any write protection from the disk before making the other-oriented query of asking the user's name. The technicality of this identification phase is reflected in the matter-of-fact attitude the users have to the query: typing in the name and the following thanks with the name, *Thanks, Tim* (the only time the user's name appeared on the screen) were passed with no commentary. In the case TASK, there was no

²⁰ So called expert systems are built in order to find out about the user as much as possible, so their design idea could be called other-orientedness.

introductory phase for the participants, which meant that the getting to know each other was dispersed within the encounter (*have you done this before?*) the program being a text-processor and the manual a neutral collection of instructions, these 'others' were not needed to be 'introduced' as if individual entities²¹. In the last case, TEACHING, there was an opening phase of introductions between the two sites in the first meeting, the rest were started in a more straightforward manner. It seems that the textually mediated 'others' in the encounters with a predesigned (interactive) program (cases TUTORIAL and TASK) were treated as more institutional, i.e. not needing a clear 'opening phase'.

3.2.2. Visual and interactional sharing

When multiparty interactions in computerised surroundings are researched, the conversation analytic tradition of incorporating the whole of the context as potentially meaningful for the interaction offers an effective method for analysing how the participants understand the technology they are surrounded by. As with any environment in which artefacts are used, gesturing is an important part of the activities: "Thus a single gesture performs different actions at different stages of its production, systematically serving to provide a framework of participation for its own performance, and working with and alongside the talk to shape the way in which the activity is dealt with, interactionally" (Heath 1992, 119). Gestures are often forward pointing: "In this respect they share the functions of other prefatory actions studied by conversation analysts" (Streeck & Hartge 1992, 137). This means that the linguistic output is better understood when the accompanying gestures are taken into account. Also, more can be learnt about the role of gestures: "The range of kinds of gesture has barely been explored, and existing typologies of "nonverbal behavior" are equivocal at best" (Streeck & Hartge 1992, 136).

The importance of nonvocal features of interaction is clearly expressed by E. Goffman, who wants to abandon the 'speaker' and 'hearer' classification as they stress the importance of sound too much.

In the management of turn-taking, in the assessment of reception through visual back-channel cues, in the paralinguistic function of gesticulation, in the synchrony of gaze shift, in the provision of evidence of attention (as in the middle-distance look), in the assessment of engrossment through evidence of side-involvements and facial expression - in all of these ways it is apparent that sight is crucial, both for the speaker and for the hearer. For the effective conduct of talk, speaker and hearer had best be in a position to *watch* each other. The fact that telephoning can be practicable without the visual channel, and that written

²¹ The selection of animated 'wizards' in Word 97 is a conscious step to individualising the guiding 'other'.

transcriptions of talk also seem effective, is not to be taken as a sign that, indeed, conveying words is the only thing that is crucial, but that reconstruction and transformation are very powerful processes.

(Goffman 1979, 6)

According to Goffman, all interaction is mediated bodily, i.e. we orient towards a unit of talk and disorientate from it also with our postures, gestures, etc. Thus nonvocal phenomena are of importance when a face-to-face situation is examined. Streeck (1988) shows that gesticulation and talk are aligned, and that gesture usually comes before the talk related to it. Goffman calls nonvocal orienting as being in a *state of talk* (1979, 7), though one may not be saying anything. So, we do not talk only through language, but through our bodies as well.

The human participant is the only one who can change the participation framework (see Goodwin & Goodwin 1992, 92) as the spatial organisation of bodies. A frame, according to Kendon, "comes to be placed around the actions and utterances of the participants, which both determines the sense in which they are to be taken and serves to define whole ranges of possible acts as irrelevant (as not to be included)" (Kendon 1985, 230). The frame that the computer-mediated 'other' and the user occupy is clearly distinct from that of copresent interactants, and the most critical difference is in the sharedness of the frame: if the 'other' is not only mediated but also generated by the computer, then the work to make the frame shared is left for the user. This resembles interaction between an adult and a toddler, because whatever the baby does the adult renders meanings to the actions.

Also, it is the human who can and will move and focus on different things in the interactional space: "The transactional segment is the space into which the individual addresses his gaze as he carries out his line of activity, whatever it may be; it is the space from which he immediately and readily reaches for whatever objects his current project may require he manipulate; it is the space immediately in front of him that the individual projects forward and keeps clear if he is moving" (Kendon 1985, 237). In the case studies, the primary objects to manipulate, or to give attention to, are 1) (talking on) the telephone, 2) the computer, 3) the computer and the manual, and 4) a selection of computers (the freedom of changing the attention increases from 1 to 4). Though telephone and computer as artefacts were not unfamiliar for the participants, the situations of cooperative interaction in new technology environments were new for the agents. This means that at least implicit learning is taking place, and, "where there is the need for learning there is room for error" (Hutchins 1995, 272). In semiotic terms, the interpretant needs constructing or adjusting. However, it is wrong to assume that interacting in new communicative environments and with textually/technologically-mediated 'others' would be one continuous failure.

An example of a Peircean semiotic indexical (see 2.3.) can be accompanied with a gesture: the pointing finger establishes a connection

between the object and the sign. Hanks (1992) discusses indexical origo, the origo that language use creates for the participant. Deictic expressions (e.g. *this, here*) are often supplemented or even replaced by gestures that link them to the referent. According to Hanks, the referential salience is dependent on "the degree to which interactants share, or fail to share, a common framework" (ibid., 67). Hanks divides sharing into participant access (spatial, perceptual, cognitive) and participant domains. The domains are the symmetric Common ground (cf. Labov's AB events), and the pragmatically asymmetric Speaker, Addressee and Other, who is "a non-Participant in current speech event" (ibid., 68). Hanks' three-fold division points at possible confusions and contradictions in the roles of the computer-mediated participant: 1) is the 'other' understood as Speaker and as Addressee alternatively? 2) is the 'other' primarily an Other? 3) are the three roles mixed and not possible to distinguish from each other? and 4) are the three roles not enough to explain how the 'other' is constituted in technology-mediated communication? As technology-mediated communication is often realised by written text or typing, the degree of representation grows when the text has been produced asynchronously, even if it appeared synchronously (cf. TELEPHONE, TUTORIAL). Hanks mentions that voice is an auditory signal of origo (ibid., 66), i.e. the speaker relates his or her deictical or indexical expressions to his or her spatial and temporal point of view. Representationality of the 'other' grows with asynchrony: 1) In face-to-face communication, the interactant is not represented but is directly observable; 2) In on-line textual communication, typing is representing the otherwise uttered language; 3) In human-computer (simulated) interaction, the pieces of text appearing on the screen are representations, giving an idea of 'as if' 2).

An 'other' usually assumes humanness, even identity (cf. Bruce 1996 on the importance of face in interaction). However, in present day Western society, there are numerous encounters between people in which the 'other' is not seen, and his or her identity remains unknown (e.g. switchboard workers). In these cases, the 'other' is considered as a representative of, for example, secretaries, though even in occasional meetings the 'other' can be quickly judged to be friendly, hostile, or having other personal traits. Thus, the 'other' can be treated as personifying a certain role.

If a human being is taken as a point of comparison, in my data from the human-computer interactions the voice mediated 'other' of the hypothetical telephone answering system becomes closest to a 'real other': a human voice interacts with the caller. In TUTORIAL the program is still fairly interactive, but the transient voice is replaced by text, thus making the system rely more on a representation of an 'other'. In TASK the 'other' is most distant as interactivity is diminished into a few interactive messages from the program in which most options are menus which the user manipulates, and the manual does not do anything on its own accord. The last case, TEACHING, goes back to interaction between temporally copresent 'others'.

Hanks' treatment of indexicality and deixis also describes the premises for copresent activities. For instance presentatives (e.g. *here* in 'Here, take this') can be used and understood if the participants have immediate sensory access, they do not require shared background knowledge of a referent as such (Hanks 1992, 68). In pairwork tasks at a computer terminal, a lot of pointing and other gestures are used to share the object of attention. This sort of sharing is qualitatively different from understanding and sharing what is meant by the entity being pointed at.

As discussed in Section 1.1.2, sharing as a basis for interaction depends on what type of intersubjectivity is at issue and what is assumed from the interactants: people can share goals, procedures and/or background knowledge; the communication can be about intentions or information.

In all the case studies, the intents of the 'others' and the recipients do not always match. In this way, the goals of the human participants, i.e. the recipients, are given prominence, and interacting with the 'other' is communication about information rather than intention. The users seem to orient to the programmed 'other' as someone or something, the abilities of which are limited: the human participants regard what they hear on the phone or see on the screen as shared, both in regard to background knowledge and communicative procedures. The role of the 'other' is that of a stranger who has to be approached with interactive caution. The distance in time from the 'other' might not only be restrictive but give more interpretive freedom (cf. telephone discussion versus use of manual). In the following four chapters, the interactive constitution of the 'other' and the ongoing task are addressed.

4. TELEPHONE: THE USE OF COMMUNICATIVE RESOURCES IN ENCOUNTERS WITH AN EXPERIMENTAL TELEPHONE DIALOGUE SYSTEM

This chapter builds on a study in which the findings and methods of conversation analysis were used to test how possibilities for local repair could enhance the use of computer systems (Raudaskoski 1990). The aim was to challenge the artificial intelligence work on modelling the cognitive processes of a user as the only way to resolve problems in human-computer interaction. The report shows how difficult it is to make predictions about the users' behaviour, even if results from research into real life human-human conversations have been used as the basis for design. Most importantly, this chapter reveals the users' methods of managing an encounter with a system that requires 'talk time' contribution from callers: each 'turn' by the system is a request for the user to say something. In exploring the use of communicative resources, the experimental nature of the data is not hidden. Instead, the impact of, for instance, an instruction sheet on the ongoing interaction is made explicit.

For the study, a hypothetical telephone answering system program was built, hypothetical because the speech recognition and production part of the system had to be simulated as the research into speech recognition is still confronted with great obstacles²². Of course the designed interface did not do what it was claimed to be able to do, viz. send messages to some persons or store and play on request messages sent to it, and the study was designed so that it was unnecessary.

The system produced repair initiators randomly from a selection of five possibilities in case the user's turn did not match what was expected in that phase of the program. The choices were:

- (1) Specify the source of trouble: "Leave what?"
- (2) Imply not hearing or understanding the user: "Sorry?"

²² In this sense, the requirement of interactivity as worded by Lombard and Ditton, for example, was fulfilled: "The larger the vocabulary of a computer speech recognition system [- -] the more interactive is the computer use experience" (Lombard & Ditton 1997, 25).

- (3) Guess what the user is trying to do (on the basis of synonymous expressions): "Do you mean leave a message?"
- (4) Give specific choices to the user: "You can leave a message or listen to messages - select either."
- (5) Precheck everything is alright: "So you want to - - - is that right?"

Out of these formats, 1 and 2 are clearly other-initiated self-repairs, i.e. prompting the user to repeat or rephrase her utterance. Although number 4 explicitly states what the user can do within the system, the format does not tie the utterance only to the previous turn, but it could function as a third turn repair-initiator as well (which often begins with *No, I mean...*). Also, because the utterance can be of no relevance to the user's turn, it is the most 'uncooperative' of the repair initiators. Both number 3, *Do you mean leave a message*, and number 5, *So you want to - - - is that right?* would not be regarded as repair initiators in traditional conversation analytical terms, as they are understanding checks, even if Frohlich and Luff regard *OK?* after an explanation as leading to an other-initiated self-repair (Frohlich & Luff 1990, 208). However, a paraphrase of the user's talk could be regarded as 'preventive' repair, as a form of uncertainty on the system's part which gives the user the last decision and choice, because "unequivocal displays of understanding can be achieved by producing a transformation or paraphrase of some prior utterance" (Heritage & Watson 1979, 129). Through repetition and tag questions, the system checks whether the user's intent has been correctly mapped from the information the system got through the keyboard, and whether misunderstanding has been avoided. Rather than describing the format as something leading to an other-initiated self-repair, I have suggested elsewhere (Raudaskoski 1992) that it could be understood as a special form of (potential) self-initiated other-repair, because the initiator is explicitly orienting to her understanding as a possible trouble source. According to Schifffrin (1990) and Schegloff (1992), intersubjectivity is a locally managed phenomenon: not being able to understand or error occurs only if participants in the conversation somehow make them relevant, especially through repair work. If intersubjectivity is achieved locally, then prechecks should be seen as fitting for that work, making sure that repair work is not needed; they operate towards future prevention, not on past repair.

The following repair initiators were allotted to the user:

- (1) Specify the source of trouble: "Do what?"
- (2) Imply mishearing: "Sorry?", "What was that?", etc.

The system was tested by three secretaries, who had never been in this kind of situation before. However, they were, because of their work, used to phoning and delivering messages or finding out about information for their bosses. Thus the situation was familiar otherwise; it was just that they were meant to be speaking to a system instead of a human being.

The participants were given a sheet of instructions (Appendix 4-1) about what they were anticipated to do: they were to send some messages and find out if there were any messages stored in the system. The participants were asked to do the following:

- I Find out if Fred has sent any messages
- II Find out if Eve has sent any messages
- III You want to get a message through to Eve, saying that you are at home
- IV You want Fred to know that you are going to a meeting at 7 pm

That fact that the system involved speech-exchange and not textual input was important for the analysis, as it shows the pressures there are in spoken interaction as opposed to those in a keyboard situation. For the transcription, both the automatically stored scripts of the sessions at the computer and the recordings of the telephone conversations were used.

4.1. Channelling the user through spoken dialogue

The experiment was preceded by preliminary testing, in the first of which only the keyboard mode was used (i.e. a student typed his turns in), and the second served as an experiment with the telephone system before the actual data gathering. The keyboard system relied on visual resources, for instance text and other uses of the screen. The spoken mode on the telephone resembled much more an ordinary human-human telephone conversation situation. However, even in this system, the user could hear the communicator type in what she said. After the data gathering all the secretaries said that the sound of the keyboard was not disturbing. The transcripts show how the user's input was modified and sometimes changed to make the typing as fast as possible and not disturb the *chronos* time of the dialogue. Maybe this is the reason why the following scenario did not happen (and as the sound of typing filled in any pauses in talk): "Small silences in a telephone exchange are read by both parties and may cause them to check both the talk and the technology ("Hello, are you there?")" (Boden & Molotch 1994, 264).

There was some overlapping of speech in the spoken mode, which occurred when the system had a longer turn that ended with a request for confirmation or a yes/no question. Overlapping often happens in speech with tag questions. For example, in 4(1) the user's *no* (line 2) does not overlap with anything, but occurs during a brief pause within the system's turn. (See Appendix 1-1 for the general transcription system and the special features of TELEPHONE.)

4(1)

- 1 S: 'So you want to leave a message for alison'
- 2 U2: <<no>>
- 3 S: '- is that right?'
- 4 U2: no I want to leave a message for eve

Another major difference between the spoken and written simulation was that speakers tended to expand their turn when a single *yes* or *no* was expected (the program allowed for expansion after 'no'); they would resort to everyday use of language on the phone.

4(2)

- 1 S: 'So you want to leave a message for <ID LIKE TO
- 2 LEAVE A MESSAGE FOR> fred - is that
- 3 [right?']
- 4 U2: [yes] <<yes i want to leave a message for fred>>

Out of 35 yes-no questions, 23 were expanded²³; the remaining 12 were answered with a single *yes*, *no*, or corresponding expression (*that's right* etc.). This lends evidence to the results of Hauptmann and Rudnicky's study (1988) where they contrasted different modes of interaction and came to the conclusion that spoken utterances tend to be longer than typed ones. Also, Severinson Eklundh noticed the same phenomenon in radiophone talk: "Many feedback answers serve to show the listener's *support* for the message in the form of an agreement; such responses also take varying forms from words like *Visst* ("Right"), *Just det* ("That's right") to longer phrases restating the message or giving a new version of it." (Severinson Eklundh 1986, 38). However, one subject almost always produced single word answers. Interactional reasons for different expansions will be discussed in the next section.

So one of the purposes of the system's repair initiators was to narrow down the users so that they gave only the expected information (and gave it in elliptical sentences). Examples show that channelling the verbose humans to give a one word answer is difficult, even with yes/no questions. Extract 4(3) serves as an example of how the user's answer to *A message from who please?* was never answered as briefly as was expected: usually it was answered without ellipsis. In comparison to the question, only *please* was left out, contrary to what was expected in the original design.

²³ Note how the user in Extract 4(2), line 4 treats the overlapping of her and the system's speech as people do in face-to-face discussion: the overlapped item (*yes*) is repeated when the user is the only speaker.

4(3)

- 1 S: 'A message from who please?'
2 U2: <<er>> a message from >>ALI<< <<alison>>

4.2. Repetition and reformulation

Examples from 4(1) to 4(3) above show that the users were resorting to repeating words or utterances spoken by the system. In 4(1) and 4(2), instead of answering simply *no* to the system's check, the user reproduced the whole utterance, and in 4(3), the system's question was repeated in its totality. Heritage and Watson (1979) treat repetition as one kind of formulation, a technical term for "demonstrating that, among other things, the conversation has been and is ongoingly self-explicating" (ibid., 123). The system's *So you want to --- - is that right?* was designed to overtly check whether the system had 'understood' the user correctly. Though the *is that right?* tag question was designed to channel the user to give a yes/no answer, the human participants tended to be more verbose. A closer look at the interactional sequence reveals that different formations by the human participant can be explained on the basis of the interactional context:

4(4)

- 1 S: 'Hello – what do you want to do?'
2 U1: hello has it that's kerttu has <<has>> fred sent any messages
3 [hello]
4 S: ['A mes]sage from who please?'
5 U1: has fred sent any messages?
6 S: 'So you want to listen to a message from <HELLO HAS> fred
7 <SENT ANY MESSAGES> – is that right?'
8 U1: y<<es can I listen to any messages from fred please>>
9 S: 'Sorry no messages. is this all?'
10 U1: >>NO<< has eve sent any messages
11 S: 'Listen to a message from who?'
12 U1: eve
13 S: 'So you want to listen to a message from eve – is that right?'
14 U1: y<<es please>>

The user's formulation of her first question in line 2 parallels direction I given to her in the instructions: *Find out if Fred has sent any messages*. The participants were well aware that despite the human voice on the telephone they would be interacting with a computer system. The decision to have a mediator between the (computer) system and the user meant that the telephone dialogues were not designed to 'test' the attitudes towards or trustworthiness of the system (cf. Wooffitt et al. 1997), but they concentrated on finding out how users managed a dialogue with what they knew to be a system, but with a human

mediator. (On the difficulty of being a mediator in real telephone mediated expert systems, see Whalen and Vinkhuyzen (forthcoming).) The beginning of the instruction sheet reminds them of the fact as well:

This experiment is being carried out to test a hypothetical telephone message storing and sending system: you can call the system which is attached to your phone to send messages or check if there are messages for you from other people's systems.

As there is no speech recognition/production system attached to the program, I will be simulating it on the phone.

It is an understandable strategy to assume that lexicon or phrases used by the system would be ones that the system would understand (cf. *hello* in line 2 as a second pair part in response to the system's *hello*). Also, orienting to the directives in the instruction sheet, shows that U1 wants to 'play safe' and aim at using the available materials as resource for interaction with the system (lines 2, 5, and 10 with the instruction sheet's formulation *has X sent any messages?*). The abundant repetition explains reciting the system's yes/no check in line 8: *listen to* appeared in the system's previous turn (line 6) for the first time (it was not used in the instruction sheet, nor had it appeared in the system's turns so far). Later on (line 13), the user does not repeat *listen to* after a similar question; the lexical item has been used in the previous check (line 11), and therefore the formulation in line 13 no longer needs confirmation (the adjacency pair in lines 11 and 12 already confirmed that it is a question of listening to a message and the check up in line 13 then concerns the person rather than the action). Later on in the same call, the user uses repetition again (the following extract continues from where Extract 4(4) ended):

4(5)

- 15 S: 'Sorry no messages from eve but there are messages from the
16 following persons: Emma, Evelyn – do you perhaps mean one
17 of them?'
- 18 U1: no <<fred or eve>>
19 S: 'OK. Is this all?'
- 20 U1: no
21 S: <WHAT ELSE DO YOU WANT TO DO?>
22 U1: i<<d>> like to <<leave>> a message >>TO<< <<for>> eve
23 <<– can you say that judy is at home – a message for eve>>
24 S: 'You can leave messages or listen to messages – select either'
25 U1: just to leave a message >>TO<< <<for>> eve <<and I'd like to
26 leave a message for fred>>
27 S: 'So you want to leave a message for eve – is that right?'
- 28 U1: <<a>> message for eve that's correct
29 S: 'What is it that you want to do then?'
- 30 U1: <<can you say that>> judy is at home
31 S: 'Do what?'
- 32 U1: judy is at home
33 S: 'You can leave messages or listen to messages – select either'
34 U1: I want to leave a message for fred

- 35 S: 'So you want to leave a message for fred — is that
 36 [right?]
 37 U1: >>Y<< <<for fr]ed that's correct>>
 38 S: 'Please give the message.'

The user orients next to the rest of the two tasks (III and IV above) in the instruction sheet (from line 22 onwards). The system does not 'understand' the first attempt by the user (lines 22 to 23), and the user's slight distress can be detected in the use of *just* (line 25) to start the turn after the system's announcement of the possibilities for use (line 24). After both of the check ups by the system (lines 27 and 35 to 36), the user orients not so much to the content of the whole utterance (leaving a message for somebody) than to the person the message is meant to be given, therefore repeating what to her is the gist of the check up: the person's identity. This could be seen from the two other interactions, as well (cf. Examples 4(1), 4(2) and 4(3) with U2). The verb *know* is used by U2 in the following, echoing the verb choice in direction number IV (*You want Fred to know that you are going to a meeting at 7 pm*):

4(6)

- 1 S: 'Sorry no messages. Is this all?'
 2 U2: <<er>> I want to know if there are any messages from eve

One of the users never returned the *hello* of the system at the beginning of the interaction:

4(7)

- 1 S: 'Hello — what do you want to do?'
 2 U3: has fred left any messages for jack
 3 S: 'So you want to leave a message for jack — is that right?'
 4 U3: fred should have left a message for jackie
 5 S: 'What is it that you want to do then?'
 6 U3: check <IF> <<that>> fred has left a message

The user's turn in line 2 is echoing the direction sheet, except that she uses *left* instead of the *sent* formulation in the instruction sheet²⁴. By giving the name of the recipient of the message, U3 in line 2 (as does U2 in Extract 4(20) below) the orientation to the system as a multi-user domain which needs not only information about who left or is going to be given a message, but also who the caller is (phoning for). This results in misunderstanding on the system's (pattern matching) part (line 3) and the user repeats her request, changing it from a question into a statement, using *should have left* which does not appear in the instruction sheet. When the system apparently does not understand the intent of the user (question in line 5), she resorts to the available materials and

²⁴ The system failed once with this user and was started from the beginning. In that response to the system's first turn, the user this time understandably did not return the greeting, but went on in her task: *please send a message to eve to say that jackie is at home*.

in line 6 uses the verb *check* (the first sentence of the instruction sheet, repeated above, was: *This experiment is being carried out to test a hypothetical telephone message storing and sending system: you can call the system which is attached to your phone to send messages or check if there are messages for you from other people's systems*). Interestingly, this user repeats the same 'mistakes' again in what follows from Example 4(7):

4(8)

- 7 S: 'A message from who?'
 8 U3: message from fred for jackie
 9 S: 'So you want to listen to a message from <MESSAGE FROM>
 10 fred for jackie — is that right?'
 11 U3: that's >>RIGHT<< <<correct>>
 12 S: 'Sorry no messages. Is this all?'
 13 U3: no has eve left any messages for jackie
 14 S: 'So you want to leave a message for jackie — is that right?'
 15 U3: eve should have left a message for jackie
 16 S: 'What is it what you want to do then?'
 17 U3: check that there is a message from eve for jackie
 18 S: 'So you want to listen to a message from eve <FOR JACKIE> is
 19 that right?'
 20 U3: that's >>RIGHT<< <<correct>>

Note that the system in a way confirms to the mentioning of the originator of the message by repeating it (automatically) in lines 9 to 10. Lines 13 to 17 are an almost exact copy of the lines 2 to 6 in Example 4(7). The strategy of this user, then, was not only to repeat the available linguistic resources, but also to repeat her (in the end) successful procedure with the system. So she assumes shared procedures with the system, but only on the basis of her own interactional experience with the same self system.

4.3. Use of pronouns

If repeating already occurred linguistic material was used as a strategy to interact with the system, then every utterance that was not available to the user as a resource becomes interesting as they can be treated as an indirect account by the user of the system's abilities. The instruction sheet warned the participants about use of pronouns: "The system does not understand pronouns like me, myself, you, he etc. so you should try and avoid using them!" This caution was meant to avoid situations in which a message is left for *him* or *her*, instead of using proper names. There are points at the interaction in which the user clearly orients to this rule:

4(9)

- 1 S: 'So you want to listen to a message from Eve — is

- 2 [that right?
3 U2: <<has she>>] has eve sent me a message

In line 3, the user changes the pronoun into the name. Maybe somewhat confusingly, the program would refer to the user as *you*, and reciprocally the user always referred to herself as *I*. This was done according to the change of indexical origo; the *you* and *I* of conversational coparticipants oscillates between the two as in Extracts 4(1), 4(2) and 4(5) above. However, the system never referred to itself as *I* (this was a conscious choice; the researcher did not want to anthropomorphise the system)²⁵. One of the users broke this symmetry by referring to the system as *you* twice (Extract 4(5), lines 23 and 30: *can you say that judy is at home*). Also in the Extract 4(9) above, the user in line 3 is referring to herself as *me*, which was the only time this usage occurred. As discussed in Section 2.3, Wiley's categorisation of 'self' could be adapted to the 'other' who in a dialogic relation is referred to as *you*; you-present-sign, you-future-interpretant, and you-past-object. My hypothesis was that the dialogic would emerge in interaction, which can be detected in these examples: *can you say that judy is at home* is being addressed to the present 'other', the system, and the formulation grew out of the immediate past interaction with the 'other' (e.g. in 4(5), line 30 as an answer to the system's question), and with the future 'other' in mind. (*Can you* portrays uncertainty about the system's abilities, an observation that seems to get support from other experiments: in Wooffitt et al. (1997), out of the data extracts given as examples, from the users' calls to the 'Wizard of Oz', many seem to employ the 'less polite' *can you* vs. *X could you*).

4.3.1. Avoidance of use of pronouns

One of the participants made a clear effort to avoid the use of pronouns (though even she used *I* to refer reciprocally to herself after the system's turns). An extract from the dialogue:

4(10)

- 1 S: '**What else do you want to do?**'
2 U2: <<uh>> **I want to find out if eve has sent any messages**
3 S: '**A message from who please?**'
4 U2: <<er>> **a message from >>ALI<< alison**

²⁵ Cf. the SUNDIAL data, collected by a 'Wizard of Oz' method, in which a human simulator did the 'role' of a computer, and in which the use of 'I' by the 'system' was not questioned or addressed at all (Wooffitt et al. 1997).

5 S: 'So you want to listen to a message from <A MESSAGE
 6 FROM> ali – is that right?'
 7 U2: no I want to leave a message for eve from >>ALI<< alison
 8 S: 'What is it that you want to do then?'
 9 U2: I want <<er I want>> to leave a message from alison <<to tell
 10 eve [that I'm]
 11 S: ['So] you want to leave a message for alison
 12 U2: <<no>>
 13 S: – is that right?'
 14 U2: no I want to leave a message for eve

In line 2 the user hesitates at the beginning (*uh*), and continues by using linguistic resources available to her in the previous turn (*you want to*) and in the instruction sheet (II *Find out if Eve has sent any messages*). The system's turn in line 3 is meant to work as a next turn repair initiator checking the identity of the message's sender. However, the user interprets this turn as 'who are you?', resulting in the system's wrong analysis (lines 5 and 6), and it takes more redirecting (line 8) and checking (line 11) from the system before the user gets her formulation into an acceptable (or parsable) form which is without the name of the user. Yet the user understandably repeated her name within her turns as she took the system's query in line 3 to be about her identity, and not a question about who the message leaver was.

4.4. The user and sequential meaning making

The users never applied the repair mechanisms available to them, though it was mentioned in the instructions that this was possible. However, the instructions were given without exact formulations: *It is possible for you to indicate mishearing and to some extent misunderstanding*. Maybe due to the vague description the users never resorted to the repair strategies; a general advice is harder to adhere to, unlike the specific one about not using pronouns (see Appendix 4-1).

As the system's interpretation of the user's turns was done via simple pattern matching, repair was initiated often 'needlessly' (cf. 4(10), line 8). But instead of challenging the repair initiation with counter repair, the user follows the system's agenda (though in 4(10), line 9, the repetition and hesitation at the beginning of the turn indicate problems with the system's previous turns). This may be due to the fact that users are very tolerant when they know that it is a computer system they are talking to. Therefore, they do not consider anything that the system may do as striking enough for them to show that they did not understand, or that the system's turn is not appropriate to the context. It also serves to show how flexible human beings are in making sense of a situation. They interpret the system's turns on the spot, sometimes giving a new meaning to a turn, quite different from that intended when the system was designed.

4(11)

- 1 S: 'So you want to leave a message for eve - is that right?'
- 2 U1: <<a>> message for eve that's correct
- 3 S: 'What is it that you want to do then?'
- 4 U1: <<can you say that>> judy is at home
- 5 S: 'Do what?'
- 6 U1: judy is at home
- 7 S: 'You can leave messages or listen to messages – select either'
- 8 U1: i want to leave a message for fred

In the excerpt shown in 4(11), the user's first turn was regarded as a negation (as it was not in an acceptable form for the parser to be a *yes*) and the system went on to ask what the user wanted to do. The user, oriented towards the dialogue on the basis of what had happened and what she had said, gave the system's turn a new interpretation: that the system was asking her for the content of her message. Even the system's next turn, which was meant as a repair initiator indicating that the system did not understand what the user wanted to do, was reinterpreted as a request to clarify the message and the user obediently repeated the message. The following system's turn was a repair initiator again, but the user's interpretation was that the system had continued after sending a message to Eve, repeating the possible actions that can be performed by the system. The user then went onto her next task of giving a message to Fred.

This imbalance between the user's and the system's abilities to make sense resulted, thanks to the user's work, in smooth communication between a user and a computer, but this had dangers. In 4(11), the user thought that her task was done while the computer was still trying to elicit an 'understandable' turn from her. The task was never completed, nor was coconstruction of misunderstanding.

Although a telephone dialogue system does not have the visual richness of a keyboard system which affects the interpretation and understanding process, this study showed that even if a computer system can dismiss the visual features of interaction, there is still lots to explore in the details of telephone conversations. Particularly in order to get results that would help designers to understand and build better computer systems for people to use. In this study, the results achieved by conversation analytic research so far were used with the assumption that these would ease the dialogue. This was the case to some extent²⁶, but definitely more attention has to be paid not just to the possible concurrent features and repeated patterns of talk, but also on

²⁶ A student of artificial intelligence who first tested the system on the phone was surprised how simple the implementation was in comparison to the level of interactivity (cf. Nass et al. 1994 on the social agency afforded to (simple) computer systems).

the actual minutiae of talk-in-interaction, whether on the phone, at a computer, or wherever more than one is involved.

A computer dialogue system usually tries to channel the user to give 'correct' answers, replies that the system is able to parse. One lesson from the design of the hypothetical telephone dialogue system is that the system's turn should not include two types of speech acts. However, already the first turn by the system included two first pair parts, a greeting and a question (*Hello – what do you want to do?*). Both greetings and questions as a first pair part of an adjacency pair normally require the production of a second pair part :

The most sequentially constraining of all utterances are the first pair parts of what Sacks and Schegloff term "adjacency pairs," where the production of a first pair part requires that a next speaker produce (1) adjacently, (2) a second pair part, (3) of the pair type previously selected by a first speaker. Thus greetings require returns, questions require answers, and offers require acceptances or rejections (or, more generally, decisions).

(Heritage & Watson 1979, 141)

Thus in human-human interaction there is a strong tendency for greetings to be returned, and this orientation was also shown in my data (cf. Example 4(4); another user replied *hello it's alison here*). Also the repeated occurrence of check-ups by the system (*So you want to X – is that right?*) and the similarity of the formulation of the closing up initiator (*Is this all?*) with *is that right?*, led into the following user formulations:

4(12)

- 1 S: 'Allright – the message - - - is to be given to fred. Is this all?'
- 2 U2: **that's correct**

In 4(12), the user's turn shows that she is interpreting the question to concern the first utterance by the system. *Is this all?* could of course be seen to question whether the content of the message to be given is enough, rather than asking whether the user wants to continue with the system, showing another example of the difficulty in predesigning one sense for an utterance.

4(13)

- 1 S: 'Allright – the message jackie is at home is to be given to
- 2 eve. Is this all?'
- 3 U3: **that's correct. I have another message**

Again, the user evaluates the system's informative turn, also giving an answer to the preclosing question.

4(14)

- 1 S: 'Allright – the message - - - is to be given to fred. Is this
- 2 [all?]
- 3 U3: [**>>Y<< <<that's correct] that's all>>**

4 S: 'Bye bye then'

4(15)

1 S: 'Allright – the message - - - is to be given to fred. Is

2 [this all?]

3 U1: [<<that's right>>] y<<es>>

4 S: 'Bye bye then'

In Examples 4(14) and 4(15), the user is overlapping with the system's *Is this all?*, clearly orienting to it as a tag question, and also giving an answer to a tag about the content of the first sentence (*that's right, that's correct*). However, when the users hear *all*, they have to give another answer to the new initiation, *Is this all* (*yes* and *that's all*).

Therefore the designers of systems in which the user should be channelled to produce one speech act at a time, should take into account the dangers of having more than one speech act in the system's turns.

4.4.1. Contesting/yielding to the 'other'

Production in the cycle production-distribution-consumption of text is often connected with literature or other printed media which are material language objects usually read in isolation; they are not used in connection with practical activities. Although reading in itself can be regarded as social action, the language is not produced, but consumed, and is not that of speech-in-action. However, in computer surroundings, it is possible to research the interplay between production and social interaction (cf. Thibault 1991); how language products are made use of and understood in the everyday surroundings, and how the predesigned meanings of these language products can be changed or contested *in situ*. Though the three secretaries in the trial phone calls could orient to the 'other' as a system which can be assumed to understand the linguistic items it produces, there were moments of mockery – which of course could be caused because of the experimental nature of the enterprise:

4(16)

1 S: 'So you want to listen to a message from Eve – is that right?'

2 U2: yes <<a message from eve>>

3 S: 'Sorry no messages from Eve but there are messages from the
4 following persons: Emma, Evelyn. Do you perhaps mean one
5 of them?'

6 U2: 'yes <<I will listen to a message from emmaly>>

7 S: <WHICH ONE?>

8 U2: >>FROM EVELY<<

9 S: 'Sorry no messages'

10 U2: <<no messages from emmalee gosh>>

For the participant U2 (as for the system), the persons called Emma and Evelyn are of even less interest than Fred or Eve as the two female names come up for the first time in the whole telephone call setting; they were not mentioned in the instructions. This is reflected in the way the user combines the names of the two into another hypothetical person. When the system then tells the user, contrary to the information in lines 3 to 5, that there are no messages, the user's reaction in line 10 exaggerates her disappointment. The use of colloquial *gosh* makes it clearly not orderly in the communicative context, and maybe it is addressed to the mediator rather than to the system. Extract 4(16) continues in Extract 4(17) below, in which the user continues challenging the system:

4(17)

- 1 S: 'Is this all?'
- 2 U2: **that's all**
- 3 S: 'What else do you want to do?'
- 4 U2: <<er>> **are there any messages from anyone <<he he:>>**
- 5 S: 'So you want to listen to a message from anyone – is
- 6 [that right?]
- 7 U2: [from anyone] <<anyone>> called fred or eve >>A<< <<or>>
- 8 **emely**
- 9 S: 'What is it that you want to do then?'
- 10 U2: <<er>> **I want to know if alison has received any messages**
- 11 **from fred eve or emely**

As the system was not designed to understand *that's all* as answering positively (*yes*) to the question in line 1, the system continues the predesigned turn-taking by asking the question in line 3. The user's hesitation (*er* in line 4) and the following question in which *anyone* instead of a name was used to ask once more the question the user had not got an answer to so far. Not only did the participant use a forbidden pronoun but she also laughs at the end of the turn. The laughter particles make it clear that the user orients to her question as something not wanted or out of order. When the system seems to 'understand' the question (unlike with several attempts to use names), the user hastens to clarify it by giving a list of names, thus producing a turn with the resources had appeared in the past (lines 7 and 8). However, the formulation that was designed to meet with the restrictions of the system fails, resulting in an answer predesigned to a *no* (or, rather, anything else but *yes* as *from anyone* in line 7 fails to fulfil a slot for *yes*). The user, after a short hesitation mark (*er*) in line 10, continues with a new trial, this time employing a new verb, *receive*.

Thus, taking up not previously used lexical items seems to be related to challenging the system if it is not 'interpreting' the user correctly; however, once the system seems to 'understand' the new item(s), the user gets back to the orderliness of the interaction. Extract 4(18) below continues from where Extract 4(17) ended:

4(18)

- 12 S: 'So you want to listen to a message from fred eve or
13 emely – is that right'
14 U2: that's correct
15 S: 'What is it that you want to do then?'
16 U2: are there any messages
17 S: 'Listen to a message from who?'
18 U2: from eve

Unfortunately the system was not programmed to parse *that's correct* as equivalent to *yes*, so the system regarded the user's answer to the check-up as *no*, and continued accordingly. The user then has to resort to another strategy to get her task done, and she uses a short formulation in which *are there* is utilised. It had already occurred before (Extract 4(16) line 4, the system uses these lexemes, and so does the user in Extract 4(17) line 4, with interactional success).

4.5. Summary

Many in our society have only recently experienced the shock that comes when one first realizes that the "person" who rang the phone is really a recording. But those among us who are already accustomed to recordings will readily acknowledge a certain process of acclimatization: as the collage of recorded and "real" voices becomes more and more intricate, and as the underlying programming responds more and more flexibly to our needs, we make less and less of a distinction between the various levels of genuineness. We are comfortable doing business with the words themselves.

(Talbot 1995, 221 – 222)

The quotation from Talbot repeats his idea of human-computer interaction being not so much of human-person, but human-words interaction. The three secretaries who tested the system used the same basic strategy which seems to be very 'word oriented': employ the lexemes or linguistic units available in the test situation. At the beginning of the phone call the pressure for identification was there: "Whatever a telephone conversation is going to be occupied with, however bureaucratic or intimate, routine or unusual, earthshaking or trivial, it and its parties will have to pass through the identification/recognition sieve as the first thing they do" (Schegloff 1979, 71). Identifying the communicator or the caller was also a way of bracketing the action ensuing, trying out the system. Through ongoing interaction with the system, the users gained knowledge of both vocabulary (either used by the system or successful trials by themselves) and the interactional procedures that would result in the task(s) getting done. Their expertise was interactionally constituted (cf. Goodwin 1986b, 292-293): the secretaries managed the task of giving a phone call to a hypothetical telephone answering system. Heavy reliance on already

existing linguistic resources led to a seemingly repetitive style in all the dialogues. However, a closer analysis of the interactional details of the dialogues revealed the users' interpretations of the system in the ongoing conversation. The repetitiveness was an outcome of managing the talk in appropriate time: the *kairos*, "the *now* whose time has come" (Erickson & Shultz 1982, 72) has to be negotiated on the basis of the available interactional resources, in *chronos* time. The overlappings of talk with the system happened in exactly the same positions as in human-human interaction: tag questions were produced to get a reaction from the other participant in the first place, therefore simultaneous speech with them was not considered rude.

The skilful management of the telephone calls also points at learning, which "can be viewed as a *feature* of practice, which might be present in all sorts of activities, not just in clear cases of training and apprenticeship", as described by Hanks (1991, 18). The users learnt the interactional scope of the 'other' by carefully expanding the repertoire of their talk on the basis of the immediate history of the phone call. As the acquisition of new interactive skills, growth of communicative experience, another layer of interpretant, can take place in any communicative environment, the fact that learning is taking place did not mean that the users would treat the 'other' as a teacher, expert, or other entity with more knowledge and skills. However, resorting to repetitions and testing out the interactive ground, they did acknowledge the 'other' as asymmetrical in the dialogic skills, and as somebody or something they have to adjust to and not vice versa. The verbose agreements to the system's checks (or self-initiated other-repairs) emphasised the system's correct interpretation as a mutual achievement. Even if the 'other' started the exchange with a transient (voiced), open question format *What do you want to do?*, the users were restricted by their task sheet to certain answers. In a way, the instruction sheet worked even as a more restricted option (as the questions are numbered) than, for instance, the written options provided by a visual computer conferencing system "What do you want to do? (Read) next letter, (Go to) next conference..." (Severinson Eklundh 1986, 22). The users in TELEPHONE attempted many times to reformulate their request into an acceptable form. By resorting to known skills (e.g. repeating words and phrases available in the situation) from the very beginning, the users exhibited uncertainty about the 'other's' abilities. Intersubjectivity at the level of the code was assumed, but the interaction consisted of repeated attempts at filling in the 'right words' by the user. She had to build the 'other's' interpretant because of the limits of the 'other': instead of coconstruction of meaning, these dialogues were examples of user-construction of meaning. The system's turns were treated only as signs or objects that the system puts out in reaction to what the user said, not as interpretants that the system would be responsible for.

The participation role of the 'other' that the dialogues analysed above hint at is an interlocutor rather than an intermediary, because the

intermediary would not be granted any knowledge of the subject, i.e. they would not be engaged in conversation. However, the users seemed to be interacting with (mediated) words in the sense that they make use of the very same words as the system to get their tasks done. This is also an indication of accommodating to the 'other's' genre, which is typical in institutional encounters; the topic and how it is talked about (e.g. by asking questions) is controlled by one of the interlocutors. Marková calls these features, actually produced in dialogue, endogenous asymmetries (1994, 333). In TELEPHONE each of the system's 'turns' was designed to prompt an answer, i.e. the system was always initiating a first pair part of an adjacency pair. Although the users were trying to make the system do things for them, the system 'took over' the dialogue by not only initiating questions and directives, but also by initiating repair. It is in the act of contesting the interpretation of the other participant by repair initiation that intersubjectivity is exhibited (Schegloff 1992). When the mediated 'other' repaired the caller, it was evident that what was being challenged was the user's wording of the instruction sheet, resulting in rephrasing action (or request), instead of undertaking another action. If the repair initiations were deemed unnecessary by the users, the evident state of no intersubjectivity brought about ridicule, which also reflected the user's inability to repair the system. Frustration was another result of not being able to achieve intersubjective meaning by negotiation and repair. Thus, repair work is not only a technique for maintaining intersubjectivity in interaction; unequal rights to the technique can result in strange constellations of meaning making. In TELEPHONE, the user was one-sidedly constructing the meaning of the encounter, but she had to do that by aligning to the system's repair initiations and other requests. The fact that the users' did not initiate repair (as sign of mishearing or – their own or the system's – misunderstanding) could be an indication that the experiment was not just measuring the 'intelligence' of the system: for the human participants, their abilities as secretaries to deal with a telephone message system was an issue as well.

The communicative resources in TELEPHONE were drastically different from communications in a written medium, such as the Advice System dialogue interface described in Frohlich and Luff (1990). In this system the users could compile sentences from a set of options, and sometimes they had to type in answers to the system's queries. In the Advice System, the possibility for the user to initiate repair by asking *What?* and *OK?* were visually depicted as buttons which could be selected at certain points of the interaction (and which also were used). Thus, the users were aware at each point of the interaction what communicative resources were available, and could reflect upon those as well as upon the system's turns. In Luff and Frohlich (1991) the consequences of the visual interface for the interaction are explained further. One clear difference between the telephony system of the present chapter and the Advice System was that in interacting with the latter the users sometimes confused the utterances designed as their potential

contributions to the ongoing dialogue as the system's turns. Thus, though visual interfaces allow for more reflection than talking to a system, the material manifestation is consequential to the interaction. In the case studies detailed in the following three chapters, special attention is paid to how encounters with language technologies are shaped by the visually available semiotic fields (Goodwin in press) that get activated.

5. TUTORIAL: (CAUSES OF) MISUNDERSTANDINGS IN TEXT- USER INTERACTION

In the previous chapter, the communicative situation involved a predesigned computer program ('telephone answering system'), a mediator between the program and the user, and three users who telephoned the system in experimental circumstances. The case study of the present chapter comes from a university setting: two native speakers of English agreed to rehearse their word processing skills with the help of a tutorial program. The data was gathered when DOS Word 5.0 was commonly used; the tutorial that came free with the word processing program was called Learning Word 5.0. As the language technology used now involves a visual interface and text, the problematics of meaning making through text become central.

5.1. Introduction

In the case of instructional texts such as printed or electronic manuals or educational documents, the author(s)'s intent and the reader(s)'s interpretation should overlap as much as possible for the best outcome. In researching causes for potential difficulties when conveying meanings textually, the focus has moved from texts as structures to their readers as active participants in the local meaning making process. Each text user has learnt the language of the text through various contacts with (similar) language or terminology, which means that the text users may have different personal intertextualities, the implications of which for their specific encounter with a text are very hard to detect. Also, each engagement with a text has an immediate history that can reveal causes for possible misunderstandings. In the following, a local abbreviation (i.e. not found outside the context of the text) is learnt and used in a word processor tutorial. However, at one point, the referent of the agreed convention changes to something else. A careful conversation analytical investigation of the local history of the encounter reveals the contingent reasons for the wrong interpretation: 1) the misleading semiotics of the visual details of the text on one screen, and 2) the general orientation of the users to the program as

logical in its performance. The technique of the chapter is to slowly reveal more and more of the context, to show that many aspects of it need to be known for a proper explanation of a phenomenon.

The social interactionist view that, for instance, M. Nystrand represents, gives the author of a text the credit of at least aiming at making the text as comprehensible as possible through elaborating an inherent "troublesource" (Nystrand 1992, 168). He gives an example of a text in which the writer and the reader meet in the textual space created by the author's active rewordings of potentially difficult terms, resulting in a growing circuit of Rommetveitian temporally shared social realities (TSSR) which equal "an expanded, shared understanding between writer and reader" (Nystrand 1992, 166).

In the case of nonfiction and especially so-called utility texts (e.g. Pilto & Rapakko 1995), it is easier for an outsider to grasp the users' interpretations, as these texts usually deal with 'exophoric' (situational) rather than 'endophoric' (textual) relations (Halliday and Hasan 1976). A set of directions can engage the reader in two sorts of activity: reading or acting upon reading. The difference between these two looks similar to what Coulter (1994) suggests: there is a distinction between understanding and interpreting a text, because understanding means that the reader is able to make the text intelligible, but interpreting is an activity in which the text is given a significance. If the difference is brought to the investigation of computerised textual environments, 'understanding' could equal the ability of the users to decipher the English language, and 'interpreting' the actions that they do on the basis of what they are reading, i.e. the practical outcome of their understanding. Also a parallel could be drawn between text linguistic research which concentrates on the text itself, the structure of which is the target of analysis ('understanding'), and interaction oriented research which treats meaning as situated and emerging ('interpreting') which cannot be determined *a priori*. Similarly in Schiffrin's (1990) view, 'understanding' (meaning) deals with the level of the sign; how the information or the actor's (author's) intent, is organised, whereas 'interpretation' is the receiver's intent.

The problem explored by several reading research projects could be summarised as "what it is that makes texts easy or difficult to understand" (Cook 1995, 9). With the advent of Information Societies, citizens all over the Western world increasingly also encounter electronic texts. In the following a case study is presented which illustrates the often preached but not so often practised analysis of readers encountering texts, or the fleeting moments of enunciation. This can inform the reader and interaction-based theories of the pitfalls of ignoring the material circumstances and the sequential developing of the 'communication'.

Acting in the world sometimes requires reading and understanding texts and other signs, and this work takes place in the spatio-temporally unfolding moments of interaction. When the process of interpreting is researched, the 'layers and layers of meaning' or other dangerously structural

sounding terms (Leppänen 1995, 200) can be shown to be in flux, not just among language users or in one person over time, but in one person within a short time span. Silverstein (1976), among others, stresses that language does not exist for primarily referring interactionally. However, the abbreviations which are potential troublesources could be described as metalanguage, as referring to the expansion formulated, and this connection can sometimes be disrupted due to the interactional particulars of each encounter with the term and the text.

The aim of this chapter is to give a concrete example of how misunderstandings can be detected and what their reasons can be. The method of investigation will be conversation analysis, which is a powerful tool for analysing (troubles in) meaning making, and which can be used to explain the semiotics of communication as well. Firstly the case study will be introduced, then the methods used, with the bulk of the chapter consisting of expanding the analysis from where a misunderstanding has occurred to the wider interactional context of the encounter.

5.2. The electronic text of the case study

The Learning Microsoft Word program (for Word 5.0) dating back to 1989 provided text-based help for users (though some elementary graphics and animation were sometimes used to illustrate a point). In its design, simple interactional and even corrective devices were implemented. These features and the whole structure of the program reflect the flow chart of computer aided instruction, CAI, (cf. Tiffin and Rajasingham 1995, 98) and the examples within brackets refer to the tutorial studied in this chapter.

1. Present a frame of knowledge (e.g. the *Entering Text* screen could show an *Overview* or a *Steps* option)
2. Test whether the knowledge can be applied (e.g. the *Practice* screen)
3. If the student response is correct, provide positive feedback (e.g. *Great!*) and go to the next frame, if not, give an explanation that relates knowledge to the problem (e.g. *Move the mouse pointer to..*)

Two fairly novice users explored the program. One (A) had been using Word 5.0, but not the mouse, nor was the participant acquainted with some special features of the program; the other (B) had used text processors before only a little, and had once been shown how Word 5.0 works. As there were two people using the program meant that meaning making was a cooperative effort, thus there was more to rely on in the analysis than just the script of what happened on the screen and the actions of one user (cf. Frohlich et al. 1994; Suchman 1987). This encounter was videotaped so that both the

participants' gaze direction, gestures and other activities were captured together with what was happening on the screen.

The intent of the authors of the learning program can safely be assumed to be the simulation of 'real' teaching of computer use. According to Witte (1992), the writer's intentions can be understood on the basis of a triangular relationship in which text, context and intertext replace the famous Peircean sign, object and interpretant relationship²⁷. In the present case, text would be the electronic tutorial text, context would be the unproblematised 'context of use' as depicted by the authors of the program, i.e. a computer setting, and intertext those occasions of text use that might be relevant for the present text (intertextuality). In the Learning Microsoft Word text this was most probably that of a manual (and a computer), and maybe an instructor near the computer user or in the front of a computer class. Also, a presumption of any teaching program is that whoever uses it is a learner, i.e. they need guidance. In the case of the Learning Microsoft Word, not only was the presupposition that the users might not have used (functions of) DOS-Word 5.0, but that they might not have used a mouse either. This happened to be the case in the present data: the users had to learn the context of educational language use (e.g. how to take instructions from a mediated 'other'), but also the material context of communication (e.g. abbreviations and other 'jargon' on the computer screen). However, in order to research the success of mediated instructions, a hypothesis of the program writers' intentions as (stable) aspects of a contextualised text does not give any idea about what readers' interpretations will be.

5.3. Adjusting interpretation: Repair work

As discussed in earlier chapters, if the stress in the analysis lies on the moment-for-moment meaning making, then conversation analysis (CA) as a method of analysing the actual semiosis is pertinent because it is able to reveal the sequential and local sense making practices. Also, adjusting meaning in talk, i.e. repair work, has been researched extensively in the CA research tradition (Schegloff et al. 1977, and many other works by Schegloff). In the present data, meaning adjusting is done between the human participants, but also by the program, and this is why it is important to remember the basics of repair, discussed in 2.2.1.

Repair work is closely related to the sequential nature of meaning making through turn-taking: every turn is at the same time an analysis of the previous turn and a new contribution. But the turn can also be coconstructed

²⁷ The turn of the century American semiologist Peirce sees the sign as a (never ending) loop, in which the three parts of a sign are interrelated. "A *sign* refers to something other than itself — the *object*, and is understood by somebody: that is, it has an effect in the mind of the user — the *interpretant*" (Fiske 1990, 42). (See 2.2.3.)

nonlinguistically: "Whether articulated through talk or body movement, the context is produced in and through a social organisation; an organisation which is both context sensitive and context-renewing. In and through this organisation, and their visual and vocal actions and activities, participants systematically shape and preserve the context at hand, repairing the indexicality of practical action 'in flight'" (Heath 1992, 122).

There seems to be an order of importance or preference in how people do repairing, and this order is closely connected with who starts the repair (self or other), and who is the target of the corrective work (self or other). Self-initiation is preferred over other-initiation, and the space for both types of initiation are within three turns (the present, at the transition relevance place and in the next turn) (Schegloff et al. 1977). In other-initiation the trouble source can be made explicit, and there is usually a short pause indicating a possibility for self-initiation thus making it visible that self-initiation (and correction) is preferred. Anything in the conversation can be a repairable.

5.4. The semiotic systems of the setting

The users of the actual tutorial could choose which module to do, and then go through it by reading about it and trying the functions out on a replica of a Word 5.0 screen. At the very beginning of the session (before the start of the actual tutorial), the users could select to do a separate mouse tutorial to learn how to use the mouse. In using the program, the human participants had to decide whether a screen was only for reading, for reading and acting upon reading (which could be bypassed), or having to act upon reading. As can be seen below, usually the last option, or rather obligation, accompanied an interactive message from the computer.

In Section 2.4, Figure 2-1 repeated below as Figure 5-1, depicted the talk between the users: the semiotic process that occurs between two competent interactants who share the time and the place of talk.

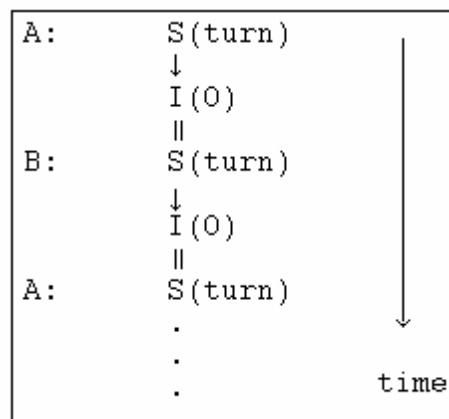


Figure 5-1

However, in the case study of the present chapter, talking, gesturing and other meaning making was undertaken by two participants side-by-side, 'ear-to-ear', in front of a computer. The program exhibited information and requests, and as mentioned earlier, the CAI (computer-aided instruction) characteristics of it meant that the user/learner's action was acceptable or not to the program. In addition to the overt acceptability of doing the practice correctly (managing the testing stage of CAI), a more covert interaction was also going on. Namely, the users' choice of *Next Screen* or *End Practice* at the lower right corner of the screen implied that they had read/seen the screen they were leaving: by pressing *Next Screen* or *End Practice* they were giving a metaphoric nod to the program, an *mm* or backchannel particle, to indicate understanding and to give a go ahead for more information, for a relevant next. In fact these two options for the users to choose in order for the 'tutor' to continue were also indicating what kind of screen was in front of them, whether they should read (*Next Screen*) or practise (*End Practice*). However, due to reasons that will be dealt with later, sometimes a problem arose when they could not see the difference of action space created for the user-reader; as a result, the users would try to do the action described in the text. This mistake would then result in a corrective message from the computer, an other-initiated other-repair (Schegloff et al. 1977), which behaved a bit differently from the corrective messages in the testing or practice phase. Unlike the 'turn-analysis' repair, the turn-taking repair (see e.g. Raudaskoski 1992) which resulted from the user's attempt to take the floor of action too early would disappear from the screen without any action required from the user. The consequences of having two sorts of repair initiators can be quite intricate; however, through careful data analysis they can be found to be reasons for some problematic understandings.

A screen called *Overview* depicting each stage before practice was static, i.e. there was no interaction with the user. In the case of actions that combined various stages such as inserting text by choosing the place with a mouse click and then entering the text, the overview was sometimes given in pieces called *Steps* screens if it concerned the practical mouse and keyboard clicks demanded. The stepwise instruction giving resulted in a confusion between reading from the *Overview* and *Steps* screens and doing in the *Practice* screen.

In the following diagram, the step-by-step appearance of instructions in the steps mode and the ensuing practice mode are schematised (*NS* standing for *Next Screen* and *EP* for *End Practice*):

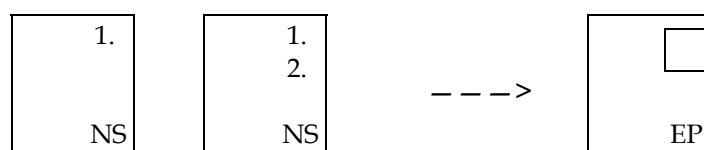


Figure 5-2

The practice mode combines the directions of the *Steps* screen(s), transforming a list of elements into one entity or screen to perform on. The possible subtasks are given at once: e.g. in the figure above the right hand box on the screen could have the following text: *Move the highlight to the d in due and type the word past.* <2 new lines> *Then press the Spacebar to add a space.* The program is in charge of the content of practice and gives feedback to the learner (e.g. *Good!*). The *Practice* screen emulates Word 5.0 but does not work fully like Word 5.0, e.g. the backspace key cannot be used at times.

The encounter between a text/author and a reader can be analysed as social interaction unfolding in time and space which, in addition to the local history of the encounter, also forms an integral context of understanding or misunderstanding. The text/computer can be seen as taking turns, but in the case of the users choosing to go forward, the turn is given to the text by its reader, and therefore s/he is in control of the encounter.²⁸ However, sometimes the program corrects the user in an interactive fashion, and in these moments the reader/user is meant to decipher the ensuing turn/action by him/her as part of a side sequence. The active role of the program in marking the problematic actions of the user makes it different from Suchman's (1987) photocopier, which relied much more on the user to find the trouble: the photocopier showed the user instructions of the relevant step on the basis of what the user was modelled to be doing. The display had a 'Help' button which was meant to be selected in case of problems.²⁹ Thus, even if the photocopier's instruction would have been a repairing one (e.g. repeating an earlier instruction because the user's action caused the program to backtrack to an earlier stage), the instruction was not marked as a repair initiator.

A careful analysis of the encounter below will show that difficulty in making the difference between commands to do and reformulated directions (repair) can be a cause of some problems in maintaining an interpretation.

The 'interaction' between the computer and the users could be sketched semiotically to be as follows:

²⁸ Actually, the system was programmed to initiate a 'Hint' or repair if the user did not do anything within one minute, allowing therefore for a 'slow motion' interaction pace. In the case study, this prompting never occurred as the users were active enough. In spoken conversation, the 'tolerance limit' for pauses apparently is one second (Jefferson 1989).

²⁹ The 'Hint' option of the Learning Microsoft Word program could be selected also by the user from behind the 'Course Controls' option — it was never adhered to, though, maybe because it was not directly available on the screen.

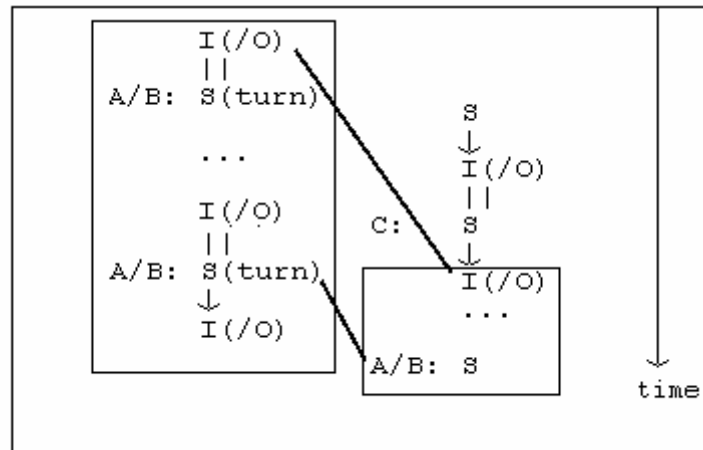


Figure 5-3

In Figure 5-3, the ongoing semiosis, or meaning making in a pairwork situation with a computer system, is given a pictorial representation. The left part of the diagram depicts the negotiation that goes on between the users (cf. Figure 5-1): it is possible for the users at each stage of the interaction either to do the relevant next action by themselves or after negotiating with the other participant. Thus, two parallel and intertwining coconstructions are at work: that between the reader(s) and the text, and that between the readers.³⁰ Even if they do go on without asking the other's opinion, the users normally give a verbal account to the other of how they understand what is going on. The arrow from sign (S) to the interpretant(/object) (I(/O)) depicts how each sign and therefore its object has to be interpreted, and this interpretation is 'externalised' in the next turn by the other participant (the two vertical lines picture an 'equals' symbol). The interactionist depiction also brings to the fore how close-knit the sign, interpretant and object are; not only the sign, but its communicative interpretant are material.

The right part of Figure 5-3 represents the interaction between the computer (text/author) and the users. As mentioned earlier, the users can select the next turn from the computer, or there can be an interactive corrective or praising comment from the program. In the instances when the reader(s) select to proceed and therefore request for a next piece of information/turn from the computer, the sign (S) is the mouse click on the lower right corner, resulting in the program's interpretation/object of showing more information. It can be preceded by negotiations and interpretations of elements of the text on the screen, and in these moments the text on the screen is usually made part of the ongoing interaction by the users through quotation.

³⁰ Cf.: "the organization of situated action is an emergent property of moment-by-moment interactions between actors, and between actors and the environments of their action" (Suchman 1987, 179).

The depiction of the semiosis in the case study concerns only an actualised encounter, and not the potential that the designers of the tutorial program had anticipated when they planned the program. The program and its user(s) are asymmetrical in their interactional abilities, both in the extent of the 'input' from the outside world (in the case of the computer, only mouse and keyboard clicks), and also in the capability of 'making sense' out of the incoming data (the program was predesigned to treat some actions by the user as correct, some not). Every screen the users encountered enabled them to do certain things. If anything different was done, there would be a corrective message. In this sense, what C produces in Figure 5-3 is preplanned and therefore extremely limited in comparison to what A and B can do in their turns.

5.4.1. *Visuality and materiality*

The written language that the users encountered on the screen formed a visual context that was surrounded by the icons, arrows, (moving) graphics and other visible surroundings of the program, as exemplified in Figure 5-4.



Figure 5-4

Thus, the interactive space was different from a text-only format prevalent in most fictional texts (also electronic ones; cf. Aarseth 1997). As discussed in 2.3, visibility and interpretation of visual phenomena have been studied by the visually oriented research traditions within semiotics (semiotics of the media, architecture, visual semiotics); however, because semiotics is traditionally a structural enterprise, the visual phenomena are often analysed in order to search for a structure, in the same way as the structure of language has been

of interest to linguists. "The third dimension: from reader to user" continues to be at a stage of "a first exploration" (Kress & van Leeuwen 1996, 242).

Petrilli's (1990) ideas of materiality and signs were discussed in 2.3. Her division into extrasign materiality (physical, instrumental) and semiotic materiality (ideological, extraintentional, signifying otherness, elaboration) maintained the distinction between instrumental and symbolic, contested by Engeström and Middleton (1996). To Petrilli, only nonverbal signs can have extrasign instrumental materiality, as verbal signs exist for meaning only. In technology-mediated communication environments, extrasign physical materiality refers to the substance, colour, shape, and other material features of the artefact. According to Petrilli, semiotic materiality proposes that no sign, be it a verbal or nonverbal one, has only one and stable meaning but that the complexities of the communicative situation influence semiosis.

Petrilli's and Witte's ideas about signifying as a contextualised and intertextual process belong to the same Bakhtinian heritage. The social semiotic point of view takes into account the material and visual semiotics of the communicative situation. When this approach is combined with the accurateness of conversation analysis to analyse ongoing interactions, the researcher has a convincing array of methodology to inspect semiosis at a computer tutorial.

5.5. Data analysis

In the present data, the extrasign physical materiality and semiotic materiality are both represented, the first as the physical and visual context for the latter to take place in. The ideological materiality comprised the context and intertext of teaching and learning how to use Microsoft DOS-Word 5.0, which took place in an experimental situation in university surroundings. The tutorial was given by a program, which the designers had programmed to take into account that the context of computer use might not be known to every user. Consequently, a potential learner could choose to use the interface to find out for instance how to use the mouse.³¹ Indeed, this option was available for the users at the very beginning of the learning program, preceding the actual tutorial (which started with the user giving their name to the computer). Most of the mouse tutorial concerned which buttons to click on the mouse and how to proceed to the next screen with the mouse. Thus local

³¹ It is noteworthy that, for instance, Microsoft Word 7.0 does not even have 'mouse' as an entry in the index for the Answer Wizard. In the *Help Options* searchable word list, mouse is not a potential problem, instead it occurs mostly in the phrase 'click the right mouse button', i.e. 'to click' invariably means clicking the left mouse button. The mouse has become a 'theme', a taken-for-granted, of text processor use; also, it is an endangered species which is being decentred by trackballs, pointing sticks and pen pads.

(i.e. not to be found outside the tutorial program) abbreviations were introduced to make referring to clicking the left or right button (or both) of the mouse more efficient. One of the most used ones was *Click-L* to refer to clicking the left button.

5.5.1. Click-L

In the mouse tutorial, *Click-L* "means to press and release the LEFT mouse button", or in the words of the summary:

When you see this:	Do This:
Click-L	press and release the Left mouse button

So, *Click-L* was a local hybrid with features of symbol, icon and index at the same time. It clearly was artificial and conventional (symbol), its design iconic in that the acronym was a symbolic metaphor of the activity (pressing and releasing: *Click*) and the object (the left mouse button: *L*). All in all, *Click-L* was an abbreviation typical of written language, a troublesource, a complex token whose value (Halliday 1994b) was given in the definition. When it was used without the definition, it was a two-morpheme construction, and composite structures "require relatively more constructive and/or interpretative work on the part of the language user so as to construe the new joint meaning which results from the combination of their constituent parts" (Thibault 1997, 283).

The temporally shared reality that *Click-L* and the reader were meant to achieve worked quite well in the tutorial, but at one point one of the users was clearly working with an incorrect notion of what *Click-L* refers to. B's interpretation is available for the analyst in the video recording, and in the transcript in line 114 which is marked by an arrow, at which point B is going to press key l on the keyboard³²:

- 5(1)
- 108 B: **oh**
- 109 [(**move the mouse pointer**)] ((whistle)) (2) y[**eah** (1)]
- 110 [((moves mouse))] [((gaze: sc, kb))]
- 111 **click l** ((hand to l on kb, gaze up, hand to rest; knits eyebrows))
- 112 A: **did you click l?**
- 113 B: [**no**]
- 114 -> [((hand towards l on the keyboard))]
- 115
- 116 A: [**the left button?**]

³² The transcription conventions can be found in Appendix 1-1.

117 [((gaze to B))]
 118 B: ((hand to rest, gaze to mouse, hand to mouse))
 119 ([])
 120 [((↓↑))]
 121 ((■))

Line 114 in Extract 5(1) is a very strong indicator of what *Click-L* 'means' to B at that point: it refers to clicking l on the keyboard, in other words the *Click* (press and release) part of the acronym was correctly interpreted, as *click* is a transitive verb meaning, according to The New Penguin English Dictionary, "to strike, move, or produce with a click". It was the *L* part which for B at this point did not refer to the left mouse button but the key l³³. It could be possible to leave this misunderstanding sorted out through a repair device by A at that, just as a description of what happened. But as there had been no problem with *Click-L* before, i.e. the users had demonstrably learnt the meaning of this acronym at the beginning of the lesson, this brief moment in the encounter offered an exception, a deviance that should be accounted for.³⁴

The interactionist/constructivist/dialogistic views of reading and communication in general are all aware of interpretation as local, but also as social, cultural, historical and intertextual. So many forces might be at work in Extract 5(1) when B seems to have attached a new meaning to *Click-L*. As the acronym was something that was learnt for the purposes of this one encounter only, it is fairly easy to see 1) its potential intertextuality (how it was introduced in the mouse tutorial, and how clicking of the left mouse button was referred to in the actual learning program, or the Help option of Word 5.0; cf. above), and 2) its actualised intertextuality (what the users encountered, e.g. from the very beginning of the session (where the instructor

³³ The connection is enforced by the fact that all the letters on the keyboard are in upper case, and therefore *L* is visually identical with the key l. it is possible for a reader to see *L* not just as an acronym, but also as a sign referring to the visible letter on the key l. This connection elevates *L* from being just a grapheme, "a constituent element of a sign" (Nöth 1995, 263). The material circumstances allow for and cater *L* to be a symbol (paraphrasable as 'key l') which, in principle entitles a metaphorical usage to refer to the left mouse button. However, as the analogy between the key l and the left mouse button resides only in the visibility of the written version (Left) and the keyboard (L), the lexical metaphor is highly arbitrary.

³⁴ This resembles conversation analysis in that in CA nothing in the data is irrelevant *a priori*, as formulated by Heritage:

The basic orientation of conversation analytic studies may be summarised in terms of four fundamental assumptions: (1) interaction is structurally organised; (2) contributions to interaction are both context-shaped and context-renewing; (3) these two properties inhere in the details of interaction so that no order of detail in conversational interaction can be dismissed a priori as disorderly, accidental or interactionally irrelevant; and (4) the study of social interaction in its details is best approached through the analysis of naturally occurring data.

(Heritage 1989, 22)

was present as well) in connection with 'clicking/pressing/flicking/doing' 'mouse buttons' or 'keyboard buttons/keys'), which also forms the local history of language use and use of *Click-L*. The two types of intertextuality³⁵ are interconnected because problems in 2) might be a result of problems in 1), i.e. the program writers must have taken into account that not everybody would go through the mouse tutorial, and therefore use *Click-L* sparingly in the actual tutorial.

A detailed analysis of the data below will show that the users were trying to find a logic in how the program worked, and that one of the complicated reasons for why in Extract 5(1) B misunderstood A's direction/repair initiator was that the routine was not there. Therefore, the present data has evidence for what Reeves & Nass conclude about how people interact with electronic media: "When people know what to expect, they can process media with a greater sense of accomplishment and enjoyment" (1996, 254). In the case of a tutoring system, it could be assumed that the interactional space should be as clear and concise as possible for the users to be able to concentrate on the content of the learning program rather than the intricacies of the program itself.

Line 114 in Extract 5(1) and the ensuing repair by A give the analyst enough data to conclude that in this particular point in the encounter for B, *click l* meant clicking the key l on the keyboard. There follows a closer look into the locally wider context in order to track the possible reasons for this confusion. Already Extract 5(1) showed that the differing semiotic systems of writing and speaking might have contributed to the misunderstanding, i.e. when the acronym *Click-L* was spoken aloud (*click l*) the potential ambiguity of it referring to clicking the key l was enforced. However, to make the point clearer, the interaction preceding Extract 5(1) is needed, and is shown in Extract 5(1').

5(1')

- | | |
|-----|--|
| 102 | C: [=Move the mouse pointer to the "d" in "due". Click-L] |
| 103 | |
| 104 | B: [o-oh] |
| 105 | A: 'move the (.) mouse' |
| 106 | B: mouse |
| 107 | A: first |
| 108 | B: oh |
| 109 | [(move the mouse pointer)] ((whistle)) (2) y[eah (1)] |
| 110 | [[((moves mouse))]] [(gaze: sc, kb)] |
| 111 | click l ((hand to l on kb, gaze up, hand to rest; knits eyebrows)) |

³⁵ By concentrating on the local history, the Saussurian 'associative solidarities' (e.g. Thibault 1997, 265) are limited to those that can be traced in the actual encounter.

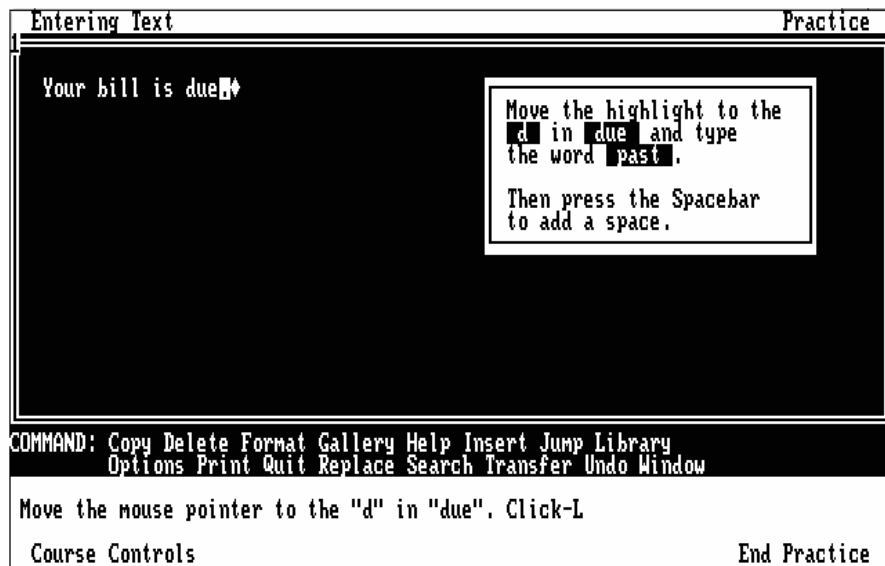
112 A: **did you click l?**
 113 B: **[no]**
 114 [(hand towards l on the keyboard))]
 115
 116 A: **[the left button?]**
 117 [(gaze to B))]

Extract 5(1') begins with the repair initiator that appeared at the lower part of the screen after a wrong action by B. Now there is a clear connection between the computer's 'turn' in line 102, and what B says in lines 109 (*move the mouse pointer*) and 111 (*click l*): he is quoting what he sees on the screen (though he is muttering the words to himself, rather than reading aloud to A). In line 111, B's actions and facial expression show that his uncertain analysis at that point is that *Click-L* refers to the key l on the keyboard. A starts an other-initiated other-repair (resulting in self-repair) in line 112 (*did you click l?*), but as *Click-L* is spoken aloud, the interpretation of it as meaning that he/they click the key l is enforced. The consequences of the difference between *click l* as the spoken form of the written *Click-L* reflect what Halliday has said about the differences between written and spoken modes of language: "Writing brings language to consciousness; and in the same process it changes its semiotic mode from the dynamic to the synoptic: from flow to stasis, from choreographic to crystalline, from syntactic intricacy to lexical density" (1993, 118). The lexical density of written text in this moment of spoken intertextuality creates problems rather than solves them. This is the reverse of the cases in Hutchins' (1995) discussion about the materiality of signs (e.g. *Light2* and *LightZ* can be mixed up as visual signs but not when spoken aloud: 'light two' and 'light z'). After B starts moving his finger towards the key l on the keyboard, A glances at him, which is a very distinctive action in an 'ear-to-ear' situation i.e. when the persons are sitting shoulder to shoulder it is not very often they turn to look at each other. A's gaze occurs with his other-initiated other-repair (*the left button?*) where the referent of *L* is made explicit. With the question intonation, A produces the other-initiated other-repair as if it was a self-initiated self-repair of his own turn (i.e. clarifying *did you click l?*). An analysis of Extract 5(1") which extends the example to show what happened next, reveals that although the problem seems to have been resolved and the side sequence that the repair initiator started has finished, B seems to still be confused.

5(1")

102	C: [=Move the mouse pointer to the "d" in "due". Click-L]
103	
104	B: [o-oh]
105	A: 'move the (.) mouse'
106	B: mouse
107	A: first
108	B: oh
109	[(move the mouse pointer)] ((whistle)) (2) y[eah (1)]

93 B: **heh** [he]
 94 |
 95 A: [oh] there we go now 'move the highlight to the d in due
 96 (.) type the word **past**'
 97 B: **that's** what's () gonna do. (move the)
 98 [(highlight)]
 99 [((moves mouse; pointer to the lower right corner))]
 100 **let's press this** <-=



101
 102 C: [=Move the mouse pointer to the "d" in "due". Click-L]
 103 |
 104 B: [o-oh]
 105 A: 'move the (.) mouse'
 106 B: **mouse**
 107 A: **first**
 108 B: **oh**
 109 [(move the mouse pointer)] ((whistle)) (2) y[eah (1)]
 110 [((moves mouse to d))] [((gaze: sc, kb))]
 111 **click** 1 ((hand to l on kb, gaze up, hand to rest; knits eyebrows))
 112 A: **did you click** I?
 113 B: [no]
 114 [((hand towards l on the keyboard))]
 115
 116 A: [the left button?]
 117 [((gaze to B))]
 118 B: ((hand to rest, gaze to mouse, hand to mouse))
 119 ([])
 120 [((↓↑))]
 121 ((■))
 122
 123 B: ((gaze: screen; moves back))
 124 ((gaze: down screen, up screen; moves mouse pointer to the right))
 125 **okay so (I want to move my) mouse pointer**
 126 [gaze: down screen; moves mouse pointer))]

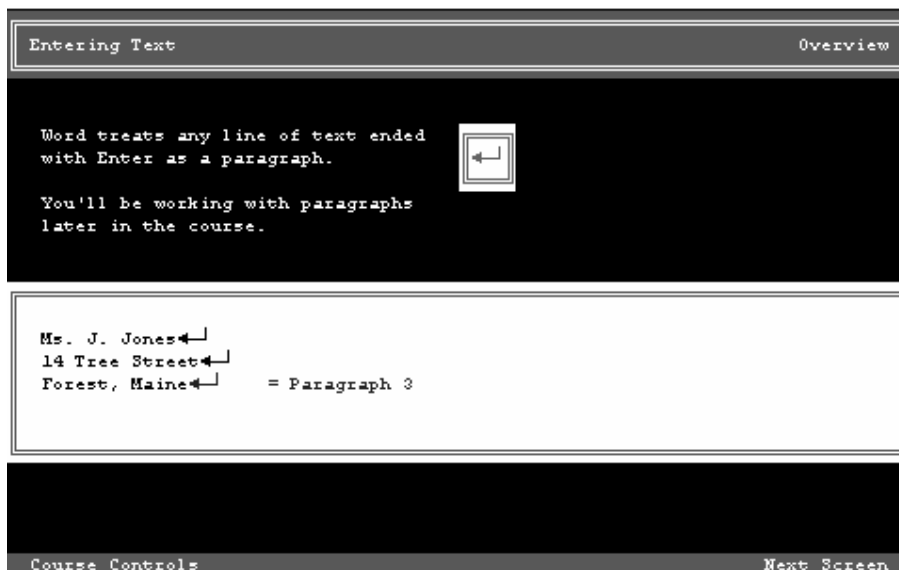
127 |
 128 A: [yeah]
 129 B: **to the d** ((gaze up screen))
 130 [((hands palms up))]
 131 |
 132 A: [**and now you can type**]
 133
 134 B: [**do you**] think?
 135 [((hands palms up))]

The directive (quoted on lines 95 and 96, and prefaced with *oh there we go now* in which *oh* conveys noticing something (Heritage 1984a)) did not explicitly mention that the highlight should be moved to by the mouse. B's *that's what's () gonna do* (line 97) makes it clear that his attention is drawn to the same piece of text. He does not use the mouse, however, but leaves the mouse pointer at the lower right corner (line 99) and tries to use the backspace key (line 100), a logical step in the sense that the use of the backspace key was taught a bit earlier. B's *o-oh* in line 103 indicates that he noticed the message appearing at the bottom of the screen and that his pressing of the backspace key was not sequentially correct. A quotes the instruction with a pause which makes *mouse* foregrounded (line 105). By repeating *mouse* (line 106), B makes it interactionally relevant as well, but does not do anything. It is only after A's added *first* to his quote that B seems to get what A means, and produces an *oh* receipt (Heritage 1984a, 319) to acknowledge he understands what A proposed, namely that the first thing to adhere to is the instruction at the bottom of the screen. In other words, B did not treat the message appearing at the bottom of the screen as a repair initiator that needs to be dealt with immediately. In lines 109 and 110, B demonstrates that he is following the first part of the new instruction, but *Click-L* seems to be a problem to be locally solved, as explained above. The question still remains: why would B have this new referent for *Click-L*, or rather *L*? B is clearly hesitant to press key *l* (line 114), and this could be an indication that he still has the mouse click as the other option. However, a closer look of what had happened earlier in this Entering Text module of the tutorial gives evidence that what B might be perplexed about is not a choice between these two interpretations, but between whether *L* refers to clicking key *l* on the keyboard or clicking *l* in the word *bill* on the screen.

The *Practice* module was preceded by *Overview* and *Steps* screens that might give an explanation for B's problematic interpretation of *Click-L*. The following extract starts with the participants quitting the practising of entering text with and without using the 'enter' key for a new line. *Click-L* appeared on that screen within the instruction to go on (*When you've finished, point to End Practice and Click-L*), which was carried out with no difficulty (see lines 1-5 in Extract 5(2)). So by this time in the encounter, there had been no problems establishing *Click-L* as referring to *press and release the left mouse button*.

5(2)

- 1 B: **okay.** [(we've finished with that)]
2 [(glances at the mouse)]
3 **end of practice (here we go)** ((glancing at the mouse, ↓↑))
4
5 **[tsh]hh**
6 [(°)]
7 ((pushes the keyboard back to the middle, right hand to the mouse))
8
9



- 10 B: **mm (where do we) paragraphs**
11 [(gaze: screen; moves the mouse pointer to *Next Screen*)]
12 |
13 A: [(gaze: screen)] '**later in the course**'
14 B: ((looks at mouse)) **u:h so we go (we go) to t[he next screen]**
15 |
16 A: [°go next screen°]
17 B: ((↓↑))

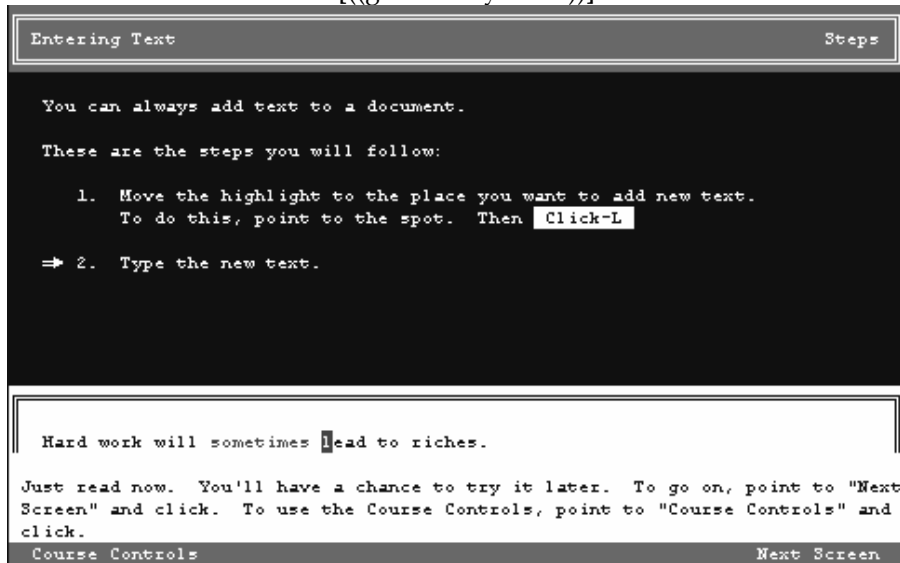


18 '(you can always add text to a document)'
19 (4)
20 (oh) ((knits eyebrows)) o:h ((g:scr)) so! [((g:scr))]
21 |
22 A: [((g:scr))] ('to the left of the
23 high[light'])
24 |
25 B: [so you] put (.)
26
27 [()]
28 [((hand to the side of neck))]
29 A: 'what you [type will appear to the left] of the highlight'
30 |
31 B: [a:h (.) so that adds a space]
32 [((hand to point to A, cheek, rest))]
33 puts a space in ((↓↑)) ((gaze: screen)) ((animation on
34 the screen 'typing' (with a rattling sound) some[*times*]))
35 |
36 A: [()]



- 37 A&B: ((gaze: screen))
 38 B: 'move the highlight to the place you want to add new text.
 39 type the new text'
 40 ['hard work will sometimes lead to riches']
 41 [(((gaze to lower part of screen with head movement down)))]
 42 ((gaze: upper part of screen))
 43 A: [is that ()?]
 44 [(((moves closer to screen)))]
 45 B: ((moves his left hand to the keyboard; looks up))
 46 A: oh [there there]
 47 [(((smiles; left h. to screen)))]
 48 h:a(h).
 49 B: we [need to put the] l (. to this
 50 [(((index finger to press)))]
 51 A: do we do that now?
 52 B: ((hand away from kb)) we wanna add the sometimes?
 53 [(((tongue smack)))]
 54 [(((moves to left)))]
 55
 56 [() last screen]
 57 [(((whispering, smiling)))]
 58 ((right index finger pointing screen))
 59 [uhm:]
 60 [(((scratches chin)))]
 61
 62 [(((finger to kb; gaze: screen)))]
 63 |
 64 A: ['type the new text'^o]
 65
 66 B: [(yeah) we will put it to the left (2) or shall I shall I figure out
 67 sometimes (on there) (2)]
 68 [(((gaze: screen; right index finger on backspace key)))]
 69 ((right hand to point screen))
 70 [do you think?]
 71 [(((moves to writing position)))]
 72 A: give it a try

73 B: (to) see what happens [s=]
 74 [((gaze to keyboard))]



75 C: =Just read now. You'll have a chance to try it later. To go on,
 76 point to "Next Screen" and click. To use the Course Controls,
 77 point to "Course Controls" and click.

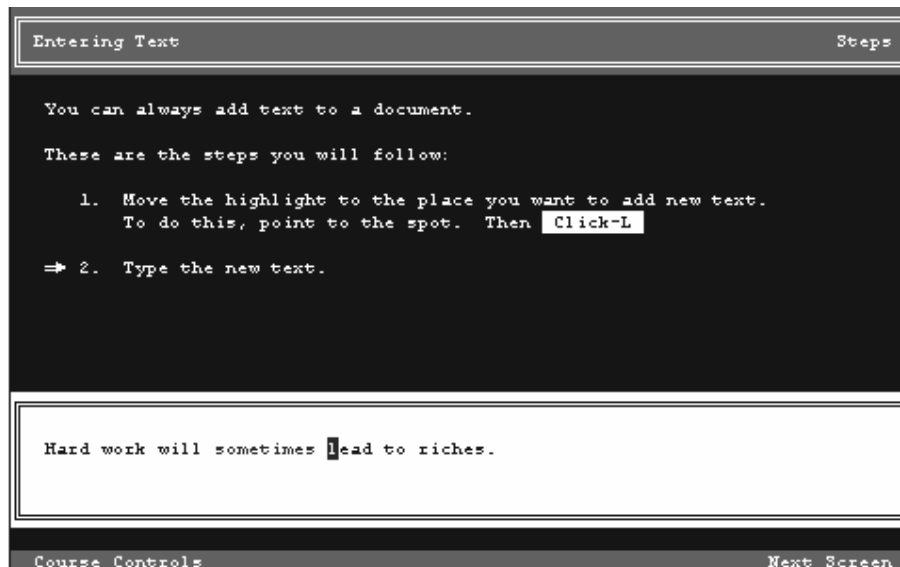
78 A: [(.) EH ha ha]
 79 |

80 B: [((gaze to screen))] wh[at happened?]
 81 |

82 A: ['just] read now' ((laughing voice))

83 B: [huh?]
 84 [((gaze: screen; hands on kb))]

85 A: **impolite thing**
 86 ((■))



87 B: [oh:]
 88 [((moves away from keyboard))] ((sigh))

89 A: so I guess we just go to the [nex- next screen=]
 90 |
 91 B: [=go [to the next scre]en.]
 92 [((↓↑))]

A and B encounter the new frame of knowledge (cf. the CAI phases) on the screen reproduced between lines 9 and 10. Lines 10 to 17 display one of the objects of continuous meaning negotiation in the encounter: is the screen for reading only or for reading and doing? B's words in line 10 seem to orient to the latter interpretation (before quoting the screen in *paragraphs?*, B says *where do we?*, which orients to doing something, as does the text before what he quotes: *You'll be working with*). This is how A interprets his turn (line 13), producing an other-initiated other-repair by transforming the screen text (*later in the course*). In line 14, B seems perplexed (*u:h. so we go we go*), the reason for which might be that he already had oriented to moving on by positioning the mouse pointer on *Next Screen* (line 11), and therefore A's contribution may appear to B to be redundant and unnecessary.

The next screen (between lines 17 and 18) which appears as a result of the mouse click on *Next Screen* (line 17) is a complicated one: there are deictical references within the text and as it is describing the steps to take to add text to a document, the imperative mode is used (in contrast with the *Overview* screen between lines 9 and 10). This screen also introduces *Click-L* for the first time in the *Entering Text* module instructions though here *Click-L* is not used to refer to clicking the right hand lower corner of the screen.

The following two subsections point out two types of complexities that language-in-use creates in this case, firstly because of its materiality, and secondly because of the read/act dichotomy instructional texts produce.

5.5.2. Potential hazards created by the materiality of the text

The first sentence (*You can always add text to a document.*) is a declarative one, and *always* makes it a description of a general state of affairs; the following one introduces what has to be done: *These are the steps you will follow*. Numbering the step as 1. ('one' or 'first'; it can be understood as a number, and a number in a sequence of numbers yet forthcoming) and attaching an (red) arrow next to the instruction can be understood to indicate that what A and B see in front of them is a first step. However, the sentence in the next line *To do this, point to the spot. Then Click-L*, in this particular screen allows for more than one referent for *L*. Firstly, the reference of the anaphoric expression *To do this* is not quite clear: does it refer to *move* or to *add*? If the referent is *Move the highlight to the place*, and it is compared with the example at the lower part of the screen, one plausible interpretation is that the spot has already been pointed to and the letter *l* of the word *lead* clicked; after all, there is an arrow pointing at the letter *l*. Also on the original screen, both the highlight,

the arrow and the box frame, together with text, were red in colour, and the arrow and the box appeared on the screen a couple of seconds later than the rest of the screen, thus being foregrounded as important. If *to do this* refers to *add*, then the interpretation of the rest of the instructions could mean: point to the spot and click key 'l' on the keyboard (so *ead* becomes *lead*); a similar interpretation would result if *To do this* referred to the whole of the preceding sentence (i.e. *Move* and *add*). To complicate the meaning potential, the word *left* appears in the box text, and *L* has been used to refer to the left mouse button earlier.

Complex intertextuality and meaning potential might be the cause of A's and B's perplexity at this point in their use of the program. A seems to be concentrating on the (red and thus foregrounded) text in the box (lines 22-23 and 29). B's interpretation of what the text in the box says also reveals that he considers the example text as depicting what has happened when the instruction in 1. is carried out (and *To do this* refers to the whole of the preceding sentence). To give an explanation to the problematic sentence in the box, B finds a logical solution: *that adds a space, puts a space in* (lines 31-33), i.e. after the letter *l* is highlighted/pressed, a space has been added to the left of the highlight (by the program).

5.5.3 *To do or not to do*

The next screen has an animation adding the (red) word *sometimes* to the text in the example, something which A misses at first, even if the animation has a sound effect of typing (the clattering sound of which does not exactly coincide with the appearing of the letters). In line 49, what was previously hypothesised from B's action, now becomes visible: B interprets *L* to refer to the key l, and number 2. instruction is an instruction to carry out the action (as number 1. was interpreted to be an example-to-read about selecting the spot and typing there). A stops him by questioning, at least on the surface, the sequential placement of the activity of pressing l (but not the activity itself, i.e. A does not directly accuse B of a wrong interpretation of the screen as one to practice on): *do we do that now?* (line 51). In fact, A's turn is ambiguous. B takes the other-initiated other-repair to be about 'doing what', rather than about 'reading, not doing'. B suggests another text to add, which is the same text as was typed last by the program (*the sometimes*, line 52). There is no acknowledgement by A of this understanding check/suggestion; A's next contribution is in line 64 which shows that A is still pondering over the instruction to type the new text. For B, who is already convinced that they have to do something, and who is waiting for A's evaluation of what to do, A is producing a request rather than quoting a repairable. B then self-repairs his pressing of key l to moving the highlight *to the left* (his finger ready to press the backspace key), but A, again, does not comment on B's suggestion (2

seconds pause in line 66), which results in B reformulating his earlier noncommented suggestion *we wanna add the sometimes? as shall I shall I figure out sometimes on there*. B is taking the responsibility and is showing that the last proposal is his intent about which he is asking A's desire/opinion (see Boyd 1992). A is still silent, so B elicits A's assessment again while moving to the writing position (*do you think?*). A's reply makes it clear that he is not sure whether this is what should be done (*give it a try*), nor is B (*to see what happens*). However, because B is in a typing position and looking at the keyboard, he misses the repair initiator that is promptly produced by the computer, and to which A immediately strongly reacts (line 78). B produces an other-initiated repair (*what happened?*), which in the circumstance is directed to A. As a repair, A quotes C (*just read now*), but he transforms the repeated text because in the process of quoting it he fits it in his own action of laughter (line 82). The computer's turn is different from the other repair initiators in that it disappears from the screen without A's or B's correct action (repair): B did not necessarily understand what was going on by the time the fairly long message vanished (even if in line 84, B displays some sort of change of state). However, though the message was temporary, it was still visible for some time and available to be used as a resource.

Repair initiators and the screen types encountered

If the local history of the encounter is to contribute to the meaning making, then it is important to know what repair messages the users had come upon and in which circumstances before those in Extracts 5(2) (*Just read now...*) and 5(1) (*Move the mouse pointer...*). The very first screen that the users encountered (after the copyright title of the program) was one that had an error message visible at the bottom of the screen.

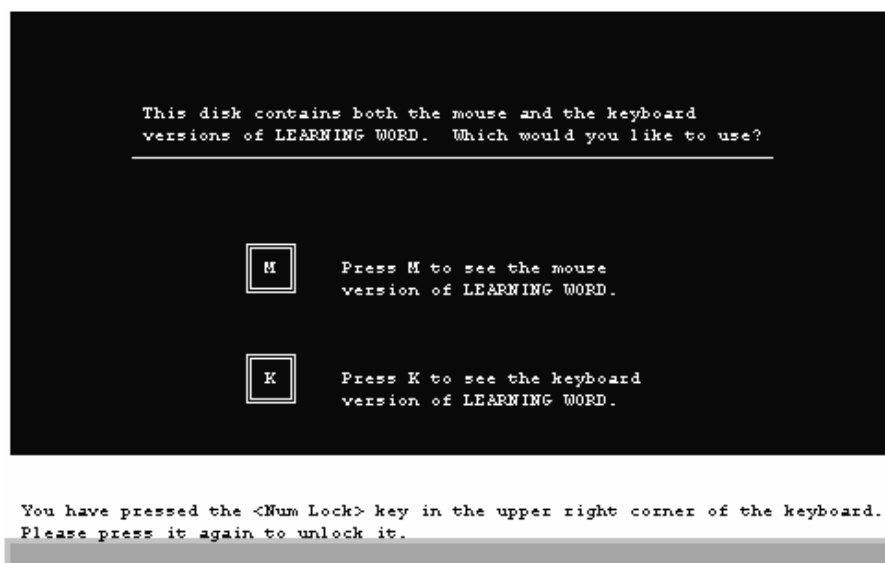


Figure 5-5

This is how the users came to find the corrective message:

In Extract 5(3), the users learn that a message at the bottom of the screen has to be dealt with first to be able to continue (the accusation '*you have pressed...*' given in the error message was wrong – as A claims in line 17, *I don't think we have* – because the users actually never had pressed the NumLock key; the program/system had activated it earlier for some reason).

5(4)



- 7 C: *Move the pointer to the 'Next Screen' that appears in the lower right corner of*
8 *the screen. Then press and release the left mouse button.*
9
- 10 B: **'pointer to the next screen (that) appears in the lower right corner of the**
11 **screen'** [((sharp movement of head twds the screen))]
12 |
13 A: [((sharp movement of head twds the screen))]
14 B: **(next) screen**
15 A: **[it must be (that)]**
16 [((points to the lower right corner with left index finger))]
17 B: **aha:. okay** ((looks towards mouse)) ((↓↑))((■))
18 C: *EXCELLENT! Try it once more.*
19 B: **h[a ha ha ha.]**
20 |
21 A: **[try it once more.] heh**
22
23 B: **[er (.) this.]**
24 [((gaze: screen; pointer to 'next screen'))]
25 A: **yeah**
26 B: **[this.]**
27 [((gaze: mouse))]
28
29 **[try this on(ce more)). ((↓↑)) ((■)) okay.]**
30 **[((gaze: screen))]**

Though the repair initiators of the kind shown on the screen between lines 6 and 7 would potentially appear every time *Next Screen* was not correctly clicked, A and B did not encounter one after this instance (i.e. they clicked the words *Next Screen* correctly). The praise and direction to do the same action again (*EXCELLENT! <new line> Try it once more.*) replaced the text to the left of the picture of the mouse, viz. *NOW YOU TRY <new line> 1. Move the mouse pointer so that... etc..* Thus, the history of the encounter revealed an inconsistency in the (outlook of the) screens for practice and for reading.

5.5.4. Click-L revisited

Extract 5(2) ends with the users entering the practice stage, which we have seen in Extracts 1-1"', to go through the local details to find possible reasons for B's misunderstanding of the reference of *Click-L*. The limited history of the users' encounter with the *Overview* and *Steps* modules showed that B never actually got to try his wrong interpretation of *L* as key *l*, and therefore he was never repaired by the program. Originally, the wrong 'pointer' seems to have been caused by the ambiguous visual layout of the *Steps* screen (see Example 5(2)), the picture between lines 17 and 18), which allowed for *L* to be connected with the letter *l* in the word *lead*. When the reference of *To do this* in the same screen also allowed for several candidate directions, one possibility was to link *L* and letter *l* of *lead*, and at the same time, to key *l* on the keyboard. Thus, the extrasign materiality and semiotic materiality of the screen was strong enough to override the carefully taught and successfully practised connection between *L* and the left mouse button: the lexical metaphor (e.g. Martin 1992) *L* in the technical term *Click-L* was unbound, and a less complicated connection (*L* = *l*) established. In other words, the new referent of *Click-L* was an example of how, also from the receiver's point of view, "specific associative patterns are evoked from the virtual associative series that constitute 'an entire latent system' in response to specific contextual contingencies" (Thibault 1997, 274).

The *Click-L* episode gives a concrete example of how the actual context of interpretation is crucial for meaning making. The semiosis of the human participants depicted in the left hand side of Figure 5-3 (and pictured also in Figure 5-1) does not capture the details of how the interpretant *I*, gets to be formed due to the context. In that it is reminiscent of other semiotic accounts of human meaning making: the grain size of the analytical object is either too big or else only hypothetical. Conversation analytical investigation was needed to get into the how and when of the layering of the interpretant.

The wrong interpretation was later maintained through the possibilities created by stepwise meaning making. The repair message *Move the mouse pointer to the "d" in "due"*. *Click-L* was an example of a typical turn-analysis repair; it was meant to correct the user's previous action. However, when the users encountered this repair message, they had just dealt with the turn-taking repair initiator of *Just read now. You'll have a chance to try it later. To go on, point to "Next Screen" and click. To use the Course Controls, point to "Course Controls" and click.* This message disappeared without the user's doing anything, which was logical in the sense that it was only meant to be read, not acted upon. However, the latter part of the message gave two options of doing, and thus did deal with 'a right action'. B treated the *Move the mouse pointer..* repair initiator as a message that also disappeared by itself rather than after the correct action. This is why B did not seem to connect the side sequence of the reference of *click l* being corrected by A's talk and the repair

message itself: B orients to the disappeared message as a relevant next. Part of this foregrounding is due to the higher rank of interactive computer messages in the order of 'mustness': it was always necessary for the user to do/stop doing something after an interactive message, whereas the practice instructions could always be bypassed (the program did not complain if the user did not do the practice; indeed, this actually happened once in the encounter).

5.5.5. Encountering an 'other'

Reeves & Nass (1996) made the observation that people treat computers and other media as if they were human. Their report was based on psychological studies on attitudes. The present report comes to the same conclusion, though at the level of social interaction. The program was treated as if it was a logical entity in that the users were trying to accommodate to its way of 'behaving' (Examples 5(1) and 5(2), also comments like *it hasn't told us to do anything yet*, attitudes: *impolite thing* (Example 5(2), line 85, etc.)³⁶. The praising comments (e.g. *Great!* and *EXCELLENT!*) always created amusement and, when appearing after a repair sequence, laughter. Evaluating responses are a category that in human-human interaction have been claimed to contribute to intersubjectivity (e.g. Severinson Eklundh 1986, 37). The participants clearly were entertained by the human feature which the tutorial program exhibits. There were moments in which Learning Word 5.0 was objectified as a program (*thing, it*, talking about going to *next screen* (e.g. Example 5(2), lines 14 and 16) or *end of practice* (Example 5(2), line 3), thus using the technical terminology³⁷ instead of, for example, 'going ahead' or 'moving on'. Also a

³⁶ Beach reports similar observations about users' comments, reflecting how "inanimate "objects" can be oriented to as, essentially, interactional partners" (Beach 1990, 213). It is interesting that a conversation analyst would draw such a conclusion from talk which took place away from the computers, in a meeting. Interestingly, similar humanising comments are not made concerning people's talk about newspapers (e.g. "The Independent tells today..." on BBC World, 5.3.1998; "says the paper; it says" 11.3.1998)

³⁷ Actually increasing the technicality by transforming the active form *End Practice* to a noun phrase *end of practice*.

mixture of both could be detected: *let's see what happens/it says* as depicting a game-like situation in which 'it' is going to react somehow. Thus, the 'other' of the program that is encountered, is not a unified concept but flexible and context sensitive. In fact, the program itself at points of interactive features exhibited its limits as an interpretant. The episode documented in Extract 5(1''') continued as follows:

5(5)

136 ((hands to keyboard, gaze to screen)) *d u e* ((back in chair; hand
137 to mouse)) **mm**

138 A: **h::**

139 B: **heh [he he]**

140 A: [**ha ha**] **I don't think that's (.) exactly (.) exactly like that**
141 ((gaze to the lower part of the screen))

142 B: ((gaze: screen)) **please format (that)**

143

144 [(((moves the pointer to 'd' in 'due' in the upper screen)))]
145 [(((back in the chair; hand to chin)))]

146 |

147 A: [**oh er: (.)**]
148 [**'press the spacebar to add a space'** ((presses spacebar))]
149 |

150 B: [**'move the highlight to the d in due and type the** word **past**
151 **I've I've**

152 A: **HEH HAH HAH [HAH HA]**

153 |

154 B: [**typed the wrong** ((laughing)) **hah ha h**] **a**
155 **wrong word due [to (it)]**

156 A: [**ju**] **st go to [end of practice]**
157 [(((points at End Practice)))]

158 **and see what it says**

159 B: ((back in the chair; gaze: screen)) **just let me let me do past to**
160 **see what it does (then)**

161 A: **okay**

162 B: **just out of curiosity. it might accept** ((takes the keyboard)) **like**
163 **this** ((gaze: kb; hands to over the kb)) **(let's) see if we can outfox**
164 **the machine**

165 A: **uhum well we should [be smarter than the machine (we have)]**

166 B: [*p a= s t*]

167 C: =Please check your spelling (making sure you typed a space after
168 "past" to separate if from "due"). Use the Backspace key to delete
169 characters you don't want.

170 A: **'please check your spelling' heh heh he**

171 B: **heh heh okay**

172 A: **yeah**

173 B: ((glances at mouse)) [(((clicks the mouse)))(**■**)]

174 A: [**go ahead**]

In line 136, B types the word 'due' (and not 'past' as was instructed by the program) to the *Practice* screen. However, at this point the predesign of the program does not take into account that a wrong word could be typed in, and therefore the program does not initiate repair at this point. Thus, the

coparticipant A, carefully orients B to reread the instruction on the screen by quoting the second piece of advice and performing the action himself (line 148). B simultaneously realises what his error was and quotes the direction, stressing the word 'past' to mark his understanding of the problem (line 150) which he specifies in lines from 151 to 155. After this, B decides to try out the correct interpretation of the direction *to see what it does then* (159-160), he wants to try if they *can outfox the machine* (163-164), as they according to A *should be smarter than the machine* (line 165). It is noteworthy that the users do not try to delete the wrong word: they orient to the program as logical because using the backspace key had caused an error message just before (line 100 in Extract 5-(1'')). In lines 167 to 169, the machine does 'say' something as a response, by interrupting B with an error message that does not make sense as only the letter 'p' out of the word 'past' actually shows on the screen. Also, now the use of the backspace key is specified as a legitimate action.

5.5.6. Summary

The problem of wrong reference or disruption in the temporally shared social reality between the reader and the meaning of the acronym that the program had managed to create was caused and maintained because the users treated the program as functioning with a particular logic. For instance, they tried to exercise the skills they had been taught in using the text processor, and therefore they used the backspace key to erase text (to do what they had seen to be the visual 'meaning' of the wordings 'to move the highlight' 'to the left', which does take place when backspace key is hit). Also, they tried to be methodological in interacting with the interactive (repair) messages: usually adhering to how the program had behaved 'the last time'. The design of the program was very modular, and the interactive features which were 'true' for one module, were not in another one. As was demonstrated above, the users were methodological, they stuck to their latest interpretations of referents and ways of interacting, exhibiting orientation to meaning as something that "entails order, pattern, regularity" (Thibault 1997, 291). The interactive force created by the repair messages was understood to be the strongest: they could not be ignored. However, this did not mean that the 'repairness', the message as beginning a corrective side sequence, was understood as such. All in all, the users oriented to the content, and also to the interactive order of the program, as if in a tutorial with a real word processor and a real tutor.

On the basis of this encounter, *Click-L* was an instantiation of a prototypical *click-something*, and the latter part got an unintended referent due to the material circumstances. It was also shown above that the wrong interpretation of *Click-L* was carried on partly because the program was unable to make clear for B which messages were repair initiators and which mere instructions. And when B had just learnt that the 'orders' of repair

messages can be oriented to, even after they disappear (*just read now -- go to next screen*), he applied this latest knowledge to the present case as well. The indistinctiveness of these two type of messages was increased by the fact that the repair initiators were not clearly connected to the action/interpretant of an instruction they were correcting. So in Example 5(1st), the screen between lines 9 and 10, the wordings of the instruction (*Move the highlight to the d in due and type the word past.*) and that of the repair initiator (*Move the mouse pointer to the "d" in "due". Click-L*) start identically. The repair initiator did not mark the repairable, the troublesource, (for example, by *To move the highlight, move the mouse..*, or even, *No, to move the highlight you have to move the mouse pointer..*) and the only hint that it is a repair initiator is that it surfaced after an (erroneous) action. The design of the program relied totally on the users' understanding that interactive messages from the computer were repairing ones. Strate regards the metatime of cyberspace as a reason for "the breakdown of the signifying chain" (Strate 1997, 373). The example of *Click-L* shows that also the sequential time of semiosis, when it is represented in a computer program, can be part of the breakdown; in face-to-face sequential interaction between humans ambiguities of meaning seldom arise (Schegloff 1984).

5.6. Discussion

The purpose of this chapter was to show that the meaning stratification, the layers of meaning, the interactive accomplishment of meaning making, the dialogic alternative to meaning negotiation, to name a few perspectives on meaning making, can be subtle and intricate. In the case study, I demonstrated that for one of the participants, the material/visual semiotics, embedded in the sequential meaning making, created another interpretant for the recently learnt *Click-L*. It is a concrete example of "the cross-coupling of the material and the semiotic domains, which is a necessary condition of all meaning making" (Thibault 1997, 287).

Thus, in instructional contexts, in which the intent surely is to be as clear as possible, indeterminacy of meaning exists. Temporal socially shared meanings are transient: the interactive contingencies disrupt the good intentions of the author(s) in ways that they might not have anticipated. This is why close inspections of empirical materials of the kind reported here are needed to connect the social semiotic to social interaction and to appreciate the complexity of the 'context of use', not as an additional separate 'factor' of meaning making, but as intertwining text, context, and intertext into aspects of sequential interpretation. When the aim is to channel the user of an electronic (instructional or otherwise interactive) text, the interpretative framework, the extrasign and semiotic materiality that affects the interpretation, has to be researched. Towards that goal, conversation analysis

as a method is needed to reveal the sensitivity of the situated and sequential meaning making process to the material and semiotic context. According to Suchman, though intelligent machines are meant to simulate human cognitive processes, "the measure of success is at bottom an interactional one" (Suchman 1987, 2), i.e. the user's encounter with the machine is smooth. The same goes for any language artefact, be it a printed manual, a word processing program, or a word processing tutorial program, that is meant to provide support to the user/reader to accomplish the task they are doing. The present chapter also showed that simply imitating a teaching process in computer aided instruction does not take into account that the basic interactive features of the process ('positive feedback', i.e. evaluation and 'explanation', i.e. repair work) in human-human interaction are seen to confirm the participants' intersubjective understanding of what is being talked about (Heritage 1984b, Schegloff 1992). This has two kinds of consequences for human-computer interaction: either the computer's evaluative 'turns' are oriented to as funny (i.e. out of place), or interactionally more consequentially (partly due to the complexities of the semiotic surroundings), when the system's next turn repair initiators are missed, the system has no capacity to continue the "routine grounding for intersubjectivity" (Schegloff 1992, 1295), i.e. to repair the user's misunderstanding of the first repair initiation

Specific textual formations, in the present case an abbreviation, help the researcher to find out how accessibility of computerised (and other) texts is not a once-and-for-all phenomenon, but in flux. When a specifically agreed convention turns out to be indexical, we have evidence for the creativity of the interpreter (see Streeck 1980), aided by the visual (cf. reading) and the sequential (cf. interaction) resources offered by the program, the latter being different from traditional (printed) text-user encounters. The question of what makes texts easy or difficult to understand is an empirical one, the answer to which can be found by using qualitative empirical methods, whatever the theoretical framework.

6. TASK: BETWEEN READING AND ENCOUNTERING IN HUMAN- (INSTRUCTIVE) TEXT INTERACTION

In the previous two case studies, the users encountered an audio dialogue answering machine (TELEPHONE), and a visual training program which instructed them by giving information and by engaging them to try out what they had just been reading about (TUTORIAL). In interacting with the answering machine, the users had no 'time off' but had to make a next step after each of the system's turns: they had to be 'doing by saying'. In TUTORIAL however, the users encountered two types of computer screens, one giving distinct instructions about what the users should do with the other screen of the practice mode. Thus, although displayed by one entity (the computer), the instruction and the activity to be taken were dealt with separately by the program. This, together with the visual semiotics of the screen and the sequential activity in the use of it, resulted, among other things, in confusion for the user(s) about when to act (*Practice* screen) and when to read (*Overview* and *Steps* screens).

6.1. Introduction

In this chapter, another type of encounter with computerised media is studied. In this case study a pair of volunteers undertake the task of producing mailing labels with a real word processor (Microsoft Word 2.0); the case is called 'TASK'. The Finnish students of English who agreed to do the assignment had a computer manual available for guidance in using this rather specialised function of the word processor. Now, the artefact to accomplish an activity and the directions to do were two separate textual objects: the computer and the manual. However, whereas in TUTORIAL the program made a distinction between instructions to read (learning *that*) and instructions to act upon (learning *how*), the written instructions of the manual in TASK collapse the two modes into one: a set of directions can be read by the user for information or they can be resorted to as a step-by-step guide during the user's activity. Unlike the learning program in TUTORIAL, the

manual could not influence what appeared on the screen at all; the reader had to be the mediator. However, the problems for the user-readers seem similar: is the text on the screen/paper informing the reader about actions or requesting the reader to undertake those actions? The differences between the cases become clear when the user's wrong interpretations have dissimilar consequences: for example in TUTORIAL, the program actively repaired the users' interpretations, whereas in TASK the users have to sort out the difficulty between themselves, with the manual and the program as static resources. So in TASK, there are three types of interaction going on: human-computer, human-manual, and human-human. The interactions with a textual object (computer, manual) deal with signs which are always produced in the past. The main activity at hand is to 'communicate' with the text processor and this work is facilitated by the human-manual interaction. Though texts do not interact as such, the notion 'human-text interaction' emphasises the intention of instruction writers to make the user act on the basis of the text. As the task was done by a pair of students, both of these normally silent participation frameworks with one human working with electronic or paper text are 'externalised' into talk and gestures to account for one's understandings to the other participant. For example, Extract 6(1)a below starts with reading in silence, but soon the readers begin to exhibit to each other either that they understand what the text is about (*mm-m*), or what they think is the relevant next by quoting "*choose new*" and/or by pointing at the text.

The students had used the text processor Word 2.0 but not to produce mailing labels: they partially knew the interactive environment. Thus, the encounter differed from that of the emergency call-takers (e.g. Smith & Whalen 1995, Zimmerman 1992) who thoroughly knew the computer system they were using, and therefore encounter the slots on the computer screen as 'requests' they understand how to answer. In Smith and Whalen's analysis, "the empty fields of the computer form are questions insistently seeking response" (Smith & Whalen 1995, 12). The bulk of the call-takers' work is to get the information needed out of the caller and in that talk the electronic form in front of them plays a role; everybody has been trained how to deal with the texts encountered and produced. However, the screen the students met posed problems, and the manual which was meant to help them could actually increase the problems, even if both the program and the manual were designed to be readable and usable by anybody who knows the English language, and therefore no specialised knowledge was required.

Due to their apparent incomprehensibility to users, computer manuals have been the target of fairly intensive research (see e.g. the Journal of Technical Writing and Communication). Manuals have been studied especially in the U.S., where technical writers belong to an established profession, employed mostly by computer companies. The task of technical writers is to produce readable texts both for inside and outside users of the

company's gadget. Much of the research on manuals or other 'necessary' texts has concentrated on the features of how information is presented in the texts, or on the linguistic properties that might make the text accessible or effective (e.g. Jordan 1994, Rogers & Brown 1993, Shubert et al. 1995, Teklinski 1993). The linear nature of the book format as such has been regarded as unproblematic, and the user's understanding as straightforward: "the user is relatively passive, with action restricted to appreciation and judgement of someone else's selection" (Barfield et al. 1994, 54-55). As in other reading research, an increasing amount of effort is spent on explicating the role of the reader and the context of reading as a crucial factor, though the wider aspects of manual writers and their normative values have also been investigated (e.g. Mårdsjö 1994). A general trend, however, has been towards understanding the 'here and now' quality of manual use, the 'enunciation of text' (Barthes 1977) in the user's act of reading.

At the same time, some researchers who are closely connected with the conversation analytical research tradition and ethnomethodology have started producing a growing body of 'beyond here and now' research. These studies want to say something exceeding discrete conversations, and consider communicative practices across modalities (e.g. Firth 1998, Goodwin 1994, Hanks 1996, Kleifen & Frenz-Belkin 1997, Linell 1998, Mulkay 1985, Smith and Whalen 1995). A special interest has been the various intertextualities of texts and conversations or other actions, which is a concern of many other communication scholars, especially those who work within the Bakhtinian tradition. Also, corporeality (gaze, gestures, facial expressions, etc.) in the meaning making process has been considered crucially important, especially in Charles and Marjorie Goodwin's work. In this approach, texts are seen as part of the activity sequence, whether they are inscribed or activated (e.g. Smith & Whalen 1995). The present chapter illustrates a case study in which both activities take place: in the students' effort to enter or create mailing labels with the help of the word processor, they activate the computer manual, the screen, and, more peripherally, the task description and the model sheet, for advice and guidance. It is the latter activity, the 'unscribing' (vs. inscribing), that is of interest here: how do the textual objects, not only as symbols but also as material artefacts, get incorporated into or steer the action. Manuals might be understandable when read as a text in their own right, but difficulties often occur when they are interpreted by user-readers as instructions for action. The 'possible worlds' or 'possible objects' that the textual time and space (e.g. Barthes 1977, Smith & Whalen 1995) create do not necessarily coincide with the interactional time and space of the encounter with 'real world' material objects. Smith and Whalen (1995) criticise the way texts (linguistic signs, numerals, symbols and figures) have been used only as data or experimental props in the human sciences, and their communicative functioning has not been taken seriously. In this chapter, the communicative functions of noninteractive texts (in contrast to the TELEPHONE and

TUTORIAL case studies in which the user was actively engaged) on screen and on paper are analysed in detail. The situated use of the traditional written manual will provide an example of how the writers tend to produce a coherent narrative instead of providing indexical advice for situated action. Thus, the present case study will supply results from yet another circumstance of language use which is aimed at supporting sequential action.

The information that manuals are meant to give to or share with the user could be considered as knowledge which is distributed. The users bring their knowledge to the situation, for instance that of encounters with guide books and machines (the semiotic idea of 'layers of interpretation' growing with people's experience was discussed in Chapter 5). For distributed knowledge to be integrated into human action, sharing and explaining are normally needed. Although the debate about intersubjectivity as sharing the code or 'norms' of language (e.g. Schelgoff 1992, Schiffrin 1990), or intersubjectivity as sharing experiences (e.g. Gumperz 1995) is an interesting one, I assume here that sharing is understood as that whether the participants have common knowledge or not, the status of it is negotiated during an encounter.

Goodwin (1994) describes archaeologists' distributed cognition as two archaeologists collaborate to 'inscribe' events they see in the earth into a category or graph onto paper. In his study of interactions between a more experienced researcher and a student, sharing and explaining also takes place, though the activities surrounding the record taking will not be visible in the written version of notes. A considerable amount of any instructive communication consists of requests which are given to guide the other in their work. And as the requestee is typically a learner, repair of her action is often necessary, which can be done verbally or by the original requester doing the action herself. In comparison with the archaeologist work notes, a manual could be regarded as a record of hypothetical instruction and explanation giving, written for an innumerable amount of ideal readers³⁸. The task of the users of a manual is to 'unscribe' the inscription back into actions. In particular, they have to make a distinction between the authors' descriptions of events on the one hand and requests for action on the other hand: knowledge *that* is accompanied by knowledge *how* (Ryle 1975 [1949]).

In their search for a better theory and method to explain text use, there is an interesting intersection to be found in the new developments combining reading and conversation analytical research. Computer manuals and computer programs, like text processors, are designed to be 'beyond here and now', ie. although they are enunciated by each occasion of their use, they are meant to manage encounters with an unknowable number of anonymous people. The same is true for the computer programs described in the case studies TELEPHONE and TUTORIAL. The predesign of the computer

³⁸ Section 2.1.1. gives a comparison between the basic 'design features' of spoken and written discourse.

programs means that however responsive to the user's actions, they are inescapably beyond here and now. However good the design of the program's interactivity, they are textual products of the past that get to be 'evoked' each time they are used. Although their interactive success can only ultimately be researched by investigating their use, as was done in Chapters 4 and 5, and is carried out below, the interpretative work or communicative practice on each occasion is an example of a 'relay' (cf. Smith and Whalen 1995, referring to M. Goodwin) from the designer(s) to the users; the design of the program, however, sometimes aims at giving the appearance to the prospective user that the computer is a 'participant' in the encounter.

The Word 2.0 screen the users had in front of them in the experiment provided an enabling and constraining work space, which in its appearance guided the users (for more on enables, or affordances, and constraints see e.g. Norman 1988). The initial space for choices was broader than in the tutorial program Learning Word 5.0 (Chapter 5), and most of the constraints appeared as 'nonclickables': when a dialogue box is open, other parts of the screen cannot be worked on. Also, the manual text could reproduce as a separate description what was visible on the computer screen. Sometimes the representation was one-to-one: a picture of the whole screen or a foregrounded part of the screen. Often the manual described the screen indirectly, as a given, by referring to its parts in the instruction for action ("description (...) is always subordinate to narration" (Aarseth 1997, 94)). Normally, only the action-relevant parts are included in manuals, but sometimes warnings are expressed against at that moment affordable alternatives, which would be hazardous ways of undertaking an action.

This chapter sets out to explore how the stable material texts of an instruction sheet (English translation in Appendix 6-1) and a manual (User's Guide. Microsoft Word for Windows™. Word Processing Program Version 2.0) are incorporated into the activity of accomplishing the task of producing mailing labels. From a human-textual object (see Section 2.4.) interaction viewpoint, we are dealing with 'frozen interaction', in the sense that none of the textual artefacts, i.e. the computer, the manual, the instruction sheet, and the exemplary sheet, actively engage the user into interaction with them. This is in contrast to the possibilities for interaction in the cases TELEPHONE and TUTORIAL: in the former the computer was interactively pressuring in 'talk time', and in the latter we found a kind of 'slow motion' interactive time³⁹. The case study examined in more detail below shows how the reconstruction

³⁹ As mentioned in TUTORIAL, The Learning Word 5.0 program was actually programmed to alert the users if they did not do anything with the system for longer than one minute. This never happened in the case study, i.e. the users would stay inside the limits of the predesigned interactive time (the pauses of which were allowed to be considerably longer than in spoken conversations: one minute vs. the putative one second for everyday talk (c.f. Jefferson 1989)).

of requests given by the manual⁴⁰ reproduces the manual text as contributing to the ongoing activity. That this transformation is always part of sequential activity means that the problematisation of accessibility (correct interpretation) of the guiding text should address, among other things, how the potential uncertainties are created by the visual semiotics and how the program has been designed to address the user.

6.2. Background

In this section, the theoretical background for how people use texts in everyday practices is discussed and introduced. The problems of intended meanings and actual interpretations in the case study TASK will be connected to the distinction between knowledge *how* and knowledge *that* (Ryle 1975 (1949)). Any text should be a coherent entity (which could be regarded as knowledge *that*). The authors, however, might have intended some parts of the text to be read as knowledge *that* (information), some as knowledge *how* (guidelines for practical action). Moreover, (and most importantly), the user-readers' local interpretation of text can treat it as knowledge *that* or knowledge *how*.

6.2.1. Affordances and constraints

In this section, the material circumstances as supporting sequential action are discussed from the point of view of what the textual objects make possible and what restrictions they set for using them. In the case study TUTORIAL (Chapter 5), the order of doing things was partly open, i.e. the participants could choose which option to explore from a menu; cf. sections and chapters in a manual. However, some actions could only be done in a certain sequence. For instance, in a subtask of learning, predestined actions had to be taken in an order. The Word 2.0 manual combines the Overview, Steps and Practice modes of the Learning Word 5.0 program in TUTORIAL into one set of instructions, and the Word 2.0 program used on the computer in TASK was not an imitation as was the tutorial program. In the TASK situation, as in any word processor usage, the user was in charge of the content of what she was doing. The result (of subtasks and the whole task) might be a success, but the system would not give any feedback (cf. *Good!* in TUTORIAL).

There are certain interfaces, e.g. installing programs, which only give the next step to the user without an overview of what is going to happen. This resembles face-to-face or 'on-line' instruction giving: the advice is conveyed

⁴⁰ The ultimate request being the instruction sheet which (re)states the aim the students had agreed to achieve.

step by step (cf. Firth, 1998). However, when using a word processor for instance, the user is faced with a collection of 'waiting' signs and icons to be clicked to get something done. In a manual, the order of doing things is usually written down in a numbered list, and the user is thus guided through the 'waiting' icons of the screen, the sequence of actions is 'preorganised' on the page of the manual. This is because though the icons and symbols on the screen would be easy to understand as such, the order of actions realised through them in relation to the task at hand might create problems.

6.2.2. Text as a requester (instigator of user's actions)

In Smith and Whalen (1995), a fairly rapid text-talk-text modification is discussed. The case study concerns emergency phone calls which are transferred to a computerised form, and sent to a dispatcher who then alerts the police/paramedics/fire brigade. Smith and Whalen state: "This organizational sequence is or can be *temporally* discontinuous" (Smith & Whalen 1995, 9). However, the talk-text-talk revision is still produced for a (temporally) relatively close next reader. Their idea of 'regulatory texts' (ibid., 32) is interesting: text can be seen positing demands or requests to the user to 'fill the slots in'. This is true of the everyday text processing programs so widely in use in the computerised world, which could be regarded, if not as requesting, at least as inviting the user to click a word or a button. Smith and Whalen come to the conclusion that the future text (when the empty slots are filled) functions as "a 'third-party' to a conversational sequence" (ibid., 32), shaping the telephone interaction between the caller and the call-taker. They explore what consequences this 'silent participant', the (electronic) format, has for the talk-in-interaction. Though dealing with a different situation of 'text work', Smith and Whalen's expansion of conversation analytic research to texts as significant for the shape of some sequences of interaction gives support to the approach in the present work.

In literary criticism, Aarseth (1997) has introduced the concept of ergodic texts to account for readers' experiences with electronic texts (e.g. hypertexts and games) and other texts which engage the readers to do something explicit in order to progress in the reading. Aarseth gives an account of the discourse planes in narrative (Progression and Event), hypertext (Progression and Event), and electronic textual formats that give a possibility for the user to 'negotiate' with the program, which he calls 'cybertext' (Progression, Negotiation, Event) (Aarseth 1997, 127). In narrative, the planes are not separated from each other (the reader progresses in the reading which unfolds the events for him or her), but in hypertext and cybertext, progression requires the reader's action: in hypertext one clicks a link, and in cybertext one gives commands and responses to queries about the progression of the computer game. Therefore in cybertext, part of the

unfolding event is negotiated with the reader, who can for instance give orders to a 'voice' which 'talks back', either stating the result of the reader's choice or asserting a problem with the reader's order. According to Aarseth's typology then, Learning Word 5.0 (Chapter 5) could be considered a cybertext: the 'voice' gives padding or repairs the user's action.

In Section 2.3.1, Jensen's semiotic ideas of (mass) communication were discussed. In his formulation of interaction⁴¹ the importance of others' power is emphasised in creating action space for, and in general defining, the social agent. If manual texts are seen as (representatives of) 'others', then they should have the same capacity; they should be able to steer the user in the task. The composition of manuals is usually very matter-of-fact, implying a trust in the readers' capacity to follow the descriptions. If something goes wrong, it is the reader's capability in using the manual rather than the manual's sensitivity to the task that usually becomes questioned.

Barthes' view that textual time is that of its enunciation was discussed in Chapter 5. Out of the four case studies examined in this dissertation, the present one, the TASK, is the only one with a manual text which is in a book format and thus materially belongs to a culturally established type of reading objects. However, the manual is very much about giving directions of actions to be taken and therefore engages its reader not only to understand the text at the moment of reading but also to interpret it (cf. Chapter 5) such that the text brings up practical action. As pointed out at the beginning of this chapter, in the use of the computer text of Chapter 5 (TUTORIAL) and the paper and electronic texts of the present chapter, a basic distinction creating uncertainties is when to read (only) and when to act upon reading: learning *that* vs. learning *how*⁴² (through written instructions). For example, manuals give the reader instructions about how to proceed in using a specific feature of a program. However, sometimes the text is meant to be general information,

⁴¹ "It is only through *communication*, however, that the two interpreting subjects engage each other in a social process of semiosis with reference of a common object of interest, thus negotiating the status of different signs to arrive at a degree of (scientific or public) intersubjectivity. In *interaction*, this process is generalized so as to include the status of other subjects on the agenda of semiosis: social agents may redescribe each other - and their purposes and contexts - as both subjects and objects of action, ends and means of society. Others' description of who or what I am, in which contexts of action, implies what I can do." (Jensen 1995, 48.)

"There are certain parallelisms between knowing *how* and knowing *that*, as well as certain divergences. We speak of learning how to play an instrument as well as of learning that something is the case; of finding out how to prune trees as well as of finding out that the Romans had a camp in a certain place; of forgetting how to tie a reef-knot as well as of forgetting that the German for 'knife' is 'Messer'. We can wonder *how* as well as wonder *whether*.

On the other hand we never speak of a person believing or opining *how*, and though it is proper to ask for the grounds or reasons for someone's acceptance of a proposition, this question cannot be asked of someone's skill at cards or prudence in investments" (Ryle 1975 (1949), 28).

and sometimes step-by-step directions, a distinction which does not always work for the user-reader.

6.2.3. Request as part of interaction

Kleifen & Frenz-Belkin (1997) researched a situation in which, due to the different expertise and experience of a working team (a pair), the asymmetrical distribution of interactive knowledge could be analysed. Instructions imply asymmetry because the one in need of guidance does not know, or remember, or trust themselves. Interactivity is the aim also in the manual instructions which are given, for example, by using numbering to mark the pieces of text that should be 'executed' (in the order of the numbering) to differentiate them from the general description.

Roy Harris calls the two type of action sequences that can result from communication 'enactive' and 'assimilative':

A may communicate with B for the very specific purpose of trying to get B to *do* something (e.g. shut the door, answer a question, take the dog for a walk, get married). But A may also communicate with B not in order to get B to do anything in this active sense, but simply in order to inform B of something or to create some impression on B. (The cases are here differentiated by reference to A's intentions, but this is inessential: what counts is the type of sequel B produces.)

If, in the first type of case, B does respond by doing whatever was requested, the sequel may be called an *enactive* sequel. If, in the second type of case, B simply notes the information, is duly impressed, etc., the sequel may be called *assimilative*. But an enactive sequel may be produced even if not overtly 'called for' by A, and similarly an assimilative sequel may be B's only response even though an enactive sequel was explicitly requested. Furthermore, it may be that in neither case is the sequel one that A expected or hoped for.

(Harris 1996, 72).

When Harris' assimilative and enactive sequels follow contrary to the speaker's intent, the case should be considered as partial (or wrong) understanding on the part of the receiver. According to Ryle, no partial knowledge of knowledge *that* is possible, whereas knowledge *how* can appear as "having a particular capacity in a limited degree" (Ryle 1975 (1949), 59), e.g. chess players can vary in their skills. Ryle considers understanding to be about knowledge *how*. Schiffrin's (1990) idea of understanding considering the speaker's/actor's intent could be seen as similar to Ryle's concept of understanding, but Schiffrin gives the receiver the role of them producing their own intent, which she considers to be interpretation. If interpretation can be expressed by action, then Ryle's understanding (knowing *how*) resembles Schiffrin's et al.'s interpretation, and knowing *that* is nearer to how, for

example, Coulter (1994) perceives understanding. (The distinction between understanding and interpretation was discussed in Chapter 5).

When asynchronous, noninteractive instructions are made to work by people using/reading them, they surpass the (difficult) boundary of knowledge *that* into knowledge *how*: information is made to work as illocution. However, instructions are such that the writer knows beforehand that somebody is trying to follow them *in situ*. In other words, there should be recipient design on the part of the writer, for the user indexicalises the knowledge *that* (information) into knowledge *how* (ability/doing). Referring to the philosophical arguments, Prawitz (1990) links knowledge *that* with propositional knowledge and knowledge *how* with practical knowledge, a distinction which resembles that between declarative and procedural knowledge (e.g. Norman 1988, 57-58) and the linguistic debate about reference and practice/indexicality (e.g. Silverstein, 1985). The interpretation worked with in the present chapter challenges also the traditional view and proposes that the distinction between the two knowledges should be based on description and observable practice, rather than knowledge *how* being understood primarily an undescribable innate ability. This is in line with Harris' quote above, and with Silverstein (1985) who argues that the indexicality of any language use foregrounds the pragmatic function as the most important for meaning making. There is an abundant discussion of indices in the semiotic literature, as well. It is claimed (e.g. Greenlee 1973) that Peirce's other famous sign types, icons and symbols can be also regarded as indices. In situated language use and interpretation it could be claimed that any sign that is in the perceivable surroundings of the interlocutor is potentially an index, a trace of somebody or something, ie. a trace from the past that is brought to the user-reader's present to be acted upon. The success and failure of textual requests seems to lie in their relation to the action: are they primarily describing the action (through narrative) rather than paying attention to the text being activated in a specific material and interactive environment?

Recontextualisation as interactive device

When a linguistic or other sign from a textual object is brought into the ongoing interaction verbally, it is recontextualised, moved from the 'textual time' to the 'real time'. This activation or interpretation necessarily renews the sign, bringing in the reader's linguistic and cultural knowledge, his or her layers of experience (cf. Chapter 5). Not only do the stable signs get revised to be a part of the ongoing activity, but also the material surroundings are interpreted on the basis of the sign user's past experience of similar environments. The division into interpretable signs such as symbols, icons and indices on one hand, and material objects such as a keyboard, computer

screen and a mouse on the other hand, may get blurred in the ongoing activity. Streeck (1996, 366) has shown how things can become part of the meaning which “does not only flow through symbols and expressive forms that form our ancient and primary tools for communication, and it is not just “contextualized” by the material environment; rather, the environment, through the interpretive use the participants make of it in their situated activities, becomes a *component* of the process of communication.” According to Streeck, this shows the human disposition toward symbolisation (which happens through material objects becoming first indexical signs) and he relates this new use of a thing or a word to metaphor (ibid., 367). He reminds the reader that language is a *material artefact*, too, “which originates from local inventions in fleeting moments of face-to-face communication” (ibid., 382). Thus, not only linguistic signs get transformed from one context to another, but reconstruction can also take place from objects to symbols. In my data, there is evidence of the reverse movement as well: the linguistic signs (as symbols) on the computer screen tend to be objectified when paraphrased in the ongoing talk. One of the dangers of computerised communication environments might lie in this tendency of users to treat language ‘clinically’, using it as a gateway for creating an effect rather than a device for (co)constructing and negotiating meanings.

Heritage and Watson regard formulations (paraphrases, repeats) as “important methods used by members for demonstrating that, among others, the conversation has been and is ongoingly self-explicating” (Heritage and Watson 1979, 123), ie. formulations are exhibits of understanding. When pieces of text are quoted by a human in a task situation, such as can be found in the present chapter’s data, the repetition is geared towards the other human participant, and the piece of text could be called, in semiotic terms, another interpretant in the human-human interaction, i.e. it is another turn in the ongoing interaction. At the same time, however, the human mediator who quotes the text ‘nonmarkedly’ (without stressing any part of the text or without paraphrasing it), makes the text a direct ‘participant’ in the ongoing interaction and action. In a pairwork situation, the text thus becomes a ‘third participant’, but in a more straightforward way in comparison to Smith and Whalen’s (1995) data. In their case study, the ‘third participant’ is a silent or only visual party to whose insistent slots the call-taker gets an answer from the caller, and the abbreviations of which the dispatcher paraphrases into English language when she transmits the information by radio waves to the party responsible to take the action needed. When a manual text is quoted in the use of the manual, the quotation often works as if the reading aloud is of what ‘the manual has to say’ at that point, though of course it is the users who decide what the manual gets to ‘say’ in the interaction.

Paraphrases are more marked as exhibits of understanding than, for instance, the fact that every turn is at the same time an analysis of the previous one and a next step in the interaction (cf. Heritage & Watson 1979).

Paraphrases give the original speaker a more direct possibility to check the interpretation. Repetitions are nearer the nonquestionable understanding displays, the so called feedback markers, e.g. *mm-m*, which do not give the 'first speaker' the possibility to check what the recipient's understanding is.

6.3. The case study TASK

The case study of this chapter was one of the many experiments that take place in university or other institutional surroundings (e.g. Frohlich et al. 1994). Pairs of students were given a task to do with Word 2.0, and both their activity on the screen display and the users themselves were videotaped. All the subjects were students of English; they had been given short instructions on a piece of paper (Appendix 6-1), a list of names and addresses for the new mailing labels, and an exemplary sheet of the mailing labels they should be producing. Also, they had a Word 2.0 manual in front of them, opened at the page where mailing labels are explained. All in all, six pairs of students took part in the experiment, and four recordings were considered good enough (i.e. the participants would proceed in the task) to be transcribed. In the present chapter, one of the pairwork tasks is analysed. However, when considered necessary, some aspects of the other three interactions are going to be discussed for comparison's sake.

In the cases TELEPHONE and TUTORIAL, the user-reader's participation was controlled and channelled by the system (the recipient design was active). Therefore, they could go under the rubric 'new' communication environments. Using a printed manual is the 'old way' of doing things; it engages the reader to be an active participant through written directions. In all the cases, texts are guiding the users' actions, but printed texts cannot detect the users' doings at all. However, the computer is becoming a 'given' as an artefact (cf. the use of mouse in Chapter 5). The participants in TASK had used computers before, so the human-computer interaction situation was not new as such, but whereas Word 2.0 as a system was familiar, the mailing label function was not known (though one had experience from a similar function in WordPerfect). Thus, the participants had differing 'layers of experience/meaning' on which to build new layers.

6.3.1. Analysis of data

In the following, I shall concentrate on formulations (paraphrasing and repeating) in the human participants' talk, especially on those instances in which the text in the manual or on the screen is 'read aloud'. This quoting can happen with or without specific reference to the 'quotee', i.e. the manual. As part of the ongoing interaction, the quotes of the manual indexicalise the text,

which is the ultimate test for the success of text as a pragmatic tool. I also make comments on the interaction in general, in which the textual objects (the manual, the instruction sheets, and the screen) are addressed or otherwise incorporated in the ongoing activity. Thus, there are three types of interactions or participation frameworks in the pairwork situation: 1) between the human participants, 2) between the users and the computer, and 3) between the readers and the instructive texts. It is possible to compare how formulations of textual materials are done in each participation framework, and thus draw conclusions about the nature of the textual 'other'.

According to Heritage and Watson (1979), formulations are an important means of exhibiting understanding of the other participants' talk. This could be expanded to human-text interaction, as well. When texts are quoted or paraphrased, the participants show to each other that and how they understand the text, and, by doing this, give the floor to the textual 'other' in the action sequence. Often the quoting is done directly, without referring to the manual as the source. This is of course easy to do as the textual object quoted is available for the participants, i.e. it is indexically present; the origo, the source of the words, does not have to be established through language use.

In the data analysis below, a special emphasis is placed on (what constitutes) formulations and how these paraphrases or repeats relate to the various interactions going on: human-computer (screen text), human-human, and human-(manual) text.

Participation frameworks: Human-human, human-computer(text) and human-manual(text) interaction

In Example 6(1)a below, in which two students are using the manual to check the first procedure of making mailing labels, we find a minimally transformed quote (line 6). Quotes tend to be accompanied by gestural pointings at the text (line 3), which visually strengthens the cooperative stance of the pairwork, but at the same time gives the other participant a possibility to challenge that the pointed at section of text on the page is a relevant next in the situation. (The transcription conventions can be found in Appendix 1-1. In the transcription, the English translations are placed in each turn such that they intervene as little as possible with the markings of overlapping speech and action.)

6(1)a

- | | | |
|---|----|--|
| 1 | | (((A & B read the page in silence))) |
| 2 | | [(5)] |
| 3 | A: | ((puts [left index finger next to text])) |
| 4 | | |
| 5 | B: | [mm-m ,]= |
| 6 | A: | = "choose new" (((right hand reaching to mouse))) |

7 B: [(**joo**/.yeah)]
 8 A: ((gaze to screen, arrow to option File, glance to the mouse))
 9
 10 [(**no!**/c'mon!)]
 11 [((clicks File))] ((clicks New))
 12 (**no ni**/okay) ((gaze to the new box that appeared on the screen,
 13 head down and up))
 14 (**a**)**ha!**
 15 [**tuo!**/that one!]
 16 [((points the arrow to MAILLABL))]
 17 >**ha**[**ha**<>**h**]**aha**<
 18 B: [**mm-m**]

The extract begins with a 'normal' reading scenario, or human-text interaction: the students are reading a page to themselves in silence. B's *mm-m* in line 5 not only exhibits her understanding of what she has just read (and therefore her orientation to the text as information, as knowledge *that*), but also as accepting A's pointing as a suggestion about a piece of text that is relevant for action (before this they had read something on the previous page and noticed that *se on vain jotain yleistä* 'it is just something general'). A's direct quote in line 6 treats the foregrounded piece of manual text as (knowledge or learning) *how* and she interprets the text by recontextualising it in the quote: the *thatness*, the textual time of the text is activated into knowledge/learning *how*, to function as a guide for action at this very moment in this particular situation. In this extract, we can see McHoul's idea of understanding (as interpretation) is at work: he claims that "we can say that understanding is a question of knowing **how** to do something or bring something about rather than a question of knowing **that** something exists or has such-and-such properties" (McHoul 1997). The transformation from text to speech does not change the quoted text: it is just read aloud in a matter of fact way without stressing any (part of a) word, for example. Since the participants have the textual object or the 'speaker' in front of them, there is no need to mention who or what is quoted.

In line 7, B agrees with A's interpretation about choosing *New* (the whole printed sentence was *From the File menu, choose New (ALT, F, N)*) being the first direction for them to initiate in the human-computer interface/interaction (*joo* 'yeah' in line 7). The quote in line 6 is now part of the ongoing interaction, which together with the pointing are treated as a suggestion by B.

Elsewhere, I have argued that "manual text is used as such as an interpretant, which makes it more economical: participants quote it there and then, without explanation, making the interpretant an important part for the unfolding situation" (Raudaskoski 1997, 541). The Peircean concept of interpretant allows for the flux in sign meaning (see Section 2.3); when combined with the conversation analytical idea of turns at talk (or action) being analyses of the previous talk, the Peircean internal interpretant can be also viewed as external conversational interpretation (see Section 2.3.1). Also,

the Halliday-Hasanian endophoric (or textual) and exophoric (or situational) relations are now one: *choose new* as a piece of text from the manual uttered in the course of action does not refer to anywhere in the manual text, nor does it refer to some hypothetical situational relation 'in the real world', but is actually used in a concerted manner in the unfolding practical activity.

The idea of economy can be also found in Heritage and Watson's article on formulations as conversational objects (1979). As already mentioned above, by formulations Heritage and Watson mean repetitions and paraphrases of earlier conversational materials. They see the economy of formulations to be in the effectiveness of checking the topic or gist of conversation, i.e. what the conversation is or has been about. Out of the points that Heritage and Watson list as topical in the management of conversations, the following have a bearing when importing the instructive text into an activity, or shifting between the human-text participation framework and that of the human-computer discussed above: 1) "Where understandings of gist are used to warrant the construction of some implicativeness, formulations of gist may be established as prefaces to or as part and parcel of the construction of some analysis of such implicativeness" (Heritage & Watson 1979, 150). And, 2) "Where the gist of some section of talk involves matters regarded as being of special import or significance, formulations may be used (a) to reattend to such issues, and (b) to mark such gists" (ibid., 150). Further, 3) "Where the trajectory of some section of talk is directed to the achievement of some practical end, the provision of a formulation may work to reestablish the collaborative achievement of that end both as a first topic and as an outcome of the conversation's course" (ibid., 151). In quoting the manual text, the participants are foregrounding this specific piece and implying that it is a relevant next (cf. points 1 and 2) in the activity, which is an outcome of the reading's course, now introduced as a first topic (cf. point 3). Heritage and Watson regard formulations as an efficient and nondisruptive way of checking understandings and doing some other work, e.g. establishing implicativeness. This is exactly what *choose new* does in line 6 in the above example: the direct quote "minimizes the period spent checking a reading without any loss of efficiency in that checking operation" (Heritage & Watson 1979, 152). To be more precise, in line 7 B confirms a formulation that implies a selection of the first step in the practical activity, i.e. suggests that they comply to the request *choose new*. B's confirmation coincides with A's reaching the mouse (line 6), a movement which suggests A's comfort with her judgement, and at the same time with her interpretation what *choose new* means in practice.

In lines from 8 to 11 A chooses *New* under *File*, i.e. she does what they had agreed would be the next step. In this work, she interacts with the computer as a tool (*no*, 'c'mon' in line 10 to comment on the difficulty of opening the *File* menu; it is a comment on human-mouse interaction). In line 11, is the 'final transformation' of the manual text, via the human-human

interaction (lines 6 and 7) into action in the world. The remark on line 12, *no ni* 'okay', marks the end of the subtask and also states the clarification of the problem with the subtask. From line 14 onwards, the users are facing the screen 'on their own', without the manual's directions. They encounter a new textual format which appeared on the screen as a result of the mouse clicks. They look at the new screen, informed by the task description: they are to produce mailing labels (their general knowledge *that*). The text on the screen in a box is objectified, out of the list of items (words), one is recognised (line 14) as relevant: *tuo* 'that' (line 15). Note that it is not quoted, and it is pointed at with a mediating device (the arrow operated by the mouse).

When the manual text content is first and foremost information about doing, the (pieces of) text produced by the text processing program on the screen are almost always invitations to choose, and, by choosing, to do something. The predesigned words on the screen mix *that* and *how* for the reader: they give the user an idea what they might be requesting from the computer if clicking that word (e.g. *Help*), and being available as clickables, they are at the same time telling the user how to proceed in doing something (i.e. they can be seen either as demanding the user to do something or offering the user a possibility to do something, cf. Kress & van Leeuwen 1996). A's words and action in lines 15 to 16 are a case in point: she points at the clickable word both verbally and by moving the arrow to it: this is the how to do next (or rather her suggestion, as B's approving *mm-m* in line 18 shows). A's laughter in line 17 suggests an orientation to the situation as if a competition: A and B have undertaken to finish a difficult task with a computer, and A has been able to propose something without the help of the 'expert', the manual. At the same time, the high tempo laughter could be seen as orienting to the fellow student as a 'competitor' in the 'game of using the word processor'.

The human participants have to manage their own shifting participation frameworks, ranging from interacting with the other participant to managing the human-computer interaction and human-text interaction. All the participation frameworks are of course interrelated, because they deal with the same topic, and most of the time their mutual interaction concerns an ongoing human-text or human-computer interaction. For instance, B's *mm-m* in line 5 shows to the other participant that she understands what she is reading but at the same time it is a reaction to the text as information. In this as in any 'focused gathering' (Goffman, 1971) M. Goodwin's remark is appropriate: "human interactants continually display to each other, in the course of interaction, their own understanding of what they are doing" (1990, 1).

Extract 6(1)b below continues from where the preceding one ended, as indicated by the line numbering. A does not want to click the selected item before checking with the manual (lines 20 to 24). For a moment, A's left hand finger moves away from the relevant place in the text in front of them (line

20), to come back next to the text (line 24) when she verbalises the checking in line 23. A and B are looking for a justification for A's choosing the *MAILLABL* as the relevant next in their interaction with the text processing program. However, there seems to be a problem: in line 31, A reacts to what she is reading with a stressed *mitä?* 'what?'. When *mm-m* in the previous example exhibited B's understanding of the text as information, A's surprised response arises from her orienting to the text as an *a posteriori* instruction about her knowledge *how*, about her skill at choosing the correct next step which is now brought under reconsideration. Thus, examining the quality of the users' responses, can give information about how "an instruction's significance with respect to action does not inhere in the instruction" (Suchman 1987, 61).

6(1)b

- 19 A: [.h:]
 20 [((gaze to text; left index finger away from text; right hand away
 21 from mouse))]
 22
 23 [onks se nyt sitten/ is it now then-]
 24 [((left index finger next to text on page))]
 25 |
 26 B: [((hand onto mouse; A & B read the page in silence))]
 27 [(2)]
 28 A: th:
 29 [((A & B read the page in silence))]
 30 [(5)]
 31 A: m:itä?/w:hat?
 32 B: "Do not use (the) toolbar to open a new file"=
 33 A: ="file because you need to use the temp[°lates .option°].
 34 B: [(h:)]
 35 A: erm ((gaze up to screen, moves back, left index finger next to text))
 36 oliko tuo nyt sitte se tool bar? (.) vissiinki./ was that then the tool
 37 bar? (.) i guess so.
 38 [eikä ku hetkinen °onks se°-/no wait a minute °is it°-]
 39 [((left index finger up from page towards screen))]
 40 B: tä[mä on toolbar ((left index finger across the toolbar)) joo.]/
 41 thi[s is the toolbar. yeah.]
 42 A: [tämä on se toolbar ((pointing on the screen)). niin onki./
 43 this is the toolbar that's right.]
 44 B: .joo/.yeah
 45 A: [elikkä ihan oikein/so quite right.]
 46 [((gaze to text))]
 47
 48 B: [joo./yeah.]
 49 [((clicks MAILLABL))]
 50 A: .joo/.yeah ((gaze to screen))
 51 B: [tuo, /that one,]
 52 [((moves arrow to OK))]
 53 |
 54 A: [joo-o, /ye-ah,]
 55 B: ((clicks OK -> box disappears from screen and hourglass appears))

The two students are reading a manual to get their task done. The text in the manual often refers to the computer screen in front of the users, as in this case. The students had successfully pressed the icon-like verbal signs *File*, *Open* on the upper left corner of the screen, when they next came across the problematic sentence repeated in lines 32 and 33. *Toolbar* refers to the row of the iconic pictorial representations of a set of commands as an alternative to the icon-like words like *File*. The dialogue boxes that clicking the pictorial or the verbal icon open differ, however. In the case of opening a file, the latter shows the user the 'templates' option mentioned in the quote. The example draws attention to two representations of actions, one linguistic, one iconic, both of them visual and situated in the upper part of the screen⁴³. However, the directive does not make the connection between the 'templates' option and the verbal icon clear. The interpretation of a horizontal positioning of words on the screen contributed to understanding the referent of that entity to be *Toolbar*: the virtual, symbolic language and its material form are intertwined in the interpretation process.

When A in line 31 expresses that she sees a problem, she still has the left index finger next to the text. And, as the participants had read the first piece of instruction first (1. *From the File menu, choose New (ALT, F, N).*), the next unread item would be beneath it: *Do not use the Toolbar to open the new file, because you need to use the Templates option.* However, though these two sentences go together, the latter being informative or an expansion on the previous instruction, addressing the hypothetical reader in a hypothetical situation, it is encountered by A who is looking for an answer to her query expressed in line 23. In her direct quote in line 32, B exhibits what or at least where in the text the problem was, i.e. she repeats the trouble source, and, at the same time, suggests her answer to the problem. The next step in the human-text interaction (about toolbar and opening a new file) did not concern the next move in the human-computer interaction (to click *MAILLABL*): the two participation frameworks were out of pace, which might explain A's strong reaction. The formulation of the *Do not use the Toolbar...* sentence was such that it could have been a repair initiator. In a way, B's reading aloud the sentence is a repetition of it because it was made relevant (finger next to text in line 24) and problematic (line 31) by A. In fact, B's reading changes a definite article (*the new file*) into an indefinite one (*a new file*), and thus misses the link between the present sentence and the previous one, which might be the gist of the problem: the sentence reproduces the *choose New* of the first instruction as *open the new file*, which is a completely different formulation. 'Opening' in PC usage many times involves clicking an item on the screen, and therefore the sentence could potentially be about clicking *MAILLABL*. However, in line 36, A shows that she understands that the sentence concerns

⁴³ It is noteworthy how inconsistent the linguistic depictions are: the sign sometimes refers to the object of the commands (*File*, *Tools*, *Table*, *Window*), sometimes to the action (*Edit*, *View*, *Insert*, *Format*, *Help*).

her previous activity and therefore is a potential repair initiator. There is a good reason to suspect this, as the first part of the directions gave instructions to use the Alt key and keyboard (ALT, F, N) to choose new, whereas A used the mouse to click *File* and *New*. A infers that her mouse clicks must have been done on a toolbar (line 32) (note that the word *Tools* is on the same list of adjacent words), but she quickly starts doubting her interpretation (line 38). B jumps in (line 40), cooperatively finishing A's sentence which agrees with A's turn⁴⁴. *Toolbar* is kept in its English version through the Finnish exchange: when establishing the referent of the word, the users keep the signifier the same, while the signified changes. There is very little translation in general from the screen and manual English to Finnish, which helps to keep the human-human interaction (in Finnish) separate from the two other, English speaking domains: the program and the manual.

A's statement in line 45 *elikkä ihan oikein* 'so quite right' and going back to text (line 46) close up the side sequence that the *Do not use the Toolbar* sentence caused. This is a prompt for B to carry out what A had already proposed in the previous extract, lines 15 to 16: B clicks to choose the *MAILLABL*. And, with the same words and actions that A had moved the arrow to *MAILLABL* in lines 15 and 16, B moves the arrow to the OK button: *tuo* 'that one' (line 51) "makes her moving of the arrow to the OK button an interpretant for the other also via language" (Raudaskoski 1997, 543). A quickly approves of B's previous turn or suggestion (line 54), and B finishes off the encounter with the dialogue box (line 55).

The interpretation of the manual text and the unfolding human-computer interaction is in its most concrete form at points when something is done with the computer. Verbalisations which can take a form of deictic pronouns like *tuo* 'that (one)' seems to guarantee the verbalised approval of the other participant (which is important since in 'ear-to-ear' situations head nods are not easily perceived). For instance, in the first part of the extract now under scrutiny, there is no feedback when A only uses the mouse (e.g. line 11), but she does get the approval in line 18 for her next action on the screen, which is accompanied by *tuo* (lines 15 and 16). The format of the deictic item used reveals the distance that there seems to be to the screen (*tuo* 'that' vs. *tämä* 'this') though they are very close to it and operate on it (but via a mediating mouse).

Indexicality is not only created by the practical situation, but some linguistic features in the manual text are geared towards highlighting an indexical, situation bound, reading. A's *mitä?* 'what?' in line 31 is a clear

⁴⁴ This resembles the use of a map (as a description) to navigate a surroundings: "Particularly where some descriptive ambiguity is found on the map, actually locating a point may serve to disambiguate that ambiguity. Thus the map-as-used may be said to exhibit reflexive properties in that it describes (e.g. 'foregrounds') various points *en route* to a destination but is, in turn, described (specified, revised, etc.) by those points as they are found" (Watson 1997, 95). In the present case, the meaning of *Toolbar* in the manual is specified by finding the referent on the screen.

indication that what she has just read is about their ongoing activity with the computer there and then. One reason for this might be the use of definite articles (*the Toolbar, the new file*) in the text. According to Hanks, the definite article (in face-to-face or shared time and space situations) is one sign of indexicality (Hanks 1996, 205). In the case of the *Do not use the Toolbar to open the new file* sentence, the writers might have envisaged the user doing the task and therefore given a warning in this 'possible specific world' instead of giving the warning as a more general description (e.g. *Do not use the Toolbar to open a new file*; actually this is how B, who does not seem to be as surprised, literally reads the sentence (aloud)). In using the definite articles, the manual writers were coherent: when they had described how to open a new file in the previous sentence, they now gave further directions about opening the (same) file. In the textual time of the manual text, the first sentence (*From the File menu, choose New (ALT, F, N)*) creates a new data file. However, *Toolbar* is mentioned for the first time in this set of instructions. According to a guide to English grammar, "the definite article presupposes an earlier mention of the item so determined. But in actual usage the relation between presupposition and the definite article may be much less overt" (Quirk & Greenbaum 1988, 72). The toolbar had been introduced at the beginning of the manual, so strictly speaking it had been mentioned earlier. However, in the chapter about mailing labels, this was the first mention, though it was expressed as free information (which can be taken for granted) and not as bound information (when new information is introduced) (Blakar 1992, 244). Therefore, there is good reason to assume that the connection was contextual rather than textual: there is something called *the Toolbar* in the working environment which is as unique as the mouse, for instance. The extract shows that textual coherence and explicitness cannot guarantee interactional success: "what a priori has been considered precise language (e.g., manuals) might turn out to be problematic in the unfolding situation" (Raudaskoski 1997, 537.)

The manual as a textual object

There are variations of how the manual is used/produced as a material object.

6(2)

- | | | |
|----|----|--|
| 1 | | (10) |
| 2 | | ((A & B reading the manual)) |
| 3 | A: | "to print one label (.) multiple lab ^{els} "= |
| 4 | B: | =((left index finger to text, following it when reading aloud)) .h nii |
| 5 | | kato tässä on/you see here is (.) "for subsequent (.) mailing labels |
| 6 | | (.) printings (.) when the main document and data file are |
| 7 | | al[ready set up] |
| 8 | A: | ["already set UP"] |
| 9 | B: | (.) you only need to use the third procedure (.) to print mailing |
| 10 | | labels" .h no mutta oikeastaan/well but actually (.) |

11 [((left index finger down left and right page))]
 12 [tää on ihan oik[ein]/this is quite [right]]
 13 A: [nii,]/ [yeh,]
 14 B: **me mennään ihan oikeassa** /we are going in quite right
 15 **järjestykses[sä (se) on varmaan täällä]** /orde[r (it) must be here]
 16 [((turns page with left hand))]
 17 |
 18 A: [(**niin minunkin mielestä** /I think so too)]
 19 B: (s-) **joo [to print on täällä]**/(yeah[to print is here]
 20 [((right index finger to sweep the page))]
 21 |
 22 A: [no niin. **justiisa.**/okay. quite right.]

A's first quoting in line 3 is done by leaving pieces of text unread and foregrounding in the citation only the two alternatives: printing one label or multiple labels. The quoted text is at the bottom of the page, whereas B's selection can be found at the top of the page. She establishes the place of the text by pointing and says *nii kato tässä on* '(oh yeah) you see here is', thus mentioning the source of the quote, but formulates the place of the quote (*tässä on* '(in) here is') rather than the source of it (e.g. "the manual says here that"). As the source of the quote is visually perceivable for both, there is no need to mention it again. To use Finnish to establish the place of the quote and then English in the actual quote highlights the fact that it is the manual that speaks, B is not giving her own interpretation at the level of the content, though, by quoting the text, she orients to it as a relevant next.

In contrast to the manual text, in TUTORIAL (Chapter 5) the computer (screen/program) is often referred to as an 'it'. Therefore the textual object of a manual is oriented to spatially rather than as a communicating entity: the piece of text quoted *is* on a page, rather than the manual saying something. The human-manual text interaction is done in accordance with the human-computer interaction. However, the feeling of the manual text as a spatial object that can be explored out of phase from the human-computer interaction is emphasised in lines 15 to 20 when B 'peeps ahead' from the present human-manual text activity to see where they should be able to find more information about printing. This reflects the 'autonomous destiny of writing': "Through typography in particular, speech abandons the confines of time to enter those of space: paragraphs, spacing, chapters, capitals, titles, subtitles. Language becomes a two-dimensional object on the page and a three-dimensional one in volume." (Hagège 1990, 62).

Recontextualising from the manual and from the screen

Smith and Whalen point out how in emergency call relays, the dispatchers who give the description to the officials in question expand in speech the shortened official version they see on the screen (Smith & Whalen 1995, 28). In my data, there is an interesting distinction in the way the users talk about or

integrate the manual text and the screen words into their speech. The language on the screen tends to be singular words on a visual metaphor of a working surface, resulting in the language being objectified as a surface tool, even when the pieces of text behave ‘interactionally’ (i.e. as dialogue boxes that can not be passed without clicking one of the options). The manual, however, provides much more information about the context of action (and in full sentences), so the language is never objectified, though in finding relevant pieces of text, the manual is oriented to as a three dimensional object.

6(3)

- 1 (13) ((A reads the manual, B reads the instruction sheet))
2 A: [mm. hetkinen mitähän tä^ossä (on)^o?/wait a minute, I wonder
3 what ^o (we have) here^o?]
4 B: [((puts the instruction sheet away and starts reading the manual
5 page))]
6 (13) ((A uses her right thumb to follow the text she is reading at the
7 bottom of the page))
8 A: [^ohm,^o]=
9 [((gaze to screen))]
10 B: ((right index finger to text at the bottom of the page))
11 .h this means that it is possible you know/
12 .h tää tarkotta[a sitä] että voi niinku=
13 A: [mm-m]
14 B: ((right index finger pointing at text at the top of the page))
15 =to choose what /
16 =[va]lita että [mitä]
17 A: [joo] [näin on/that's right]
18 B: =haluaa noista ottaa ^oettä^o/ one wants to take of those ^oso^o
19 A: joo
20 (1)
21 B: should we just try and reply that yes/
22 kokkeillaanko vaan vastata et[tä jes].
23 A: [että jes/that yes]
24 [ku se o(h)o(h)ttaa sitä kerran/because it is w(h)ait(h)ing for it]
25 [((finger pointing at screen, glance at B))]
26 B: ((clicks the Yes button; new box appears. A & B look at screen))
27 A: joo. vastataan jees. no nii./yeah. let's answer yes. okay.
28 (2) ((new dialogue box appears))
29 A: mhm. jaHA/ mhm. I SEE
30 B: [katoppas nyt mitä(pä)s tää (llä)/let's see now what () he(re)]?
31 [((right index finger to point at manual text))]
32 no tässä on/well here is
33 "Word dis[plays the attach data]= ((moves finger near line))
34 A: [(lay the attach da)] ((creaky voice, moves finger on line))
35 B: [=file=]
36 |
37 A: [joo./yeah]
38 B: [=dialog box"] elikkä/in other words [tämä on-/this is] (1)
39 [((gaze to screen))] [((points at screen))]
40 |
41 A: [((gaze to screen))]
42 B: onko tämä oikia,=/is this the correct one,

43 =((A & B: gaze to manual))
44 A: "Header file" ((whispering))
45 (2) ((both reading the manual))
46 B: **oisko se sitteki ollu toisin päin?**/
47 should it have been the other way round, anyway?
48 ((right little finger to text))
49 **(jos)- tää näyttää nyt "attach data [file" (jos) se ois nä-]/**
50 **(if)- this displays now "attach data [file" (if) it would have dis-]**
51 [(gaze to screen, rapid vertical
52 movement of right middle finger at
53 screen))]
54 **[jos se ois sanonu vaan-/ if it had said only-]**
55 **[(rapid vertical movement of right index finger at the screen))]**
56 **that only solely those- °the other one°/**
57 **että [vaan niin] pelkästään vaan noi- °sen toisen°**
58 A: **[niin,/right]** ((still reading the manual))
59 (7) ((A reading manual, B looking at screen))
60 B: ((right hand onto mouse))
61 **[.hh] °päästäsköhän me=/'I wonder if we could get**
62 |
63 A: **[(gaze to screen))]**
64 B: **[=känselillä°-/ with cancel° ((clicks Cancel))]**
65 A: **[panepa(s) kansselia] siihen nytte niin katotaan/**
66 **why don't you put (some) cancel there (into it) now so we'll see]**
67 B: **['the macro has been interrup- ()']**
68 **[(moves the arrow on the lines))] ((clicks OK))**
69 **[(sighs))]**
70 |
71 A: **['ookoo'./'o k'.]**
72 B: **[no niin./okay. ((↓↑ File New)) alu(h)sta/ from the beg(h)inning]**
73 A: **[alu- alusta!/ from the be- from the beginning!]**
74
75 B: **[(clicks MAILABL))] [(clicks OK))]**
76 A: **[siitä ja-/there and-] [no-in,/ so-o,]**
77
78 B: **ja laaser./ and laser**
79 **[(clicks Laser))]**
80 A: **[joo,/yeah,] [ja siihen okkia,/and ok to there,]**
81 |
82 B: **[ja oo k(oo)/and o k]**
83 **[(clicks OK))]**

A detailed analysis of this fairly long extract reveals some interesting aspects of the different interactions going on in the situation. For example, the users are perplexed about the 'yes/no question' posed by the program, reproduced in Figure 6-1.

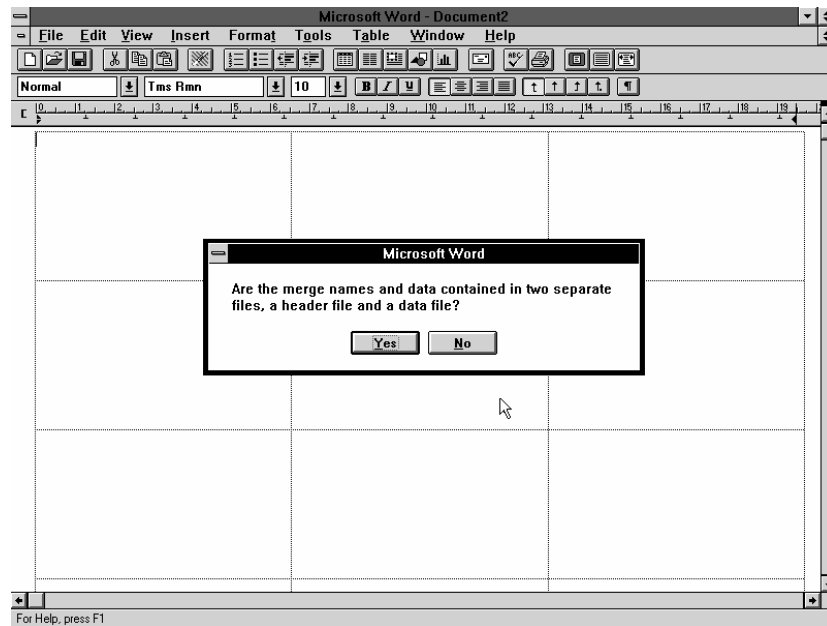


Figure 6-1 The screen at the beginning of Extract 6(3)

A and B resort to A-manual and B-instruction sheet interaction to solve the problem of 'what to answer to that then' (*mitä tuohon nyt sitten vastataan*) as formulated by A when the question box first came up. Spatial deictic markers are used to refer to the text in the manual (*tässä* in line 2, translation in lines 2 and 3: 'what we have here', or word-for-word 'what is here').

A's general locative *tässä* 'here' is transformed into *tää* 'this' in line 12 when B gives her 'translation' of what a piece of text on the page is about. Her accompanying pointing gesture specifies and singles out the piece of text she is talking about. She refers to a visual depiction of data slots on the page as 'those' (vs. 'these') (line 18), making thus a distinction between the printed text and a figure, out of which the text is nearer to her on the bottom of the page. B's explanation or recontextualisation of the text is given in Finnish, which is quite rare in the situation. It is understandable in this case because it serves as a summary, as a piece of information (knowledge *that*), a type of aside, rather than being an integral part of an ongoing practical activity.

A different type of movement happens in line 22 where B suggests what their answer to the long-waiting 'yes/no question' on the screen should be. The program could wait for the reply forever, as it is not designed to prompt the users or ask why the answer is delayed: the screen insists that the users answer the question but it allows for the answerers to take their time. In line 22, B uses the English word *yes* instead of the Finnish 'kyllä' when she suggests they answer positively to the question (*kokkeillaanko vaan vastata että jes/'should we just try and reply that yes'*). A finishes B's turn cooperatively, and also she uses *yes* in her overlapping turn (line 23). So, both A and B use the same formulation as they see on the screen, *yes*. In line 24, A chuckles, making a remark referring to 'yes' being the default answer (because the Yes

button was preselected by the program). This observation as an account of or reasoning for a 'yes' answer also highlights their orientation to the visual metaphor of buttons used by the program. The preselected answer draws attention to *Yes* as a label of a button, though they are still talking about answering or replying to a request or a yes/no question.

The following dialogue box, given in Figure 6-2, is perplexing as well (line 28).

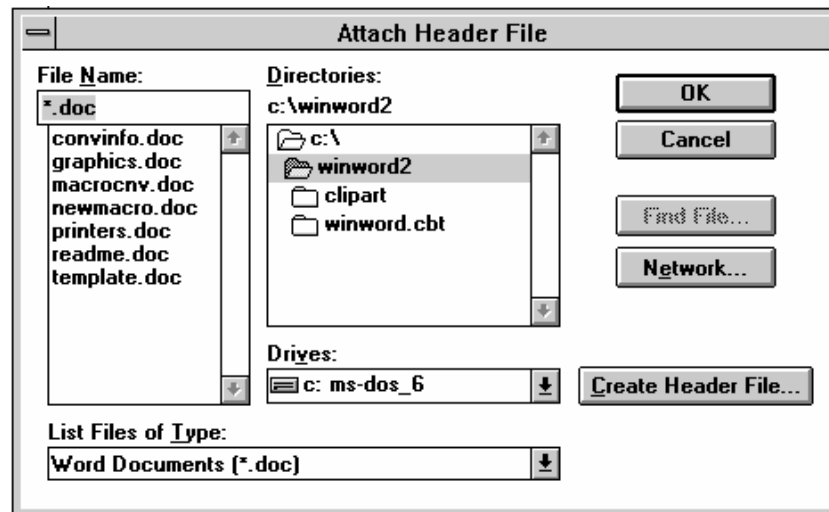


Figure 6-2 Dialogue box (appearing on the screen in line 28 in Extract 6(3))

B resorts to interacting with the manual at this difficult point, and again uses the spatial pronoun *täällä*, '(around) here' when she turns to the manual (line 30), and *tässä* '(in) here' when she spots the line she's going to quote (line 32). Both are used with the verb *on* 'is' (vs. e.g. *sanoo* 'says'), strengthening the feeling of space. B's reading of the quote is accompanied by A's almost unheard recital of the same line (lines 33-38). In line 37 A displays her understanding of the information given in the sentence (knowledge *that*). B finishes the reading in line 38, reciting the last two words (*dialog box*) when turning to see the screen; she moves from human-text interaction to human-computer interaction, a shift from understanding the manual text to interpreting whether the information matches the phase of their ongoing human-computer interaction (knowledge *how*). This transformation is marked by *elikkä* 'in other words' (line 38), followed by a gestural (line 39) and a verbal (*tämä on* 'this is', line 38) indexical pointing at the screen. In line 42, B's statement is self-repaired into a question about the correctness of *tämä* 'this', ie. whether the dialogue box in front of them is what it should be: the problem is not how to 'fill in the form' of various slots displayed by the program but whether their interaction with the program has resulted in it posing a pertinent 'question' at this point.

Immediately after B's questioning remark, both turn to the manual for an answer (line 43), A quoting to herself the gist of the problem or one of the

topics: *header file* (the other option being data file). In line 46, B starts repairing their earlier actions, ie. that they should have answered to the earlier yes/no question 'the other way around' (lines 46 to 47). She accounts for her remark by marking a piece of text in the manual by pointing to it (line 48) and talking about it as *tää näyttää nyt* 'this displays now' (lines 49 ad 50). It is an interesting way of referring to the text about what Word does if the instructions are followed (*After you've answered the final prompt, Word displays the Attach Data File dialog box.*) as if she was not talking about the text depicting what happens but about a screen on which she can see the *Attach Data File* dialogue box. The formulation is no doubt prompted by the word *display* which appears translated (*näyttää*) in her Finnish sentence. When she moves to talk about a hypothetical state of the box visible on the screen, she starts with the same formulation (line 49) but changes it into *sanonut* 'said' (line 54), changing the manual's way of referring to the screen as displaying something into an 'it' that 'says' something. At the end of her turn, B self-repairs her reference to the hypothetical screen from the indexical *noi* 'those' to a nonindexical *sen toisen* 'the other one' (lines 56 and 57). She does not combine pointing at the screen at all to this line, which emphasises that she is not talking about anything tangible in the situation. A gives her first feedback in line 58, and it is noteworthy that she has not lifted her head from the manual ever since B first marked the place in the text with her finger and her deictic *tää* 'this' (lines 48-49). Even if in lines 52 and 55 B uses pointing fingers to aid her talk about the screen, A never looks up, maybe because B does not mark her speech with deictic expressions (*jos se ois* 'if it had' in line 49, and 54; *sen toisen* 'the other one' in line 57). A only lifts her gaze to the screen after B puts her right hand onto the mouse, indicating that she is about to do something.

Again, A and B seem to have the same idea about what to do next, and both refer to the cancelling of the dialogue box with the English name of the button (lines 64 and 65). B's formulation talks about the *Cancel* button (*känselillä* 'with cancel'); though the English word is assimilated to the Finnish sentence by adding a Finnish ending to it (in vowel harmony), *cancel* is pronounced the English way. A, however, talks about the *Cancel* button with a Finnish way of reading the word (*kansselia*) though she would be capable of producing the English pronunciation as well. Thus, she highlights the label character of the word by giving it a Finnish pronunciation, and by talking about 'putting cancel' 'into it' or 'there', an idiomatic and maybe even dialectical way of talking of choosing something or showing something (e.g. in a game of cards). Again, in line 71, A accepts or suggests the pressing of the OK button by reading the two letters with a Finnish pronunciation.

Their cancelling of one of the dialogue boxes caused the whole procedure to start from the beginning (i.e. a repair of the last step only was not possible). The speed with which the users go through the first steps is completely different now (from 72 to 83), demonstrating that the first part of

the procedure was easily reproducible (and therefore learnt) to gain their old position in the procedure. Now A is not suggesting any clicks, but accepts what B is doing (line 76), referring to the click of MAILABL as *siitä* 'there' or 'from there' and clicking of OK as *noin* 'so'. The first of these 'vocalised clicks' concentrates on the place on the screen where the word MAILABL is, and already anticipates clicking of OK (*ja* 'and'), the acceptance of which refers to the activity rather than the place of the button on the screen.

In line 78, B continues the vocalising started by A, this time referring to the name of the button, but giving it again a Finnish pronunciation [la:ser]. A accepts the selection with *joo* 'yeah' (line 81), and now both verbalise the next mouse click (A in line 80: *ja siihen okkia* 'and ok to there'; B in line 82: *ja oo koo* 'and o k'). B, who does the mouse click 'reads the label on the button' whereas A now makes OK into a word (instead of pronouncing *o* and *k* separately) which she reads in the Finnish partitive case, *okkia*.

So, there is a tendency to recontextualise the words on the screen in a way that turns them into labels on objects that are talked about, rather than talking about the action they are meant to do. This seems to be especially true for one word labels of buttons like *Yes* and *OK*; the longer pieces of text in the dialogue boxes were quoted with an English pronunciation (cf. line 67 above). Thus, the fact that they are placed on the visual metaphor of buttons which have to be pressed or chosen⁴⁵ has an effect in the shape of the human-computer interaction. If the user had to produce the 'answer' by themselves by typing it in, for example, these sorts of 'labellings' would hardly have arisen in the human-human interaction. Thus, when Streeck argues that objects can be symbolised (see 6.2), here we have a reverse case: symbols, words, are made into object like entities.

The manual text was never transformed in the same way, maybe because its role was more strongly found to be of knowledge *that*, i.e. information about what to do. As mentioned earlier, in line 49, B interestingly refers to the manual text as if the description of what should be on the screen was the screen (*tää näyttää nyt* 'this displays now'). However, this seems to be due to translating part of the same sentence she had quoted already in lines 33 to 38, which in the new recontextualisation is integrated into the Finnish conversation. The name of the dialogue box is kept in its English format (and pronunciation), though. So, the practical activity of engaging with the program, which uses metaphoric ways of presenting data and interacting with the user, encourages their objectifying of the words used on the screen. The manual has more a status of a mediated instructor to which the users turn to find answers to the problem at hand. In this work, the words are also placed somewhere, not on visual metaphors of labels (unless reproduced in a depiction of the screen), but on two dimensional pages and in a three dimensional book. This results in the use of deictic *tässä*, *tämä* 'here, this', etc.

⁴⁵ "Like the label on an icon in desktop interface, the text on a button is as much operational as referential" (Bolter 1997, 106)

in the human-human interaction the aim of which is to find the right instructive words for the problem at hand. The deictic expressions confirm Aarseth's (1997) claim about the hypertextual nature of the traditional book format: the users are pointing at the 'links' (often also with gestures). When the words are found, they are not objectified in themselves, but integrated into the practical activity through human mediation (reading aloud).

As discussed in earlier chapters, every act of reading a text means recontextualising it, because the interpretations of the reader are based on her previous experiences (layers of interpretants), i.e. everybody reads on the basis of their private knowledge *that/how*. In this sense, reading does not differ from conversing: the past experiences (knowledge *that*) of people with similar situations increase their ability to manage the situated encounter (knowledge *how*). However, in a cooperative situation, quoting a text is fitted into an ongoing activity, so the reader exhibits at the same time her understanding of the ongoing activity and the relevant next in it, and not just her private interpretation of the piece of text.

Thus, texts or words seem to be interpreted not only according to the structure of the text itself, but also on the basis of the material and interactional surroundings. In the following analysis, a fragment of Extract 6(1)b above will be returned to in order to examine how the different material particulars (appearing in a book format or on a computer screen) and the ongoing practical activity and interaction affect the interpretation process. To connect the issue of sense making to a larger, societal frame, the following quote from Hanks is useful:

Words, like other valued objects, circulate in social groups. Many may have access to them and use them, but there are elements of their value that only a part of the group will have access to. What makes communication possible is not the perfect sharedness posited by Saussure and Chomsky but the modes of cooperation among different actors. Moreover, one corollary of this premise is that a given word has more than one possible meaning, depending upon the public to which it is directed or the participants who produce it. This is another type of mediation that impinges on linguistic practice: the intervention of social organization as a defining factor in the relation of language forms to their meanings. This intervention, or mediation, introduces a double division among participating publics.

(Hanks 1996, 217)

In the case of the present data, the students are dealing with electronic and paper texts and objects produced by Microsoft. In Extract 6(1)b, the relevant part of which is repeated in Example 6(4) below, one word in the manual seemed to be a trouble source: *Toolbar*.

6(4)

- 31 A: m:itā?/w:hat?
 32 B: "Do not use (the) toolbar to open a new file"
 33 A: ="file because you need to use the temp^olates .option^o"]

- 34 B: [(h:)]
 35 A: erm ((gaze up to screen, moves back, left index finger next to text))
 36 oliko tuo nyt sitte se tool bar? (.) vissiinki./ was that then the tool
 37 bar? (.) i guess so.
 38 [eikä ku hetkinen onks se-/no wait a minute 'is it'-]
 39 [((left index finger up from page towards screen))]
 40 B: tä[mä on toolbar ((left index finger across the toolbar)) joo.]/
 41 thi[s is the toolbar. yeah.]
 42 A: [tämä on se toolbar ((pointing on the screen)). niin onki./
 43 this the toolbar that's right.]
 44 B: joo/.yeah
 45 A: [elikkä ihan oikein/so quite right.]
 46 [((gaze to text))]

The word *Toolbar*, the meaning of which is easily inferable by the parts of the compound, cannot be found in dictionaries (nor in the Word 7.0 thesaurus), ie. it is not an everyday word. For instance, the Collins Cobuild Dictionary, which is based on real language use (in print media), does not have it as an entry. However, the manual writers and designers of Word 2.0 seemed to have found the word a good name for a visual metaphor of adjacent buttons visible on the screen. Thus, not only was *Toolbar* a nonestablished compound, but the nonnative speakers of English and nonexperts of computer use had to find out what the reference of the word was. The extract shows an instant of how, this time, the cooperation of two actors resolved the problem that arose in their encounter of a word the referent of which had to be identified: the social organization Microsoft had used an unusual language form *Toolbar* with a meaning which had to be worked out by the public using the Word 2.0 program and manual. The aim of the corporation in using the concept hardly was to create any difficulty as the meaning should be shared for the practical activity with the program to succeed. However, *Toolbar* is an innocent example of how language forms get new meanings or new referents: later in the pairwork situation A used the word as a 'given' in another connection. Thus, the data provided an example of how a 'troublesource' was solved through interaction and a student learnt some 'Microsoftese' or 'computerese'.

In Chapter 5, Thibault's idea was introduced about the difficulty composite structures pose to language users; they "require relatively more constructive and/or interpretative work on the part of the language user so as to construe the new joint meaning which results from the combination of their constituent parts" (Thibault 1997, 283). In the case of *Toolbar*, more problematic than the concept of a toolbar as such, is that it is 'combined' with new environments of use: the metaphor of a toolbar is first read on the manual page, and in an action sequence which makes the whole sentence problematic, as argued above; its referent is something on a computer screen which provides a visual interface to a word processing program that relies on visual metaphors of buttons etc., representations choosing and clicking of which are in fact commands to the program.

One text, several ways of reading

The detailed analysis above shows how the material and interactional setting effects meaning making, giving support to Smith and Whalen's argument about the term 'identical text'. Sameness of a text or a document for readers cannot be decided beforehand, but depends on the context and use (Smith & Whalen 1995, 6).

In all four experiments, there were restarts of the task from the very beginning. Above, A and B read through the first page (the 'general' part) quickly and then concentrated on the (numbered) directions on the following page, combining the reading and doing. The same pattern could be found at the beginning of one other pairwork (C and D): the general text is read through, and it is the itemised instructions on the following page that get the participants started:

6(5)a

- | | | |
|---|----|---|
| 1 | C: | °elikkä "Setting up"° (3,5)/ °so "Setting up"° (3,5) |
| 2 | | (° []) että° New/°() that° New |
| 3 | | |
| 4 | D: | [joo./yeah] |
| 5 | | [((clicks File New))] |

The exchange is almost identical with lines 6 and 7 in Example 6(1)a, but it is D instead of the quoter C who starts using the program. In their retrials, C and D concentrate on the 'general text'. Also, their reading and doing are separate activities. For instance, the exchange above continues (as indicated by line numbering) as follows:

6(5)b

- | | | |
|----|----|--|
| 6 | C: | no-in./like that, (1) |
| 7 | | [katoppa löytyykö sieltä-/see if you can find there-] |
| 8 | | [((right index finger up and down in front of dialog box))] |
| 9 | | (.) |
| 10 | D: | mikä?/what?(.) |
| 11 | C: | [maillab dot. ((Finnish pron.)) (.) onko se siellä/is it there.] |
| 12 | | [((points to the word MAILLABL.DOT in the manual))] |

Now C who quoted the relevant next in line 1, prompts D to do something (line 7); in Example 6(1)a the quoter and the doer were the same person. C and D talk about the instruction and the program instead of incorporating the instruction into the ongoing activity: the manual's voice is lost, it is the users' interpretations that prevail.

Another pair (E and F) did not read the page with numbered instructions at all before starting, maybe because one of them could decipher what the general directions meant. They had to restart once, but even in the

second round, the detailed instructions were not adhered to. Only when the dialogue box seen in Figure 6-2 came up, did the participants read the detailed instructions; they only read after doing:

6(6)

- 1 F: ["Setting up and printing multiple mailing labels."]
 2 [((pointing towards page with a paper sheet))]
 3 (3)
 4 E: okei elikkä siis: (1)/okay so: (1) **noin**,/like that,
 5 [elikkä valitaan uus (1)/so a new one is chosen (1)]
 6 [((pointing at instruction number 1 in the manual))]
 7 ((following the lines in the manual:))
 8 **sitten tuotah**-(2) /and then-(2.0)
 9 "because you need to use the templates opti[on]".
 10 F: [hmm,]
 11 E: **noin**,/like that,
 12 (5)
 13 F: ((coughs)) **hetkine**/just a moment
 14 [täällä vitosessa on tuo./ here in five is that.]
 15 [((E points at the page))]

The third pair (G and H) had to restart several times; at first they would read the instructions carefully and then do, in the second trial they combined reading and doing, and in the last two rounds discussed the referent of *Toolbar* which was finally figured out. The following extract comes from the second attempt. At this point the 'Use Template' dialogue box is open.

6(7)

- 1 H: () "From the file menu choose new. Do not use toolbar to open-"
 2 **eihän tässä mitään tuulbaaria missään**/but there's no toolbar
 3 nowhere here
 4 "templates option. (In the) use template box-"
 5 G: **mikähän tässä ois**-/I wonder what here would be-
 6 H: **ei. em mää tiä. ei tässä**-/no. I don't know. there's no-
 7 ((selects MAILLABL))
 8 G: **kun en mää tajua mitä tuo tarkoittaa tuo template** ((with Finnish
 9 pronunciation.))/
 10 cause I don't get what that means that template.

This extract shows clearly that also for G and H, the *Do not use the Toolbar* sentence concerned the 'Use Template' dialogue box rather than the *From the File menu, choose New* sentence. And, because the *Toolbar* sentence was still found problematic even if this action sequence had actually been completed correctly, the extract above serves to illustrate what Hutchins has marked about following written instructions: "-- a user who does not understand the domain of action may know and be able to recall what a step "says" without having any idea of what it "means" " (Hutchins 1995, 299). H and G did not manage to complete the task, and were 'troubleshooting' this specific

sequence of actions, the description of which was not quite clear to them (cf. Isomursu 1997, 92). Had they succeeded with the task, the exact meaning of step 1, the *Do not use the Toolbar* sentence, would most probably not have been an issue.

6.4. Discussion

Computer manuals are inscriptions or records of instructions and descriptions geared to a (computerised) world-wide audience of hypothetical 'others'. The data in the present chapter shows how step-by-step following of instructions during an activity can result in a mismatch between the human-computer and human-text interaction. The material and communicative circumstances cause it to happen, whereas in a telephone dialogue system, for instance, the relevant next is offered to the user instead of the user having to find the relevant 'next'.

One reason for the miscommunication was the expert versus nonexpert language groups (cf. Hanks 1996, 220); two student pairs did not use the manual in a step-by-step fashion, nor did they exhibit any problems with the first two (numbered) directions. The difference is not attributable to difficulties in understanding the English language, as the students were fluent speakers of English.

In TUTORIAL (Chapter 5), the users had difficulties in making a distinction between 'to read' and 'to act' screens. This resembles what happened with the *Do not use the Toolbar* sentence: the user(s) who were going through the directions in a step-by-step fashion, and therefore had to interpret the sentence, took it to be repairing the action of choosing *File* and *New*. So, they took the instruction to be about their (possibly) having used the toolbar, about knowledge *how*, rather than information about why they should have not used the toolbar, knowledge *that*.

As was shown above, *Do not use the Toolbar to open the new file* as an instruction caused 'readers' intents' (Schiffrin 1990) ranging from the sentence being a repair initiator to dealing with the following, not the preceding instruction. Its status as a warning was only 'unscribed' when the referent of the word *Toolbar* was established.

As for the referent of *Toolbar*, Microsoft used it to point to a bar of icons on the screen (knowledge *that*). Those students that did not know or guess the connection immediately had to learn the connection (learning *that*). When they encountered *Toolbar* in their practical action or interaction with the manual text and the screen, during the knowledge *how*, it became part of their learning *how*: proceeding in the task was dependent on interpreting the word correctly. The division into knowledge *that/how* and learning *that/how* helps pin down the difference between information (*that*) and communication (*how*) in advice giving.

When the students were reading the manual to find help for their practical problem, the *mm-m* type of feedback showed to the other participant that the reader had understood or maybe even learnt something (learning *that*). Direct quotes always bring out a salient piece of text, either to proceed in the task (*choose new* in Example 6(1)a) or to vocalise problematic information, knowledge *that* (e.g. *to print one label multiple labels* in Example 6(2)). The quote given as an ‘answer’ is , then, framed (e.g. *niin kato tässä on* ‘yeah you see here is’; *joo to print on täällä* ‘yes to print is here’) to highlight the informative aspect of the text, the knowledge *that*: the information search is over. When learning or knowledge *how* becomes problematic in the text, the reaction is to question: *mitä?*, ‘what?’. To answer queries about ability or *how*, direct quoting is in place, again: *Do not use...* . The quote serves two purposes: to repeat the trouble source and to suggest an answer to the problem. This resembles TUTORIAL (Chapter 5) in which the repair initiators of the computer were also in the form of declaratives (e.g. *Move the mouse pointer...*). In both cases, the other human participant foregrounded the relevant piece of text by reading it aloud ‘for’ the textual object. Sometimes the quote would be transformed by laughter, additional comments, pitch and other devices: the inanimate ‘other’ is given a voice and brought into the sequential action.

All quotes served to manage or forward the task. Hanks writes:

When reference is made to the *current* situation, through first or second person pronouns, present tense forms, or proximate deictics, the two planes are superposed: The event being talked about is (or is part of) the one in which the talk occurs.

and

... the speaker is simultaneously defining himself or his current situation and being defined by it. Not only does the choice of terms in which to describe oneself depend upon the situation, but the description projects the speaker as a participant in the world talked about. And this anchors the self-expression in the world of objects.

and further, he defines quoted speech as

in which one speaker reproduces the words of another speaker, with appropriate attribution

(Hanks 1996, 206).

Following Hanks, we can see that the speaker, through quoting the manual words as relevant either as the next step or as the answer to the problematic point in the interaction, constructs the situation as instruction, with the manual as the expert.

To sum up, this section has explored the practice of word processor and manual use. In a pairwork situation, the participants make their actions

accountable to the other user-reader. This gives the researcher a natural access to the interpretations, unlike in solitary, silent reading and acting upon texts.

In pairwork, text can be quoted directly: the knowledge *that*, the textual time of the text is activated into knowledge/learning *how*. It is possible to do the quoting without referring to the 'speaker' (i.e. direct quoting in its strictest sense in which the speaker is identified, e.g. 'she said') because the textual object is present; however, sometimes the users do frame the quote (*tässä on 'here is'*). The framing orients to the manual as a material knowledge object which can be entered from any place, and which also allows for 'peeping ahead'. That the text is used as such in the turn-taking makes it as an integral part of what people are doing there and then: the endophoric or textual and exophoric or situational relations (Halliday and Hasan 1976) will merge into 'endexophoric' when texts are used in a concerted manner in the unfolding interaction; also, the 'expert' manual text does not the work on its own, but the expertise is interactionally constituted (cf. Goodwin 1986b).

Sometimes text is reacted to directly, without quoting it. The feedback reflects two types of knowledge put forward by the text: 1) text as information (knowledge/learning *that*) which is reacted to with mm-m; 2) (same) text as ability (knowledge/learning *how*) to give a possibility to reconsider understanding/interpretation (i.e. what text meant or what the user did): *mitä?*.

The words appearing on the computer screen, especially on the metaphoric buttons (e.g. *Yes*, *Cancel*, and *OK*), lose some of their symbolic and interactive power and are treated as material objects, rather than as contributions in the ongoing interaction.

6.4.1. Shape of repair work

When participants make their interpretations accountable to the other, they can also do or instigate repair work to adjust interpretations. These points of interaction are important for the researcher as they reveal difficulties in understanding and difficulties in repairing the understandings. This section concentrates on the shape of repair work emerging in the pairwork situation of using a manual to help use a word processor.

Frohlich, Drew and Monk (1994) researched the characteristics of repair in human-computer interaction in which the 'participants' were the user, the assistant and the computer (Frohlich et al. 1994, 395). According to Frohlich et al., the most common repair forms in human-computer interaction are 1) Self-initiated self-repair in initial position (User: Repairable + Repair), 2) Other-initiated self-repair in next position: (User: Repairable, Computer: Next-turn repair initiator (error message), U: Repair), and 3) Self-initiated self-repair in third-position (User: Repairable, Computer: Response, User: Repair) (ibid., 389). Though both the first and the last type could be found in their corpus,

Frohlich et al. concentrated on number three especially, and described how a prior granting is undone (during a pause) or a prior request is redone, or a new request done (ibid., 413). In their article, Frohlich et al. do not analyse in detail the cooperation between the user and the assistant, and in the final analysis of repair sequences concentrate on the user's actions on the screen as the 'user input'. The system under examination was "a direct-manipulation style of interaction in which all possible user actions are legitimate from the point of view of the system" (Frohlich et al. 1994, 417), which meant that the system did not give any error messages. In this respect, the word processing program used in TASK was similar. However, in TASK, in addition to the computer and the users, the manual was also a party in the ongoing interaction. In this complex situation, the third type of repair (self-initiated self-repair in third position) is shaped differently: with the lack of feedback from the system in the form of 'padding' or repair initiators, the users are faced with continuous uncertainty about whether they 'have done the right thing'. This is why prechecks and 'postchecks' of mouse clicks and other activities are done, and in this work the manual is the source of 'correct knowledge'. In Example 6(1)b, line 23, A checks whether she is about to do the correct mouse click, and in Extract 6(3), line 30 onwards, B resorts to the manual to see whether their 'guess' of choosing *Yes*, was correct. In both cases the users have no notion whether they have done or are about to produce a repairable; they have to find the answer to their pre or post check from the manual. In actual fact, all the requests by the users are guided by the manual: they are recontextualised or interpreted instructions activated by the user in each stage of human-computer interaction.

6.4.2. *Reading vs. encountering*

In the case of TUTORIAL, text-user interaction was easy to decipher because computers are used primarily, not read, whereas in TASK the manual was an object which looks like a book, or a written document which is first of all read, not used. In relation to knowledge/learning *that* and knowledge/learning *how*, there are the following possibilities: the manual can describe knowledge *that* or knowledge *how*, and the reader can learn *that* (when reading) or learn *how* (when using the knowledge in practice).

In the case of text on screen, the activity/action was foregrounded, but maybe it would be useful to concentrate on the manual also as a 'party' in the activity that is going on. There were specific points in the interaction in which the text of the manual is 'recontextualised' by being realised as action (cf. choosing *File New* with the mouse, which was actually not an exact rendering of the text which told the reader to use the *Alt*, *F*, and *N* keys).

The role of the manual as an 'instruction giver' in the activity was especially clear when the user took the manual's 'next' to be logically about

the 'next' in the human-computer interaction (e.g. the sentence *Do not use the Toolbar...*). The tension between the reading as learning *that* (acquiring information) and as learning *how* (acquiring a skill, i.e. doing) is constantly there. Confusion results if the visual semiotics and the sequential placing or otherwise marking (in the case of the manual, numbering was not enough as an indicator of 'this next') of what is meant to be read as 'do this now' (ability, knowledge *how*) and what as 'read this for general information' (knowledge *that*) are not successful. Of course, a knowledge *how* piece of text can be read as a knowledge *that* one (when we read the instructions without doing anything). In both TUTORIAL and in TASK, this was not a problem, but when a knowledge *that* text is used as knowledge *how*, problems arise.

6.5. Conclusion

In both TUTORIAL and TASK, the texts of the manual and of the tutoring program were dialogical: there were lots of references to the reader as 'you'. However, there was no reference to an 'I' or 'we' on the part of the program or the manual: the 'author' is Microsoft, and the 'narrator' only implicitly present. When the learning program in TUTORIAL was usable and readable only at the computer, the manual could be read apart from the computer program it was describing. Indeed, the manual text was compiled as a descriptive one, even at points of numbered instructions. The descriptions of what happens on the screen (*Word opens a new document*), are what in TUTORIAL take place in the (virtual) program: the former is a textual, informative representation, the latter a visual, interactive representation of an event. The title *To use MAILLABL.DOT and run the mailing label macro* was the main sentence which the numbered instructions connected to⁴⁶. But, unlike in the general descriptions (*When you want to create...*), there is no dialogic *you* in the instructions, though the formulation of the heading would allow for *you* to be used (e.g. 'To use MAILLABL.DOT and run the mailing label macro <new line> 1. from the file menu, you choose New'). Now, a combination of imperatives with omitted *you* (cf. TUTORIAL *Practice* screen) and general description with *you* (cf. TUTORIAL *Overview* and *Steps* screens) is offered to the reader.

So, though the manual had numbered instructions at points which were designed to be the guidelines for practical activities, even these segments were more descriptions of activities which work well and are coherent in the textual time of reading them, but can cause disruptions when used as step-by-step instructions in the real time of the activity. According to Smith:

⁴⁶ It would be very informative to show a scanned page from the manual here, but unfortunately Microsoft Finland did not give me a permission to do that.

In textual time, the process of working up the formulation becomes invisible.
The account comes to stand in for the actuality it claims to represent.
(Smith 1990, 74-75)

Smith's claim also holds true for very 'transparent' texts, such as manuals which are descriptions of the computer and its workings. However, the internal time of the text becomes consequential in the practice of text use when the writers are describing the hypothetical use of the computer. Above, the problematics of the *Do not use the Toolbar to open the new file* sentence was extensively dealt with. One aspect of the sentence was that, in the textual time of not only the text but also the writers, the previous command *From the file menu, choose New (Alt, F, N)*, created *the new file*. At the time of the text activation, this coherence was lost, though, because the screen was another party in the interaction, and according to the conversation analytic principle, each turn both determines the sense of the previous utterance and limits the scope of interpretation of the next one. When the human-screen interaction is out of phase from the human-text one, the internal coherence of the textual time will cause difficulties in the practical action. This shows how difficult it is to make the present tense of the text work when the time of the reading has to break out from the textual time: within the text, the time of the reference and the time of the 'speaker' is the same (see e.g. Silverstein 1976, 24). However, some students would use the general description as a guideline, which shows that they oriented to the manual as a dialogic narrative (in the present tense) rather than as something with clear cut descriptive and prescriptive parts. Whichever the approach, the manual was adhered to in search for knowledge to enable interaction with the computer. The manual offered the knowledge (*that*) mainly in linguistic signs, but also as numbers and pictures depicting the screen. The user had to interpret these to enhance his or her interaction with the computer. The continuous linking of the manual text and the screen became visible in the users' postural orientation, gaze, and pointing which, together with quoting and other talk, connect the two domains: dual point (Goodwin in press) was a basic format of referring.

Aarseth defines discourse planes in narrative, hypertext, and cybertext (Aarseth 1997, 126) as combining (in narrative) and separating (in hypertext) Progression and Event, and, as separating Progression, Event, and Negotiation (in cybertext). In comparison with this classification the discourse planes of the manual text remain as reader-oriented Progression (narration of events; cf. the 'general text' in the manual) and Event in which the unfolding of the computer use is described (in a numbered fashion). The visual and logical placement of the *Do not use the Toolbar* sentence is an example of this: it is a remark in Event rather than a (manual) user-computer interaction oriented warning. In his search for a literary theory to cover the new interactive media which are not satisfactorily explained by the traditional literary approach to narrative, Aarseth introduces 'ergodics' (the reader's actions to progress in or create the narrative). He briefly mentions

hypertextual user's manuals as examples of descriptive and ergodic (i.e. the 'Event' is described, and the user progresses by clicking something on the screen) (Aarseth 1997, 95). But — whether in electronic or paper format — manuals 'work' is not just to create a text-inherent 'narrative': they have to guide 'real world' activities. The challenge is to make this 'external negotiation', lacking from Aarseth's theory, to work. However, ergonomics helps to understand the problem of the manual writers: they write a descriptive text to a narratee-reader, and therefore do not address a narratee-user who has to act in the world on the basis of the text.

The manual as an artefact allows the reader to use it in a liberal fashion. The search for the page to look at is limited on the basis of the job that is done with the program: in TASK, this meant that the manual was already open on the page with the start of the description of how mailing labels are produced. In a way, this freedom of moving in the text resembles hypertext, and the index at the end of the manual serves as a list of 'nodes to click'. The manual offers above all information, be it about knowledge *that* or knowledge *how*: it is a form of hypertext that, a collection of 'clickables' that inform the reader about the program, and about the use of the program.

However, the screen of Word 2.0 that the manual refers to could be considered to be hypertext *how*. The reason for this is that the screen offers options (in the form of buttons and menus) from which choose to do something (instead of getting more information) — with the exception of the Help menu which basically is a node to get to a hypertext format of instructions like the printed manual. In a word processing program, the linguistic or other icons to be chosen are closed: they refer to some specific function the user can 'tell' the program to do; they take the interaction forward. On the basis of the user's interpretation of the linguistic or other sign, he or she clicks or otherwise chooses the linguistic or iconic depiction of the function.

On the basis of the present chapter, the question of what makes the human-manual interaction go awry can be attempted to be answered. It seems to be important to 'keep relevant parts together', both visually and in the order of expressing information. It is not enough to have a coherent text when read in textual time: when instructions for practical action in the world are designed, the relevant parts of the instruction concern one action. Now, the warning about the very first step (*Do not use...*) described in the manual about opening a new file from the *File* menu was separated both visually and topically on the page from the action taken. In fact, a written sentence on a page of a manual can 'behave' as if a repair initiator that appears interactively on the screen (cf. TUTORIAL). The reasons for this possibility were: the placement (after a command), visual layout (an empty line between the command and the sentence, though no numbering), and thus lack of linking it to the preceding direction as an explanation (rather than warning or repair) by embedding it in a virtual problem state (e.g. 'If you do not have the Use

Template option visible, you have clicked the Toolbar icon' etc.) (see also Raudaskoski 1997, 542-543). This shows the power of interactive language when it is used in a general situation of learning *how*. It would have had different consequences in a learning *that* situation, when read separately from the user's practical action with the computer.

The materiality of text on pages in a book format and hence the 'autonomous destiny of writing' was discussed above. In text-user interaction with manuals or other directions (of use), however, the limits of both time and space become relevant. The writing cannot be independent of the sequential nature of acting in the world anymore: it becomes incorporated in practical action, and in this 'bringing it to life', the material characteristics of the language become important.

Agre (1988b) also points out how a set of written instructions are interpreted anew on each occasion of use. My analysis has concentrated on not only how difficult it is to design 'identical texts' for different practical circumstances, but how the material appearance of the writing and the whole setting is part of the 'anew' aspect. According to Suchman, expert systems are built to go around the problem of instruction giving: "It is this lack of "recipient design" in the written instruction manual that the expert help system is designed to redress." (Suchman 1987, 132). As was shown earlier in Chapters 4 and 5, the recipient design in the form of repairing the user at the time of the problem, or relevant entity, is not easy either. However, separate, written instructions can be geared towards supporting the practical action rather than the internal coherence of the instruction text.

The fact that the manual was 'another party' in the interaction certainly had an effect on the intensity of the human-computer interaction. But there was another reason: the word processing program relied on the visual metaphor of a work space, the interaction in/with which mainly consisted of choosing from a predefined set of answers: the users were interacting with buttons, and the buttons were objectified such that the human-screen interaction changed from human-text/'other' (representation) into human-button 'interaction'. There was no direct interacting with an entity.⁴⁷ According to Biocca, "the body enters cyberspace with the creation of the humble mouse" (Biocca, 1997) because the mouse movement is qualitatively different from typing on the keyboard. "Why not use the keyboard someone might ask? They keyboard was primarily a symbolic input device for textual "conversation" with the computer. The keyboard did not map the movement of the body in space to cyberspace. So it was conversational input, not a

⁴⁷ Andersen (1990) in his taxonomy of computer signs rejects the Peircean division (index, icon, symbol) for a classification which is based on how the sign appears to the user. Though Anderson manages to stress the importance of the material surroundings, he makes an *a priori* and strict division into linguistic interaction and tool use, though – from the user's point of view – computer signs could represent both categories.

somatic input.” (ibid.) However, when the somatic input is part of interacting with a conversational or symbolic device, i.e. with language, then the ‘somaticity’ coupled with ‘conversational’ results in managing words (superficially), not conversing.

The knowledge *that/how* distinction was used to get to grips with the difficulties of using text to instruct with a practical activity. Knowledge *that* is connected with literacy and the written mode, knowledge *how* with practical action. When *how* should be done with the help of *that* (written mode), difficulties arise. The problem seems to be the tendency for writers of manuals to produce narrative (knowledge *that*) which does not always give enough support to the activity-in-progress. When the reader of ergodic literature (Aarseth 1997) is able to create a narrative by engaging in practical activity with the text, the authors of manuals have a much more demanding task: they have to be able to guide the practical activities of the reader, outside the text. When a manual is successful as a coherent and accessible text, the authors have helped the reader produce a good narrative. Aarseth uses the concept ‘textual machine’ to combine the language (verbal sign), medium, and interpreter (operator) in one interconnected whole (Aarseth 1997, 21). However, to bring about a coherent and successful interaction with another textual artefact, a good text-internal narrative is not enough. Instead, a context sensitive dialogue or interaction is required: one ‘textual machine’ (computer as medium) is interacted with by the one and same ‘operator’ that is involved in the ‘textual machine’ of manual-mediated language.

However, descriptions are still important in the manual: in the form of a) Visual depiction of the screen or a dialogue box, or b) verbal (*From the File menu...*), by referring to the salient feature on the screen. They foreground the necessary requirements for or results of an action which is brought about by the reader’s action-in-the world (external vs. internal experience) on the basis of requests/instructions of the manual. In the present case, the numbered instructions (knowledge *how*) were like the general description parts (knowledge *that*) in that both were dialogic (talking to a ‘you’), and more importantly, also the step-by-step orders given were in a way a past ‘testimony’ of an activity (e.g. away from a computer, describing to somebody in detail how to use a function of a program) instead of ‘giving directions’ to somebody who is engaged in using a computer in a ‘temporally shared social reality’ (see Chapter 5). Thus, the phenomenon resembles that between referring and indexical functions of language use discussed by Silverstein (1985): though the latter is primary for a language user, the former tends to be understood to be fundamental by linguists and language philosophers. The present case study showed that if reference (description of events, knowledge *that*) is the mode in which indexical actions (knowledge *how*) in the world are guided, problems arise if the text user orients to the text as indexical (and treats the text as guiding his or her unfolding practical action) rather than informative.

So, if an improvement in the interactivity of guiding advice/help manuals is the aim, the focus of research should move from explicating the features that improve the potential reader's understanding of the instruction's content as such (e.g. whether long or short sentences are easier to grasp) to the level of use, of sequential interpretation: when directions are to be 'unscribed' into action, then writers have not only to write clear and concise commands, but to take into consideration that they will be followed as a support of human-machine interaction, which – similarly to every activity – takes place in a sequential manner. Successful doing (interpretation) on the basis of instruction seems to create different demands to the text than successful reading (and understanding) of the same text. Instruction as a speech act category merges knowledge *that* and knowledge *how*: Instruction is always prescriptive, telling the user how to do things, but whether it is interpreted as a request to read now or request to act upon reading now (cf. Chapter 5) is blurred by the ultimate narrativeness of the text. It 'represents' the activity of instruction giving in a textually coherent manner. Thus, the reader as an instruction taker in an ongoing activity has to work to fit the narrative into practical action: the less experienced the reader is with the domain of action, the more difficult the transformation becomes, as shown in the examples above.

Conversation analysis worked well in this case study, too, as a way of capturing and understanding the complex interplay between the semiotic fields activated and the sequentiality of the participants actions.

7. TEACHING: THE USE OF COMMUNICATIVE RESOURCES IN INTERNET VIDEO CONFERENCING

The last case study, TEACHING, presented in this chapter is qualitatively different from the others: the situation was authentic, i.e. the data was collected from a set of 'naturally occurring' university seminars. Like in TUTORIAL, the process and the product of the interaction were intimately connected: learning took place throughout the encounter. The 'others' met on the computer screens were in this case real people who were mediated to the classroom via video picture, sound and text: the computer was a mediator rather than originator of language (see e.g. Nass and Stauer 1993 for a discussion about computer as a medium or as a source of language). The research material thus comes from a very complex semiotic situation providing an opportunity to investigate the relationship between ongoing visual, aural and written interactions. The data shows how language use constructs the situation as happening in two places, and how the video picture shapes the interaction. The claims about asynchronous or 'interactive' written versus spoken guidance for action, put forward in previous chapters, can be investigated now in a language technology environment that allows for (almost) synchronous ('talk time') video and audio connection with a possibility to communicate also by typing on the video picture. By closely examining the ongoing practical activity, the analyst can reveal how people interpret and accomplish the sense of the mediated pictures/sound/text *in situ*, instead of making analyses divorced from the actual 'reading'.

7.1. Introduction

The coming together of social actors in the (post)modern world of technologies is increasingly mediated, and the Internet can bring 'others' from far away into a daily contact, distorting the here/now, there/then equations: "The development of global cities disrupts the time-space co-ordinates of natural space" (Lash & Urry 1994, 55). This chapter reports on a study of human-human interactions mediated through desktop video conferencing and argues that the activity differs in interesting ways from face-to-face and

human-computer or human-text interactions. The study particularly focuses on the consequences of the special interactional setting for the emerging activities and talk vs. typing. Desktop video conferencing is becoming increasingly available across institutional and private settings as a cheaper solution than traditional video conferencing facilities (Ehlers & Steinfiel 1992). Therefore, it is important to understand the interactional dynamics of this new medium of real time synchronous communication.

The corner stone of information society, the Internet, has brought text-based communication towards synchronism. For example, the Talk program in Unix allows two users to share the screen for simultaneous typing and the text version of Internet Relay Chat can connect a number of discussants at the same time. However, recently desktop video conferencing systems have become available for use over modem lines and the Internet. They provide a moving image of a participant: the 'other' is there also visually (as themselves, not as a 'blockie', 'avatar' or other such representation of a person in the visual multi-user domains; e.g. Bowers et al. 1996). In CU-SeeMe video conferencing, it is possible to link to a site via video picture, sound and text (typed either on the video picture itself or in a separate Talk/Chat window). This is a fairly new form of interaction, the conventions of which are in the making. My data gives examples of how in CU-SeeMe environment, spoken and written texts are combined, differing from purely spoken and purely written communication, especially from the point of view of possible trajectories of text interpretation (a comparison of design features for speech and writing is given in 2.1.1).

Desktop video conferencing differs from watching TV or reading newspaper texts and pictures in that as a participant, one is actively involved in the process oneself. Also, the physical distance between the user and the video conferencing facility is much narrower when the video conferencing program is run on a computer rather than being watched in a special video conferencing studio with TV screens. In past studies of computer-mediated communication environments, video conferencing has been regarded primarily as enhancing the feeling of shared space (as the time can be shared in writing only systems, as well). Therefore, video conferencing has been seen especially useful for 'person-centred' activities in order to enhance social presence (e.g. Svenning & Ruchinskaskas 1984, 242).

Schutz (1980) gives a comprehensive account of directly experienced social reality. He describes how in a face-to-face situation we are aware of the other(s) and how we mutually monitor each other, but at the same time cannot possibly 'see what they can see'. However, with video conferencing systems, we are suddenly provided with the same video picture (literally) as the one our interlocutors are seeing of us. Also, not only has the researcher an available 'facial' view of the participants (de Fornel 1992, 171), but also the same visual perspective of the mediated 'others' that participants in a video conference have of their distant interactants. In my research on desktop video

conference interaction, special attention is directed towards the potential that this 'sending picture' will bring to the interaction in general, and to the interpretation of directives.

In the CU-SeeMe version used in the present case study, it was possible to type on the video picture. In other versions, a separate Chat box can be opened for text-based communication. Thus, desktop video conferencing becomes a multimodal means of communication: the video picture and sound can be used at the same time as textual communication (on the video picture or in a separate Chat box). The communicative possibilities are extended beyond the dichotomy of face-to-face 'now' and textual 'past' and the data makes it possible to investigate textual on-line interaction as another mode in visual/sound interaction. This has not been an issue in studies on such text-based computer-mediated communication systems as asynchronous computer conferencing (e.g. Davis & Brewer 1997, Sorenson 1997), and synchronous split-screen discussions (e.g. McIlvenny 1990). In fact, video conferencing tends to be studied either as emulating face-to-face interaction or improving mediated interaction. The aim is understandable as many of the researchers work for companies that design and sell the systems. This is why the use of 'proper' expensive, high quality video and audio systems for group meetings (video studios) and personal interaction (desktop video conferencing/phones) have been studied quite extensively. Usually, the interactions researched have been (futuristic) experiments or the participants had or could have met face-to-face (e.g. geographically close or distant workplaces). Finn et al. (1997) gives a comprehensive overview of studies into video-mediated communication.

In my data, it is possible to show the power of text-as-produced-as-a-turn-in-interaction as a contribution to the interaction at a later stage, to have a glimpse of the 'textual machine' (Aarseth 1997) being produced by the 'interactional machinery' of a 'face-to-face' interaction with an extended set of resources. The literal overlapping of text and video conference mediated talk-in-interaction makes it possible investigate the 'past' quality of text-as-an-interactional-device: the online video communication becomes a site in which the transformation from synchronous to asynchronous text can be witnessed.

7.2. Background

For decades, estimations of the rapid expand in the use of video conferencing products have been common (e.g. Suoknuuti 1991). However, maybe only a portable video phone will be able to make video conferencing an everyday way of connecting people. Types of and reasons for video conferencing use are manifold, and thus the research has concentrated on differing issues and with differing methods (Finn et al. 1997). The aim of using video conferencing has been creating social presence to a distant link (Svenning & Roschinskas

1984), or more generally, maintaining face-to-face contact (Sallio et al. 1983), or, giving an impression of personal interaction (Santamäki 1984). Video conferencing is meant to enhance informality (Fish et al 1992) or just the awareness of a mediated 'other' (Dourish & Bly 1992). All these aims and claims are based on comparing the video link with other mediated communications, e.g. telephone or computer conferencing. However, the critics of new technologies list video conferencing as another medium which cannot replace the intimacy of face-to-face interaction (e.g. Boden and Molotch 1994, 258). Then again, some researchers regard collaborative learning tools, video conferencing among others, to be furthering feeling for and thinking with people who first and foremost share the same interests, rather than the same physical space (Bonk & Cunningham forthcoming).

The quest for having 'as if' face-to-face situations has naturally led to a comparison of the mediated and real copresent situations. For instance, the effectiveness of presentations via a video link has been checked against the same talks given in a shared space (Isaacs & Tang 1997). Studies of this type aim to offer their experiences and solutions to the problem of "trying to be in two places at the same time" (Barnes 1997, 209). However, in the multimedia packages which are used for collaborative work and learning, it is not so much the video picture but the objects that can be changed, pointed at, or otherwise manipulated as if shared, that allow one to be (or share an object in) two places at the same time. The video conference itself is still very much about trying to see two (or more) places at the same time (and cyberspace about trying to be in another place at the same time).

Apart from the highly sophisticated laboratory versions of video conference programs (see e.g. Finn et al. 1997 for an overview), the everyday use versions in educational settings, for instance, can be divided into video studio and desktop ones. Video conference studios are special rooms, often with fixed positions of television sets. The situation resembles very much that of watching television, because the participants are fairly far away from the TV-sets, making video conferencing a literally distant educational tool. The receivers are passive watchers of 'talking heads' (cf. Sulkunen 1992). However, some researchers do not see any problem in that video conferences are like live TV-programmes, and at the same time they call the encounter face-to-face communication (e.g. Rahko 1991). Desktop video conferencing programs are sometimes called video phones, maybe due to the resemblance to telephones: the contact is often one-to-one, and much more intimate also because the video-mediated 'other' is physically nearer to oneself on the screen.

Both types of video conferencing have their problems, and research attention has been paid to get rid of the 'lecture room/TV-watching' sense of video studios. Also, the strangeness of seeing oneself and other participants on the desktop video conferencing screen has been mentioned as an issue that might be important in designing video conference systems: "you appear in

the same format as everyone else, in contrast to ordinary life, in which you view other people from a radically different perspective than that from which you view yourself" (Borning & Travers 1991, 18). I discuss this issue in Raudaskoski (1996). Via the feedback screen, it is possible to monitor what the other party can see of ourselves, and *vice versa*. This is of course quite an extension to our normal awareness of the other. For instance, Bakhtin's conception of necessarily different views in and of face-to-face situation is violated:

The simplest way to state the difference between us is to say that you see things about me (such as, at the most elementary level, my forehead) and the world (such as the wall behind my back) which are out of my sight. The fact that I cannot see such things does not mean they do not exist; we are so arranged that I simply cannot see them.

(Holquist 1990, 36)

So, video conferencing already is what some developers hope it to be, namely 'beyond being there' (Hollan & Stornetta 1992). With video conferencing, the locus of oneself changes from the 'I' and its here/this/now to the there/that/then of the screen, side-by-side with the 'others'. As the desktop video conferencing picture normally lags behind, the participants can watch themselves, they can be 'eavesdropping' themselves in the immediate "objective past" (Wiley 1994, 45). To make the mediated 'others' as close to our normal face-to-face experiences as possible, some researchers have envisaged and implemented video environments in which the mediated 'other(s)' are part of the context in the same way as a copresent other would be. (Gaver et al. 1992). To test other ways of having multiparty video conferences, video pictures have been moved from the screen onto separate monitors (with separate audio channels) (Sellen 1992). Buxton calls these environments 'ubiquitous' video: video conferencing does not take place in a separate room, nor on one computer with one camera but several monitors are mounted in offices: one for visitors, another for office mates etc., to enhance the feeling of everyday practice (Buxton 1997).

Sometimes video links are more important for sharing the activity space rather than personal space, for instance to convey a shared surface for drawing together (Ishii & Kobayashi 1992) or to give the feel of three dimensional objects that are manipulated according to the directions of one of the participants Kuzuoka (1992). These studies on media spaces are often trials done in laboratory settings. The experience from real work environments seems to be that not only is the video picture of the other party important, but seeing one's own picture makes it possible to estimate what the view at the other end is of oneself, especially if it is possible for the other party to adjust the video picture of oneself (e.g. Heath et al. 1997).

In the present case study TEACHING, a 'personal' video conferencing program on the Internet (CU-SeeMe) was used to link up two student groups, one in Finland and one in Sweden. Thus, the institutionalised video studio

was not used, not only because of the higher cost but also due to the design of the local university video studio which was fairly fixed and was predicted to create a TV-watching/lecture listening atmosphere rather than a seminar with active participants. The low cost video conferencing solution resembles those used in other educational contexts, as reported in Sattler (1995). The data recorded from this 'virtual seminar' makes it possible, not only to see how the seminar was managed via the link, but also to find out what was new and different from face-to-face and audio interactions (or even video studio and TV-watching activities). One aspect of interaction that to my knowledge has not been researched is typing on the video screen itself, even when the audio connection was working. This feature of the CU-SeeMe video conferencing system does not work in the latest versions.

7.2.1. The CU-SeeMe video conferencing system

CU-SeeMe (pronounced See-You-See-Me) was developed in Cornell University, USA, to enable anybody connected to the Internet via Macintosh or Windows to participate in video conferences. Thus, the CU-SeeMe environment is one of the latest forms of synchronous computer-supported communications across the world and, according to some experts, "it may be the harbinger of things to come" (Angiolillo et al. 1997, 64). The potential number of CU-SeeMe users is huge, as the number of Internet users is rapidly growing. It is possible to have dyadic sessions, in which two computers are directly linked (as in the present study), or to contact so called reflectors in which a limited number (usually not more than 20) of participants can have a chat with each other or monitor others having conversations. So, unlike traditional video conferencing in which people gather in special rooms within a strict time limit for the activity, desktop Internet video conferencing allows participants to pop in and out of interactions across the world.

The video pictures in the present study were black and white (there is a colour version of CU-SeeMe nowadays), and appeared on the screen as a double size version (see Appendix 7-2 which shows two double sized and one normal size video picture on a 14 inch monitor). The version of CU-SeeMe used was an earlier freeware version; in Sweden, a Mac version was used, and the Finnish site used a PC version. The then PC version did not have the Chat box (called Talk menu in the Mac version) available for written communication. Since then (1995), CU-SeeMe has launched commercial versions of the program.

In the search for 'life-like' video conferencing systems (e.g. Gaver et al. 1992, Buxton 1997), typing on one's video picture, which emphasises the picture as a mediated representation of space, has not been of interest. The possibility of typing on one's picture has been considered as a 'primitive' form of communication (Sattler 1995, 112) in comparison to a separate box for

typing entries or to an audio link resulting in the abolishment of the possibility from newer versions of CU-SeeMe. In addition to the mediated video and audio, together with the Chat box (called Talk in the Mac version), the use of semi-permanent scrolling text on the video picture gives another dimension to the interpretation of what is going on. When the participants use the possibility to type on their sending picture instead of using the Chat/Talk window, the receivers do not have to find the correct video frame to match the speaker. So, the text on the picture has a property in common with the voice of the person: the receiver is able to locate the source of the words at once. But, unlike with natural voice sounds, the text does not disappear immediately: as more words are added, the text runs out of the single scrolling line on the video frame. Often the last part of text is left on the screen, becoming a disembodied piece of language as the participants move forward to the next activity in the situation.

In their book on the virtual classroom, J. Tiffin and L. Rajasingham see the ease of use of desktop video conferencing as answering to some of the problems of traditional video conferencing in education (1995, 112). What is happening in the U.S. schools, for instance, is that CU-SeeMe is used to combine one-to-many and many-to-many discussions: the Global SchoolHouse project has employed CU-SeeMe to 'bring' celebrities to schools all over the country, allowing school children to talk to the specialists but also to other schools (Sattler 1995, 164).

7.3. Interactionist approach

Many studies on human-computer interaction deal with the strength or weakness of computer-mediated communication to create a feeling of sharing, be that sharing the linguistic code as a tool to act with the computer, or, especially in the case of video conferencing, the feeling of shared presence. Often the research is conducted by asking the users afterwards how they felt about the encounter (e.g. Muhlbach et al. 1995). My choice has been to analyse these semiotically complex encounters in detail to grasp the interpretative work being done in the situation, in other words, to treat signs as communication and dialogue. The approach is that of interaction studies, which is an umbrella term for analyses in which different human interactional environments are studied to better understand how the individual realises the communicative potential in their use of language, gaze direction, posture, and gestures. In this view, language is strongly rooted in the situation, and its meaning cannot be divorced from the moment of its use. Thus, negotiation of meaning is taken very seriously and studied as a local and emerging phenomenon in ongoing human practices in real time (see Section 2.4.)

7.4. The data

My data comes from CU-SeeMe video conferencing during a one month teaching experiment which was undertaken in a Finnish university, whereby a (British) teacher spent a month in a department within a Swedish university in Spring 1995 (see McIlvenny 1995 for a project report). The World Wide Web was used as a database, Mosaic as the navigating program to access different sites and to share the documents, and CU-SeeMe as the desktop video conference link for face-to-face interaction. Electronic mail was used for additional exchange of ideas and other communication between the participants. The groups met via the video conference link once a week for a two hour seminar; one group of students were Finns in Finland and the other Swedes in Sweden; English was used as the *lingua franca*. The weekly video conference sessions gave the students a possibility to give presentations, to provide direct, 'face-to-face' feedback, and in general to talk to each other and to see each other in a seminar type situation⁴⁸ (Figure 7-1).

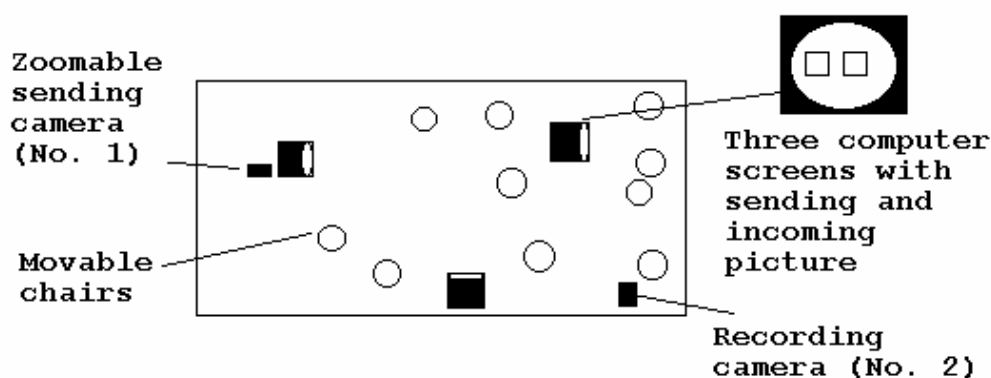


Figure 7-1

This type of communicative environment was new to all the participating students, though they were experienced in attending seminars, using personal computers and talking to foreigners. I videotaped the video-mediated exchanges between the two sites. One recording monitored the camera which was used in Finland to send the video picture to Sweden (No. 1), and a second recording was made with another camera in a corner of the room in Finland (No. 2). Both of the cameras were stationary, though the view of the sending one (No. 1) could be zoomed to focus on individual participants. During the seminars, in addition to the two recordings made in the room in Finland (Figure 7-1), a third one was collected from another site that was connected to a so-called reflector. That recording made it possible to see clearly what was written on the video pictures. This additional data was crucial to be able to

⁴⁸ See Appendix 7-1 for a snapshot of a screen in Sweden and Appendix 7-2 for a snapshot of screen activity in the middle of a seminar taken from the recording camera No. 2.

investigate the role of written interaction in the seminars. As the Finnish end was video recorded with the sending camera and another camera in the corner of the room, all talk and typing sounds were available for the analysis. The third recording could be used to fill in the information of what was typed on the screen; Appendix 7-3 shows an exemplary frame from this video tape.

CU-SeeMe and also Maven, an audio conference program used in some sessions, were freeware from the Internet. They enabled on-line meetings between the two student groups who otherwise would not and could not have 'met' each other 'face-to-face', at least not in the framework of the university seminar in question. Thus, the setting of the conferencing was 'unorthodox' in the sense that desktop video conferencing was used to link two student groups (from 6 to 12 in number at each end), using several monitors to convey the black and white video pictures to facilitate the interaction between the two sites – at the Finnish end, in an office-turned-to-a seminar/video conference room and at the Swedish end, in a computer class.

The need of a video camera to send the video picture made data collection easy and undisturbing. As most detailed data was gathered from the Finnish site, the analysis will consequently be based mostly on what happened in the room in Finland (cf. Meier who regards it "sufficient to document what is going on at one site rather than both sides" (Meier 1998, 10)).

7.4.1. CU-SeeMe as MeSee(what)UC

One of the important differences between our 'real' face-to-face discussions and those through the desktop or in a video conference studio is that, via the feedback screen, we can monitor what the other can see of ourselves, and *vice versa*. This is of course quite an extension to our normal awareness of the other. Schutz writes:

First of all, let us remember that in the face-to-face situation I literally see my partner in front of me. As I watch his face and his gestures and listen to the tone of his voice, I become aware of much more than what he is deliberately trying to communicate to me. My observations keep pace with each moment of his stream of consciousness as it transpires. The result is that I am incomparably better attuned to him than I am to myself. I may indeed be more aware of my own past (to the extent that the latter can be captured in retrospect) than I am of my partner's. Yet I have never been face to face with myself as I am with him now; hence I have never caught myself in the act of actually living through an experience.

(Schutz 1980, 169)

CU-SeeMe with its monitoring window and other video conference systems with feedback monitors make it possible for us to see ourselves as the other

sees us. However, we cannot hear ourselves as the other can; also, this 'seeing' differs qualitatively from a face-to-face situation. Therefore, in video conference surroundings the idea of non-identical experiences is challenged to an extent, because, as there are two video frames on the computer screen, it is not only a question of CU-SeeMe (See-You-See-Me) but also of What I See is What You Get. Of course this is still far from 'being in another's shoes' but we know almost exactly what the other sees of us in the situation. However, we cannot know what the screen on which our video picture is situated looks like, nor do we know if the participants at the other end of the video link are looking at our video picture, though we have some idea of their circumstances if we monitor their video picture⁴⁹.

According to Peirce, all the signs which have the feature of likeness can be classified as icons (see e.g. Nöth 1995, 121 for a good summary). If we follow Peirce's terminology, we can see that the video pictures are icons: they remind us of the real situation and carry something about it with them, "the signifier looks or sounds like the signified". (Fiske 1990, 46). Now, there is a difference between the two video pictures in terms of accompanying audio channel: the other's sound is mediated, but the participants mostly do not hear their own voices as mediated. However, sometimes the signal is slow and they can hear themselves at the other end. It is in these situations that the speakers can be assured of what in face-to-face situation is taken for granted: their voice can be heard by the other.

In a fairly stable communication situation, the sending picture will remain pretty much the same all the time: there are small heads on one's screen, including that of oneself, and it is possible to monitor the video picture of oneself as well as that of the others on the screen. This is different from the situation when there are more people in a room (e.g. classroom) and somebody is acting as a 'camera person', modifying the view that is being sent: it is necessary to attend to the 'information value' of one's own video picture much more, as the situation is strange compared to a face-to-face one: it is much more uncertain what the other's view of us is. However, once the picture of oneself is seen on the screen, one has an identical image to the receiving end of oneself. Although there is uncertainty about what the other's point of view of the space is, once the sending picture is attended to, this uncertainty changes into a glimpse of how the others see us as real objects, something which we cannot normally experience. Heath et al. (1995) report on their studies of collaborative environments in which video was used to enhance working together in separate locations. They describe difficulties participants had in orienting to the sending picture as the one the other end can see. However, an attentiveness to the other's view of us seems to be crucial for acting together in a shared space. This awareness is then incorporated into video-mediated environments via a feedback monitor. So

⁴⁹ Actually, as reported by Heath et al. 1997 (324), people might prefer looking at their own video picture rather than at the face of their coconversationalist.

especially in video conferencing with a changing camera view (i.e. altered by someone else than the participant), information of one's own video picture is required because otherwise, unlike in a face-to-face situation, it is impossible to estimate how the others see oneself.

7.5. Data analysis

The observations discussed in 7.5.1 come from the first session, a broadcast with a fixed camera view (2,5 hours of data). In Section 7.5.2, orientations to the incoming picture or the sending picture are described on the basis of the interaction when the camera view was zoomed in and out to give group shots or head shots of specific individuals (5 hours of data). A separate entity (Section 7.5.3) concentrates on how writing on the video screen was used as resource in the ongoing interaction (with both stable and changing video picture). During the seminars, there were both pre-allocated turns-at-talk (introductions, talks) and on-line, i.e. 'conversational', informal and unpredictable actions.

7.5.1. General observations

In their interaction, the participants usually constructed the situation as happening in two places: *here/there, our/your, this end, that end, (you) in Finland* etc. (of course the binary indexicals could also be used in a shared space, e.g. when there are two groups in a big hall). Only the teacher would use the inclusive *we* when he was outlining what would happen next in the seminar (*maybe we could...*). The audio connection worked such that only one end at a time could talk, and a fair amount of typing on the single scrolling line on the video picture was done as a way of communicating. When in the very first seminar the Finnish students had to come to the front to introduce themselves to the students in Sweden, most of them would look at the screen instead of the camera above the screen. Students would learn fairly soon that looking at the camera gives an impression of looking at the recipients of the picture, but at the same time they did not seem to understand that the sending picture is the one the Swedes can see: *ois hauska tietää että miltä me näytetään*⁵⁰ ('it would be nice to know what we look like') (cf. Heath et al. 1995 in which

⁵⁰ The comments the students made in Finnish could not be heard in Sweden. They are picked from the discussions between the Finnish students during the seminars, i.e. they were not elicited in separate interviews, for example.

similar observations were made)⁵¹. There were some comments made about the appearance of the Finnish students⁵²:

7(1)

- 1 S-All: **hah hah hah ha**
2 (1)
3 F-Harri: **hm** ((through nose))((smiling, turns to Manna))
4 °(nuilla)° on ihan **hauskaa**/ °(Those guys)° have **fun**
5 ((turns to)) **siel[lä]**/the[re]
6 F-Meerit: [°(kyl-lä)°]/[°(ye-es)°]
7 F-Harri: (.) **me istut(aan [vaan] täällä,**]/**We are**[(just) sit(ing) here,]
8 [(hands between the thighs, cramped position;
9 turns head, nodding, from Meerit to the screen))]
10 F-All: [(laughter))]

At times the audio connection would be cut and the students in Sweden and the teacher would discuss what to say next. This, and especially the fact that they could not hear their teacher all the time, was commented on as a negative feature of the setting: *tuntuu niin kuin olis mukana muttei kuitenkaan ole mukana* ('you feel as if you are and are not involved at the same time'). The silent video pictures from Sweden were rather agonising for the Finnish students because of the lack of sound. Thus, the video picture did not seem to convey enough information to Sweden either, as the teacher at the end of one seminar session addressed two students who were not there at all! The video picture was not a reliable conveyer of the other place: people might be out of the frame. According to Bakhtin, "if existence is shared, it will manifest itself as the condition of being addressed" (Holquist 1990, 27). The teacher speaking to absent students is tolerated in a virtually shared space; had he talked to non-participating students in Sweden, his status as a fully-fledged member of a teaching profession might have become questionable.

In Example 7(1), the sending video picture is objectified to make comments about what the Finnish crowd looked like. Towards the end of the session, the Finnish students would try out the slow motion of the sending video picture, moving a leg or an arm and noting that a big, fast movement would not show on the screen at all. This shows that the feedback video screen is extremely important for the participants to understand how their actions appear to the others; the presupposition of a common frame of reference cannot be assumed though one can see each other via a video link. For instance, Heath and Luff (1991, 102) mention the difficulty to estimate the

⁵¹ It is understandable that the extension of one's normal awareness of the other is hard to understand (see the quote from Schutz above).

⁵² Transcription conventions are in Appendix 1-1

impact of one's gestures, smaller of which tend to go unnoticed, and dramatic ones become "larger than life". In the present data, the size of the gesture (and thus its extremity) might be lost altogether if it is done too quickly: the connection only broadcasted slow enough gestures.

The Swedish video picture was also treated as an object: the Finnish students would point at the picture and make whispering comments. They would make remarks about not being able to see the expressions on the faces of the Swedes though there were some close-ups. The video picture had a slight time lag, resulting in incongruent sound and picture, which explains why some expressions, though they could be seen, would come too late for the words.

There were comments on (objectifying) the sound as well: *rupeaa keskittymään vaan siihen miten se katkoo tai miltä se kuulostaa* ('you just start concentrating on how there are cuts in the transmission or what it sounds like'). In general, the sound was clearly very important for where the Swedes were 'located' in the room: even if a student was sitting with his back to the monitor at the front, he would point behind him when he made comments about 'them starting again', though there was a screen next to him displaying exactly the same video pictures as on the monitor behind him.

The students were involved in a seminar with one topic and shared time but separate locations, and the other location was visually accessible through the video picture. Sometimes communicating through video conferencing resulted in specific orientations of gaze and posture in the room; the participation framework of the Finnish students near the computer with a sending camera (No. 1 in Figure 7-1), loudspeakers, and video picture to type on was clearly more of sharing the space. To give an example, at one point, two female students turned their heads away from the screen to the speaker in Finland at exactly the same time; a closer examination showed that the head turns coincided with a student in Sweden in front of the camera lifting her gaze from her papers up to the camera. This was a strong indication that the students interpreted the videoed picture of the Swedish student in the same way as in a shared space with a non-active participant: we know that gazing at a non-active participant can cause discomfort to the target of the gaze (e.g. Heath et al. 1997, 328). Maybe this was also an explanation for the strange participation framework (for a participant in a video conference) of the two students.

7.5.2. Orienting to the incoming and outgoing video pictures

The analysis supports the observations given above that, depending on the material setting of the room and the position of computers/camera/loudspeakers, there will emerge different zones of interaction also within a 'real' room.

i) Orienting to the incoming video picture

Saana in Example 7(2) was clearly orienting to the incoming picture (both of the video pictures can be seen on the screen of the computer) as she reacted to changes in the picture (e.g. line 3). Saana occasionally glanced to the front of the room, especially when she was listening or if she was puzzled. The picture on the left in the transcript is the Finnish frame, and the one on the right Swedish.

7(2)



Finland



Sweden

- 1 F-Saana: (---) **but (.) I thought this would be an easy extract (.) to**
 2 **analyse but it wasn't so easy (.) °as I expected°.** ((glance at
 3 screen, gaze up towards camera & down to microphone,
 4 [up to screen smiling, points at screen)) °mitä nyt?° / °what
 5 now?°]
 6 |
 7 S-View: [((camera zoomed to Katrina:))]



- 8 F-Saana: ((glance down to microphone, back to screen, smile))
 9 (3)
 10 S-View: ((panning away from close up, towards ceiling))
 11 F-Saana: **hmph!**

In a way Saana forgot about the Finnish students and devoted most of her attention to talking to the Swedes; her participation framework excluded the Finns. That the orientation was picked up at the other end of the line, as well, became clear when Saana gave directions to the seminar participants: *okay, read extract number two*. A student in Sweden replied *and in suomi as well*. Only sporadically did Saana's gaze meet that of the others in the room, and this

happened usually when there was a problem in the connection or in her understanding (cf. Extract 7(8)). Saana was very observant of how the camera was moving or what movements she could see in the Swedish location: *mitä nyt?*/'what now?' with a pointing finger (line 4) shows that the camera in Sweden zooming to a student is meaningful to Saana; she just does not know what is going to happen. Her *hmpf!* (line 11) indicates that she is not happy about the camera movement towards the ceiling, and losing the visual contact to the Swedish group.

Example 7(3) shows how the emerging participation framework fluctuates so that talking to Sweden is done either by looking at the screen or by the posture and gaze being oriented to the group in Finland.

7(3)

- | | | |
|----|-----------|--|
| 1 | S-Karina: | er I would like to speak to Marjo but she isn't there is |
| 2 | | she? |
| 3 | | (1) |
| 4 | F-Saana: | [((head shake; glances Santtu; silent 'no'))] |
| 5 | F-Santtu: | [((gaze to Saana and back to screen)) no no she is not] here. |
| 6 | S-Karina: | so (.) could someone tell me what \nsubseteq writing [about?] |
| 7 | | |
| 8 | F-Saana: | [((turns head away from screen; gaze to front))] |
| 9 | | ((eyes to right and then to meet Mirja's gaze on the left)) |
| 10 | S-Karina: | do you know? |
| 11 | F-Saana: | $^{\circ}$ mää oon opponettina $^{\circ}$ /'I'm the opponent' |
| 12 | | [mutta en tiijä] /but I don't know] |
| 13 | | [((eyebrows up, quick head nods))] |
| 14 | | |
| 15 | | [((smile))] |
| 16 | | [((gaze from Mirja to Molla and back to Mirja))] |
| 17 | | |
| 18 | | ((turning to Tarja)) mistä Marjo tekkee? /what's Marjo's |
| 19 | | topic? ((clears throat)) |
| 20 | | [((turns to Molla))] |
| 21 | | |
| 22 | F-Molla: | [$^{\circ}$ (eikö) Marjo tee Kosmoista mutta mitä ()] $^{\circ}$ / |
| 23 | | [$^{\circ}$ (isn't) Marjo doing hers on Cosmo's but what ()] $^{\circ}$ |
| 24 | F-Santtu: | ((mic twrds mouth)) we(II) we are not really ((glance: cam.)) |
| 25 | | sure here er: (.) w- what she is writing about, |

The question in lines 1 and 2 is different from the switchboard requests on a telephone line of the type "Is X there" (Schegloff 1979, 41): Karina gives the conclusion she had made on the basis of the visual information available that Marjo was not in the Finnish room. The teacher in another session used a formulation which was more similar to the telephone one: *is jonna there today?*. Both formulations would polarise the situation taking place in two places (rather than, for instance, asking if a person was 'with us'). The Swedish student's turn design with *someone* (line 6) means that she is not necessarily asking Santtu to answer the question though he is holding the microphone.

This is also demonstrated by Saana's response: Saana seems to be very alert to the question (which was later explained by the fact that she should have known what Marjo is writing about (line 11)). Saana's head shake and silent *no* (line 3) are a skilful way of informing at the same time both the Swedish end and Santtu (Saana glances at him) that Marjo is not around. Of course Santtu did not need to be confirmed, as he could see for himself that Marjo was not there. So, Saana's contribution has a double interactional function: to prompt Santtu to say *no* and to visually convey to Sweden an answer. Both roles are emphasised especially in the silent 'no': it is not meant to be heard (but seen; note that she 'says' *no* in English, not 'ei' in Finnish) in Sweden, and it is like whispering to Santtu his next line.

In line 5, Santtu is talking to the screen, but thereafter he is looking around the room and when delivering the answer (line 24), glancing at the camera as well. This seems to be due to the previous activities: from directing his words to the screen, Santtu is engaged in a conversation or monitoring conversation between his fellow students in Finland. He continues talking to the Swedes with his head up from the screen. Thus, that his answer is a summary of the negotiation (lines 8-13) is shown not only in the use of *we*, but also in his bodily orientation in the copresent space; he is still with the Finnish group.

In the beginning of Extract 7(3), the Swedish student is asking whether a certain student is present in Finland. This had to be done because she could not be sure whether there was somebody else in the room, outside the scope of the camera lens. Also the teacher checked often who is or is not present (students might come into the room late and not be visible in the picture): there was a continuous uncertainty of who were the participants in the session. The awareness of incomplete visual information also of oneself would effect the use of indexicals (see below).

The uncertainty that the video and the audio link created for the participants in the allotting of turns, and turn junctions in general, was made visible in the way the students and the teacher resort to explicit turn signals such as *over* at the end of their turn even if the material situation was such that it is not needed. Extract number 7(4) comes from the last session in which *over* was not generally used because the regulating of the audio line was undertaken by switching the microphone on and off to stop the echoing effect. Notice how at the very beginning, Kristen does not start speaking until the Swedish video picture stops moving and consequentially the movement in the Finnish picture ceases (*er* in line 1 indicating her willingness to start a turn).

7(4)

- | | | |
|---|------------|--|
| 1 | S-Kristen: | (ER) ((video picture moves to left, up, down)) |
| 2 | F-Berit: | ((hand with mic twds Harri)) ((smile)) |
| 3 | S-Kristen: | () WANT TO KNOW (.) CONSIDERING (THAT) THERE |
| 4 | | ARE MOSTLY <u>TEENAGERS</u> READING THIS |
| 5 | | NEWSPAPER IF YOU REALLY THINK IT'S WRONG |

6 NOT TO TRY TO (.) HELP THEM WITH THEIR SELF
 7 CONFIDENCE AND STRENGTHEN THEM A BIT (.)
 8 EVEN IF IT'S A BO- (.) PERHAPS EXAGGERATED TO (.)
 9 FREE YOURSELF OF EVERYTHING BUT (1) (IT)'D BE
 10 GOOD IN A WAY (WHEN) CONSIDERING WHEN
 11 ∅('RE) SO YOUNG? (.) OVER(h) ((smiling))?
 12 F-All: ((laughter))

In Example 7(4), there are several transition relevance points (TRPs) in Kristen's turn. The participants in Finland were motionless; there was no microphone movement and they did not look at the camera during Kirsten's long turn. We can therefore assume that the use of *over* (line 11), as a clear marking of giving the turn to Finland, is added to make sure the audience understands what is going on, i.e. that Kirsten selects Beerit in Finland to be the next speaker. Thus, the incoming video picture had an impact on the turn design and lexical choice. It is obvious that the Finnish party regards *over* as funny because they start laughing after it was produced; the actual question is not amusing. There might be different reasons for why they find *over* entertaining, but laughter is too strong a reaction to just a smile which is visible on the Swedish student's face.

Over, together with *your turn*, was used for explicit turn allocation, both verbally (with voice) and through writing on the video picture (e.g. *YOUR TURN* and *SUOMI OVER*). Turn boundaries were negotiated and specifically marked in the first session in which the audio line of the CU-SeeMe program was used, an option which turned out to be too slow even for a fairly formal seminar. (In the later sessions, a separate program (Maven) was used for the audio link.) In the session, the use of *over* started filtering to other talk, as well. At one point, somebody in Finland answered *over* to their fellow student's question whether there were any more replies to a question from Sweden, and the person with the microphone then said *over* as an answer to Sweden. When the students in Finland were waiting for the Sweden to join them, in the discussion of the Finns, *over* was used as a humorous way to mark the end of a turn. Here we have a living glimpse of how technologies can shape language use: the 'cyborg' register develops if we understand 'self' as socially constructed in communication also with mediated 'others'.

ii) Orienting to incomplete visual information of oneself in the sending picture

Again, the video link had an effect on how turns were formulated; the technology shaped the interaction. As not everybody was necessarily seen at the other end through the videopicture, the virtual classroom was extended to the non-seeable one by the speakers introducing themselves if they were not in the picture.

7(5)

- 1 (8)
 2 F-Santtu: **well this is Santtu >°here°<. I think I(h)- Kerttu hah**
 3 **running zooming the camera**

In Extract 7(5), Santtu is not in the picture at all (not even in the background) and he notes that his image is absent and wants to be in the front region, not only orally, but also visually⁵³. The use of deictic expressions shows how the participant did not treat the 'virtual space' as mutually accessible. His use of *this is Santtu* replaces the missing indexical image, and *here* is constructing the situation as two separate sites connected, but not sharing, a space via technology. The comment about Kerttu's activity reveals that he knows his face is soon to be visible in the sending video picture.

If the students thought that they are seeable and thus recognisable, they did not introduce themselves, but knowing who is speaking is important in a seminar: the responsibility of one's words belonging to the carrier of the proper name: "It is only from that site that we can speak" (Holquist 1990, 167). Thus the turn-at-talk of Santtu was shaped according to the status of the sending picture.

A bit later, the Swedish end swapped speakers, Kristen leaving the seat in front of the camera and Karina coming forth to sit down.

7(6)

- 1 S-Karina: **hel>lo<?**
 2 F-Santtu: **hi,**



Finland



Sweden

- 3 S-Karina: **hi. this is ()- ∅**

⁵³This also relates to what Goffman has called foreground/background, the emergence of which has been reported by M.H. Goodwin (1995) in a study of an airport working environment.

In Extract 7(6), the participant in Sweden moves to come to the front of the room to be closer to the camera. In this case, she is moving and introducing herself with *this is*; sometimes the same formulation would be used to make other participants move out from between the camera and the speaker: *okay this is - this is Phil here now* (after which the Swedish student moved away from in front of the camera). *Now* emphasised the speaker change more than 'nowness', or time, and this same interactional work got the student out of the way. Thus, the state of one's visibility in the sending picture affected the turn design.

By using *this is* the participants in Finland and Sweden oriented to the video picture; when the use of 'I' reflects "the uniqueness of the location of each body according to the principles of materiality, such that only one body be at a place at a moment, and that no body can be in more than one place at the same time" (Harré 1991, 59), this is reflected the distortion: the 'I' of the speaker was also on the screen, and for the addressee, only on the screen. The shared location of the screen was constructed through language use, as well.

iii) Orienting to sound

The sound on the audio channel of the video conferencing system was also an important resource for ensuring that the audio connection was working: if the students heard the echo of their own sound coming back from the other end, then they could be sure that their voice was heard there. However, sometimes this echoing, the slow transferring of sound on the net, would result in a virtual 'self' through the audio channel. Thus, not only could one see oneself as a mediated video image; one could also hear one's mediated voice:

7(7)

- | | | | |
|----|-----------|---|-------------------------------------|
| 1 | S-Maria: | (| NOW) |
| 2 | | (.) | |
| 3 | F-Tarja: | excuse me- | |
| 4 | S-Maria: | (| ABOUT) |
| 5 | | (.) | |
| 6 | F-Tarja: | excuse me (.) could you speak a little more slowl(ier) (.) we | |
| 7 | | can't e:r quite (.) understand you. | |
| 8 | F-Kerttu: | and e:r- and er vesa is complaining about a- echo in °his (.) | |
| 9 | | room°. ((gives microphone back to Tarja, whispering | |
| 10 | | something)) | |
| 11 | F-Tarja: | ((to the camera:)) okay, | |
| 12 | S-Maria: | (| IF I) SPEAK SLOWLIER. IS THIS GOOD? |
| 13 | (F-Tarja: | okay,) | |
| 14 | S-Maria: | OKAY, (HEH) (M:) | |

Tarja's second *okay* (line 13) is heard from Sweden as a delayed sound. To Maria, it sounds like an answer to her question (line 12), which is shown by

her repetition of Tarja's 'answer'. So, in this case the resource of hearing oneself at the other end results in Tarja and other students in Finland listening to the virtual Tarja participating in a discussion. Thus, a 'real' voice from the immediate past is heard as an answer to an indexical question *Is this good?* (*this* referring to the pace of the talking). That *okay* does not have a falling intonation must be the source of amusement as it is repeated with the same intonation (line 14). The consequence is laughter at the Swedish end of the line. The impact is different from that of the delayed written messages ('voices of the past') in computer conferencing systems which, according to Severinson Ekhlund (1986), create uncertainty because the topic can never be regarded as closed, i.e. anybody can send a new message to a discussion, even weeks after it first took place.

It is interesting that Tarja formulates the trouble as concerning the whole group (lines 6 to 7). Her shaping of the utterance has potentially multiple meanings, as the request to talk more slowly and the mentioning of understanding could be read as 'we cannot understand English spoken fast'. Later on in this situation Tarja used the form *can you repeat?* She got another explanation, after which Tarja reformulated what she thought was meant: *so (--), did you mean that?* Again, later, after hearing a question she checked *do you mean...* And, *sorry, who were you talking about?* After this Tarja made accountable the uncertainty that students feel about having to check by saying *se luulee etten mä tajuunnu* 'she thinks that I didn't comprehend'. So, the bad sound connection in this *lingua franca* situation resulted in mishearing or not hearing something. The fact that the students were non-native speakers of English would make requests for repetition topical in the sense that there was an inherent uncertainty, in spite of the intermittent bad audio connection, over whether the students did not understand instead of not hearing.

The choice to use *we* instead of 'I' about hearing problems was clearly illustrated in the following extract, line 12, in which Saana, a bit hesitantly, restarts her turn and emphasises the mishearing.

7(8)

- | | | |
|----|-----------|---|
| 1 | F-Saana: | ((microphone towards mouth)) |
| 2 | | [((glance to front))] |
| 3 | | |
| 4 | S-?: | [WE] JUST WANT TO $\nless{}$ (KNOW IF WE ARE) |
| 5 | | [RIGHT? HAH HAH HAH] |
| 6 | | |
| 7 | F-Saana: | [((knits eyebrows; mouth open))] ((gaze to front)) |
| 8 | | <i>mitä?</i> /what? ((gaze to a student, to left, to |
| 9 | | [Kerttu])) |
| 10 | | |
| 11 | F-Kerttu: | [say sorry jos et(h)-/say sorry if you do(h)n't-] |
| 12 | F-Saana: | ((gaze down)) sorry °I-° we- we can't <u>hear</u> you. ((gaze to screen)) |

In this extract, it is notable that Saana changes her formulation into *we* even if the preceding (mixed language) 'order' by Kerttu (line 11) is shaped to make the hearing problem Saana's alone (*et* is the 2nd person singular form of 'no'). However, the use of *we* is 'authorised' because Saana does not get any answers to her question on line 8, i.e. nobody else in the room shows that they had made out what was said.

The use of *we* about hearing problems as concerning the whole group would sometimes be flouted by other members of the group. For instance, at one point Santtu said *could you repeat the question because we didn't hear you?*, but after a short pause Tarja started to answer *well, I think..* which demonstrated that she must have heard the question.

Importance of sound

Sound seemed to be a more important indicator of locus than the picture. As discussed above, the students would refer to the computer with the loudspeakers in the front of the room (on the left in Figure 7-1) as 'them', or orient to this computer when in a listening mode, as in Example 7(2). Moreover, when the students in Finland were waiting for the Swedish end to start talking to them, there were cascades of attention (one student turning and the others following) to the computer with the loudspeakers. This shows that the front of the room was oriented to as if it was 'there' (in Sweden), as a virtual space within the room in Finland. As a result, the room in Finland with the computer to which the loudspeakers were connected, which was also near the camera, seemed to create a zone which affected the participation frameworks in the room. Proximity to this zone gave a feeling of one being more a participant in the situation. In contrast, proximity to the other computers gave a feeling of being more of an observer, with the result that even if one was actively participating in the seminar, one would be orienting to the incoming picture.

Features of telephone conversation

This subsection will deal with those features of interaction in the video conferencing sessions which echoed telephone conversation characteristics. First, a typical feature of telephone conversation openings and contact rechecking, viz. *hello?/hullo?* is discussed.

Both formulations were used during the seminars. A distinction has to be made between *hello* (flat intonation or falling intonation) and the rising or question intonation format. The first type was used as a greeting before introductions, for example in Extracts 7(18) and 7(19). In Extract 7(25) below, *hello* was used together with the name of the addressee, as a greeting during the seminar. The first two examples of *hello* took place during the beginning of

the first video seminar; the students at both ends introduced themselves (by sitting down in front of the camera), and the greeting was a general one with no specific addressee, a situation that could be conceived as happening in a shared seminar as well. The latter one was part of a series of ‘commentator introductions’ (the teacher told his students in Finland that *I’ll just introduce you to them*). This took place at the end of the first seminar. Thus, the students had been introduced as persons already. Compared with a face-to-face situation, the formulation *hullo, santtu*, however, sounds like a first meeting, and that in fact it was, between Santtu and Kristo. Also the first student introduced as a commentator greeted Santtu: *hi santtu*. After Kristo’s turn, the teacher said *hullo?* and *here’s ulla..* who started her turn *I’ll be watching the silence of the lambs..*, i.e. no greeting included. Unlike Kristo and Kristen, Ulla did not move up from her seat to talk, but she looked at the camera and spoke from where she was seated. So, it looks like movement, foregrounding oneself, is related to formulating greetings in the ongoing seminar.

Hello? and *hullo?* can also work as a summons, for instance, when walking into somebody’s (empty) space (“*hullo, anybody home?*”). And in a telephone conversation, *hullo?* during the ongoing talk is used if in doubt about the sound connection. The ‘checking sound connection’ use of *hello* (with a rising or flat intonation) was a common phenomenon (cf. 7(10), line 1).

The following extract shows how sometimes both the ‘greeting’ and ‘sound checking’ qualities of *hullo?* would be oriented to: Santtu is prompted (line 3) for a second pair part of an adjacency pair after a compliment (line 1). The prompting is done with *hullo?* which can be heard as ‘do you hear me’ or as ‘hello’ (as in a summons-answer/greeting sequence). Santtu treats it as the latter by returning it, and as the former by indirectly giving an account to the fact that he did not say anything: *what did you s(h)ay?*

7(9)

- | | | |
|---|-----------|--|
| 1 | S-Karina: | (very nice) speaking to you. |
| 2 | F-Santtu: | ((gaze to Molla)) |
| 3 | S-Karina: | hullo? = |
| 4 | F-Santtu: | =((gaze to screen)) hullo, e:r what did you s(h)ay? |
| 5 | | [heh heh] |
| 6 | F-Saana: | [t(h)h:] |

So, *hullo?* was treated as a summons, though it was different from the re-checking contact (Hopper 1991, 223) ones because the parties were visible to each other and had been in the conversation already. The exchange of hellos looks like a typical telephone conversation opening, though unlike in telephone conversation openings, *hullo?* was not used for recognition of any of the parties (Schegloff 1979).

However, sometimes *hullo?* was produced (by the teacher) after a long pause in the connection, for instance when the two groups were discussing

something separately at each end. That *hullo?* was treated by the other end as a hearing check which can be seen from the following extracts.

Extract 7(10), in which both typing and talking were used, shows how *can you hear us?* starts becoming a regular way of ensuring whether the audio connection is working. In line 3, after having waited for 14 seconds for the Finnish end to start speaking, Marion starts writing on the screen. Though he only types one letter, *c*, before stopping, it is quite clear that he was going to write the *can you hear us?* sentence, or at least something about the connection as he stops once a Finn starts speaking.

7(10)

- 1 S-Marion: **it's all yours**
- 2 (14)
- 3 S-Marion: (*c*)
- 4 F-Kerttu: **no comments from here so far, (2) over,**
- 5 (14)
- 6 F-Kerttu: YOU MUST GO ON, NO COMMENTS FROM HERE
- 7 (47)
- 8 S-Phil: **hullo? (1) [so there're no comments from that end?]**
- 9 |
- 10 F-Kerttu: [WE CAN]
- 11 <- <- ┘ NO ┘
- 12 S-Phil: **about**
- 13 F-Kerttu: YOUR VOICE KEEPS SWITCHING OFF ((deleted after a long
- 14 while; teacher talking and clearly not orienting to the message))

In line 6, the Finnish end shows the same orientation to Sweden: the 14 seconds of silence after the explicit marking of turn allocation (*over*, line 4), Kerttu types her request to the screen (though it scrolled away when the explanation and at the same time repetition of the spoken turn filled the picture). Phil's summing up (*so*) query in line 8 about there being no comments from Finland, with a questioning intonation, at the same time orients to the Finnish end as an active group and to the uncertainty of incorporating text on the screen into the ongoing interaction as a relevant turn from the producers of the text (as some time had passed since it was typed in). Kerttu's overlapping typing in line 10 is the beginning of a reply to Phil's *hullo?* as checking if they could hear one another (i.e. 'can you hear?'), but as he continues after a short pause, its interpretation changes from a hearing check to a summons and Kerttu deletes her typing so far to answer the question.

The teacher used *hullo?* frequently when re-establishing the contact between him and the Finnish students, e.g. *hullo? right.* (- - -) and *hullo? in suomi?* ((looks up to camera)) (- - -), and also *hullo? and here's kristen* (- - -). It seems that *hullo?* was used equivalent to *okay* or *right*, (which was chained with *hullo?* above). As shown in 7(9) and 7(10), unlike with *okay/so/right*, there was a tendency to reply to the summons *hullo?* The teacher used

hullo?/hello? like ‘okay/so/right’ that did not require a second pair part (he did not wait to hear it), but which at the same time oriented to a possible trouble in the sound connection. And indeed, the same lexical item was used at times of trouble in the sound connection, as shown by Examples 7(28)a and 7(28)b. It seems that *hello?/hullo?* at the junctures in the topic or connection would guarantee also that there was a visual feedback from Finland (the students would look up at the computer at the front as discussed earlier): they would know that *hello?/hullo?* was aimed at them, not at the Swedish students, and the teacher would see that the summons had an effect in Finland and thus was heard there.

Within a turn, okay/so, instead of *hullo?/hello?*, was used by the teacher. For example:

7(11)

- | | | |
|---|-----------|--|
| 1 | S-Phil: | (- - -) any questions. okay. so, ((teacher talking to Swedish |
| 2 | | students, nobody looking at camera/computer in Sweden)) |
| 3 | F-Santtu: | (w'1) this is [Santtu], can you hear me? [yeah (h).] okay. |
| 4 | | |
| 5 | S-Phil: | [[((turns to camera/computer))]] [[((looks at camera/co., smiles))]] |

This extract is also an example of the use of *this is*. As in other occasions (e.g. Extracts 7(5) and 7(6)), it is used to introduce oneself. And also in Extract 7(11) the question is about the speaker’s visibility: though Santtu is in the video picture, he is not visible to the Swedish end simply because they are not looking at the Finnish video picture (and he can see it).

This is X as a way of introducing oneself can also be found in the opening phases of telephone conversations (Schegloff 1979), in which visibility and therefore speaker identity is not immediately available. Interestingly, the format ‘it’s X’, (not to speak of ‘it’s me’), frequent in telephone openings (cf. the telephone opening referred to in 4.4: *hello it’s alison here*), was not used in the video conference. The reason could be that ‘it’s X’ is more addressee oriented, in a way answering the question ‘who is it that rings (or knocks on the door)?’, whereas *this is X* describes the summoner from their own perspective. Also, *it’s X* assumes that the addressee knows the person, whereas *this is X* (or *my name is X*) is a formulation used when the person is not known (e.g. Schegloff 1979, 46 and Section 7.6.2 below), i.e. in introductions. Thus, the use of *this is X* in the video conferencing not only was recipient designed in orienting to the missing visual information (as conveyed by the video pictures), but also to the possibility that the voice of the speaker would not necessarily convey the identity as the situation was from many-to-many. *My name is X* could not be used again as it is only used once in the first official introduction (see Section 7.6.2.). ‘It’s X’ does not get used because the saying of the name is at the same time the summons, i.e. there was no telephone ring or knock on a door which then would be explained as a deed of an invisible (but known) summoner who has to identify oneself as ‘it’s X’.

Also, in a situation in which there are several recipients, *this is X*, comes from the aspect of one person; 'it's X' would be orienting to several persons' question about the speaker's identity. At the same time, it can keep the communication one-way instead of reciprocal: it is easier to speak from one's location, rather than trying to incorporate the others' point of view to the talk. All these formats of announcing oneself, would, however, not be used in a face-to-face situation.

Thus, the use of *this is* and *hello* both constructed the common seminar as one which is mediated by a limited visual and aural channel. At the same time, they occupied places in interaction which were 'another beginnings', either of introducing a new speaker or a new topic, or simply getting the attention of the mediated 'others'. When the same lexical items in telephone conversation openings were used in the identification sequence due to lack of visual access (Schegloff 1979, 24-25), their use constructed the seminar as one of continuous reintroductions: the contact and the identities of the multiple participant video conference were not as self-evident as in face-to-face communication.

Summary

Complicated participation frameworks emerged because of the material setting in the room. The virtual classroom between Finland and Sweden also affected interactions within the room in Finland. A very typical feature was the prompting and negotiating of answers or comments to be given to Sweden: the students were mutually constructing the next turn. So the students (and the teacher's helper) regularly told others to say something or the Finns would confer about what to say to Sweden or discussed what they thought the Swedes were asking or commenting about. *We* in the video conferencing situation did many types of work. It would mark a turn as a result of negotiation, especially if the speaker was part of the group. Using *we* was also used to distinguish hearing difficulties from understanding difficulties, though there was no preceding negotiation (cf. Extract 7(7)), or even if the speaker of the turn had been prompted to talk for herself (Extract 7(8)). Every instance of *we* (and *you*) would be polarising the situation.

Since the 'virtual' seminar took place in two geographically remote locations, the other of which was seeable via a video picture, for the Finnish students who were speaking there were two options to orient to those present in Sweden with a gaze: either to look at the camera or to look at the video picture showing the Swedish students. In the first case, the current speaker would not be so much distracted away from other Finns, and the glances at the camera would be seen at the Sweden site as glances at them: this would be the 'normal' way for a speaker to behave (i.e. not to look at the addressee all the time). However, the tendency was to 'talk to the screen', and thus forget

about the audience in Finland. Only if an answer was negotiated by the whole group in Finland, would the gaze of the current speaker shift from the screen to the camera. Therefore, the attention structure of an active participant in Finland was geared towards Sweden, unless their turn was negotiated in Finland.

Video conferencing can also influence turn design and lexical choices. A motionless video picture from one of the sites would enhance marked turn allocation, especially the use of *over* to give the turn to the mediated 'others'. The video picture was an uncertain carrier of the visual information to the other site (the speaker can be out of the picture or moving into the picture), which was reflected in the speakers' orientation to their video picture when he or she 'introduced the current speaker' by formulating the turn beginning as *this is X (here)*. Another lexical item typical for telephone conversations, *hello?/hullo?*, was doing the work of summons and line check. This resembles the division Streeck shows for summoning in a classroom: if the person is not present, the classmates will treat the summons as an information question and "respond by an informative statement" (Streeck 1980, 143). In the present data, the summons was simultaneously getting attention and inquiring about the audio connection.

Often the virtual visual depiction of the participants was accompanied by virtual sound: when the sound connection was slow, the turn of a Finnish speaker could be heard 'conversing' later at the Swedish end.

When video pictures of oneself appeared on the screen of the two computers on the right in Figure 7-1, they were treated as metonyms, as icons of oneself. Though not having a changed identity made possible by text or graphic based computer-mediated-communication systems, the students in Finland still could "construct and encounter other versions" (Strate 1997, 371) of themselves. This is important and interesting for research on so-called virtual reality (VR), as well, because we have to understand the effects of the visual interface to communication. When some research shows that incorporating both audio and visual stimuli will create a more intense sense of copresence than only having one of them available (Short et al. 1976), it is important to know how the significance of different media in actual encounters becomes visible: that sound is more important than video picture in my data goes against the normal hierarchy of visual over audio and audio over written media in creating the feeling of social presence (see Lombard & Ditton 1997, 16). My research confirms the importance of aural dimensionality in creating the sense of presence, on which, according to Lombard & Ditton (1997), not much empirical evidence has been available. It could be that the importance of the audio connection is strengthened because of the relatively small size of the video pictures which occupy only "a fraction of the visual field" (Biocca 1997, 10).

The designers of three-dimensional VR are hoping to free people from the limits of space by visually representing the participants with a computer

image in synchronous interactions. To be successful in this, they need to know how visibility combined with other forms of mediated 'others' affects face-to-face and video conferencing interactions. Though each computer screen in the room in Finland was in principle an object of attention, a space to orient to (cf. Kendon 1985, 237), the computer in the front was the one addressed.

7.5.3. *Typing on the video picture: The synchronous/asynchronous division in the making*

In CU-SeeMe video conferencing encounters, as conveyed by the video pictures in Sattler (1995), the possibility to type on the picture can be used to indicate the name(s) of the person(s) in the picture like on TV, or to greet the other(s) (*hello!*), or further to give a caption to the picture (the name of the place, what is happening etc.). Typed talk on the video picture does not disappear unless the 'enter' key is hit, or disappears slowly into the left border of the screen if the typist keys in a longer sentence. When typing on the video screen is used as a way of communicating on-line, as in the present case study, there is a close-knit connection between what is going on (in the video picture): the written message is produced for a moment in the interaction and not as a general persistent message. However, typed talk on the video picture does not disappear unless the 'enter' or 'backspace' key is hit, or if the typist keys in a longer sentence which disappears slowly into the left border of the screen. But, unlike text-only synchronous computer conferencing programs, also visual information is available for the receiver of the message. The time of the writing can also be the time of the reading, unlike in asynchronous written communication. Thus, typing on a video picture in on-line video conferencing differs considerably from, for example, e-mail messages which are received, edited and responded to asynchronously (cf. Bowers & Churcher 1988, 129). Like Talk in Unix, editing on the video picture is done on-line, which is unlike the Chat/Talk box in CU-SeeMe that allows for separate writing and sending of one's contribution. With the narrow, one-line space for the typing, it is visually linear, reminiscent of speech, because there is no trace of the precedent discussion (versus Chat/Talk box in which a trace of whole interaction is recoverable). Unlike the sound of the speaking voice which is available for the whole of the audience (even if they would not want to hear it), textual information, like signed languages, has to be seen to be integrated in the ongoing situation. Written texts are not usually immediately available for the onlookers, i.e. one has to concentrate on the computer screen if one wants to 'hear' what the others say. But sound can bypass even visual barriers.

7(12)

- 1 F-Saana: **hello, hello** ((high pitch))
- 2 F-Kerttu: **Saana huua si(nne) että we can't- we ca sano vielä**

3 **kerran**/Saana shout there that we can't- *we ca* say once
4 more
5 F-Saana: **we can't hear you. are you talking? talking to me?**
6 F-Kerttu: *n't hear you!!!*
7 F-All: ((laughter))
8 F-Kerttu: () (.) **mitä se sannoo siinä**-/what does it say on it- ((looking
9 at the computer with maven connection)) *you're not on mo* ↓
10 [*aven!*]
11 |
12 S-Karina: [*system erreo <- <- or*]
13 F-Harri: '**system error**'
14 F-Kerttu: **system error, missä.**
15 S-Karina: *we [can hear you]*
16 F-Harri: ['**we can (.) hear (.) you.**']
17 [**you**]
18 F-Kerttu: [**he- hei**,/he- hey] **kuuleeko ne meijjät.**/do they hear us.
19 F-Mirja: **kuulee**=/they hear us=
20 F-Harri: **kuulee**/they hear us
21 F-Kerttu: **you are not on maven (.) at all**, ((turns to look at the Swedish
22 video picture)) (.) **you are not on the list (.) phil?**
23 S-Karina: [*so you better <- <- <- <- <- <- <- <- <- we know .*
24 F-Harri: '**we know.**'
25 F-Kerttu: **why are you off?**
26 S-Karina: *Phil is re-connecting us. Tell us something about*
27 F-Mirja: '**reconnecting**'
28 S-Karina: [*the course while we are*]
29 F-Mirja: ['**tell us something about (.) the (.) course (.) while (.) we (.)**
30 **are (.)**']
31 S-Karina: *waiting*

The situation in Extract 7(12) is asymmetrical as far as receiving language from the other end goes: in Sweden it happens via sound waves which are simultaneously available to the whole of the audience; in Finland, the reception relies on the visual, on seeing a computer screen with (relatively small, see Appendix 7-2) video pictures and the text on them. This results in co-operation by the participants in Finland, one reading aloud essential bits of the information (line 27) or the whole of the turn when it emerges on the screen (e.g. line 16).

Though writing is considered as rudimentary if audio and video contact are possible, the data provides intriguing material for researching the multimodal use of and borderline between spoken and written communication.

i) Typed text ignored

Often a remnant of previous interaction was left on the screen for a while, text which did not become interactionally relevant for the receiver. For instance, when the students came to the room in Finland, the Finnish video picture had

the following question typed on it: *where are the masks?*. Nobody commented on, or otherwise oriented to the writing. However, sometimes a piece of text got ignored which was produced during the seminar and thus was clearly interactionally meaningful. For instance the Finnish helper typed *phil, we are running out of time!* at the end of the first seminar. The teacher did not react to the message (which became, due to the limited width of the video picture, an elliptical sentence), and it was deleted later when another attempt at getting the teacher's attention was made. In general, either meaningless or meaningful pieces of texts on the screen which were not integrated into the ongoing interaction, were, however, not 'noise' in the sense that they did not disturb what was happening.

ii) Double check

As there were sometimes unexpected cuts in the sound, many times 'sound checking', or a confirmation of sound check was done textually. This happened at the beginning of Extract 7(8), in which the work was done by two separate persons: Kerttu typing and telling Saana to 'shout there' that they cannot be heard. In the following extract, there was a request for the microphone to be put on, to which the answer was given, by one person, both verbally and by typing.

7(13)

- | | | |
|----|-----------|---|
| 1 | F-Saana: | open the microphone please, ((grin)) |
| 2 | S-Karina: | ((picture: speaking to the microphone, no sound)) |
| 3 | | [and: (.) could be both- two men: and a man an' woman, |
| 4 | | right] |
| 5 | | [[((picture: lips forming 'it's on' hands up, mouth open, |
| 6 | | moves to type] |
| 7 | | () mistake(n) (.) when I read it. I |
| 8 | | [thought it could only be a man and a-] |
| 9 | | [[((picture: typing <i>it's</i>] |
| 10 | (F-Saana: | [open the microphone please,]) |
| 11 | S-Karina: | [[((picture: typing <i>on</i>] |
| 12 | | [oh but it's on!] |

In the session from which Extract 7(13) comes from, the sound was heard considerably later than the movements in the video picture, as can be detected in line 10 where Saana's contribution in Finland (line 1) is heard in Sweden, with the ensuing answer in line 12 (which is visually produced already in line 5). The reverse case, video picture lagging behind the sound, seemed to be tolerated much better.

Also, more implicit turn allocations could be 'double checked' by typing the explicit form on the video picture. In the following extract, Meerit asks a question, which clearly is a first pair part of an adjacency pair that requires a

contribution from the addressee. Meerit picks the teacher as a specific person that her question was directed to. When there is no answer for 9 seconds, the teacher's helper types *YOURTURN* on the video picture, and after 13 seconds without any answer, she adds another format of telling the Swedish end that the Finns have no more to say (line 5).

7(14)

- 1 F-Meerit: (---) **do you understand me? Phil?**
 2 (9)
 3 F-Kerttu: *YOURTURN*
 4 (13)
 5 F-Kerttu: ↵ *SUOMI OVER*

In the following extract, the status of the first pair part of the adjacency pair is brought into doubt because no answer is coming from Sweden.

7(15)

- 1 F-Harri: **there was this point that (.) "she is as good as gold" (.) er it's**
 2 **like er (.) er he's talking to a child (.) or °something°. what**
 3 **do you think about that?**
 4 (9)
 5 F-Kerttu: *YOUR TURN*
 6 (9)
 7 *DID YOU HEAR THE QUESTION?*

In line 7, Kerttu queries about not hearing in general, but about hearing the question. She makes it clear that the expectation in Finland is that Harri's question should be answered.

Sometimes there is no question at all; the turn ends and if no reaction comes from Sweden, a distinct turn allocation is made:

7(16)

- 1 F-Meerit: (--) **and the way they talk about (.) male mo- models is (.)**
 2 **really sexist. so (.) there has been (.) some development.**
 3 (6)
 4 F-Kerttu: ↵ *OVER*

iii) Additional information

In the very first video conference session, the students introduced themselves first, giving their names and a short introduction to their interests and their research topic in the seminar. To enable the fellow students to see the name written down, as well as to hear and see the person in question, the teacher in Sweden and his helper in Finland typed the name on top of the video picture.

Only the first names were given, and typing them was a sign of a cross-cultural situation rather than doubts about the audio line: Finns and Swedes do not necessarily know how the names from each country are spelt.

7(17)

1 S-Phil: *silvia first*

In 7(17), the teacher not only types the name of the student, but also her status as the first person to be introduced. Out of the nine Swedish students, three would include the typed name in their introduction at some level. The rest would say *I'm Sibylla*⁵⁴ or *my name is Kristen*, i.e. normal introductions preceded by *ok*, or *hello*. The first one to implicitly involve the typed name in her introduction was Tamara who was the fourth one to introduce herself. She also happens to have Finnish parents so she started in Finnish:

7(18)

1 S-Phil: *tamara*
 2 S-Tamara: ((with heavy accent:)) **no terve(h) (.) .h er: minä olen siis tamara**
 3 **mutta tämä taitaa olla englannin kurssi että pittää sitä puhua**
 4 **vähän englannin kieltä nytten,/**
 5 hi there(h) (.) .h er: so I'm tamara but I guess this is an English
 6 course such that one has to speak a bit of English now then,

The Finnish *siis* ('so') in Tamara's turn-at-talk refers to the fact that her name was written on top of the video picture as well.

The next speaker to refer to the written name is Marion. He said:

7(19)

1 Marion: **hello all of you space age er students. my name ɹ (3) beard**
 2 **here. not Marion** ((pronunciation: [marion]))

The fact that the name is written on the video picture must have influenced Marion's explanation about how his name is not pronounced in the English way.

The last one of the Swedes to introduce herself, Margit, also refers to the text:

⁵⁴The names of the participants have been changed. The place names have been modified, too: the Finnish site is called 'Suomi' (the Finnish for Finland), and the Swedish one 'Sverige' (the Swedish for Sweden).

- 1 S-Phil: *margit*
- 2 S-Margit: **hello. as I guess you can see my name is Margit.**

All the Finns introduced themselves with the same format: GREETING/*ok.* + *my name is X.* None of them was referring to the typed name on the video picture (even though one of them had to correct the typist who enquired about the spelling of her name while she was introducing herself).

So, some of the students would react to the text on the screen which, though not produced by themselves, referred to their 'label', to their name. All of the students were near the camera and the screen so it was easy for them to see their name written on the video picture. In semiotic terms, the video picture of the 'other' and of oneself is a metonymical representation, an icon: not only a static picture or a television document or a film, but an iconic representation of the ongoing semiosis, the ongoing interaction. When a name is typed on top of this picture, identifying who is talking, who is the sender in the communicative situation, the visual signified is thus labelled with a signifier. Of course a lot of other things were going on: the students were giving some background knowledge to the fellow students across the sea, but the typed name stayed on top of the picture of the talking head, reminding the others who she was. It is understandable that most of the students did not refer to their typed name in any way; its primary function was to inform those who meet them 'face-to-face' for the first time how the name is written, not just the spoken form. So, it was a mere graphical version of what they say aloud, giving extra information in that sense. There was a one-to-one relation between the typed name and the face, even if there were two girls in Sweden with the same name!

Out of those who did include the typed name in their turn, Tamara with one tiny word (*siis*, 'so') and Margit specifically (*as I guess you can see*) referred to the visual information, and therefore to the redundancy of their said words about their name. Marion's reaction was strongest because of the possibility of mispronouncing the non-English name the way the similarly spelled English version of the name would be pronounced.

When considering the accessibility of texts, the typed name on the video picture of a person could quite safely be considered a very straightforward one. The text on the screen cannot be fully understood without the right person in the picture behind the name. However, even in this case the name was fully accessible only to those who have been in the unfolding situation. (If a Margit had walked into the seminar room in Finland, seeing her name on the screen and a woman speaking, she still might have considered that somebody, maybe the teacher, from Sweden wants her attention, or that she had been or was a topic of the interaction at some level.)

That these first introductions started with *my name is* or *I am* contrasts clearly with how the same students would take turns later in the video

conference sessions: this is (see Extract 7(6), for example). The use of this is highlights the indexicality of the situation, in which the student has to introduce herself or himself into, not in, the ongoing situation: the visual connection changes the use of indexicals and thus affects the participation framework (cf. Wortham 1996).

In some versions of virtual environments, the name of the representative of the person is a sign attached to the virtual self. For instance, in the world described in Bowers et al. (1996), the blockie had on top of it a name which can be seen from behind if the blockie was seen from behind: the written name was material in the same way as are signs on roofs. The data extracts in Bowers et al. (ibid.) show that in their introductions, no reference to the seeable first name could be detected, maybe because everybody expanded from that, mentioning their surname or place. In this case, the blockie and the name were both representations of somebody whose voice brought the real 'other' to the scene, connecting the name and the virtual representation to the real person: what could be seen on the screen was not what the name above referred to.

iv) Integral part of the ongoing interaction

In this mode, typing on the screen was basically replacing speech, not adding to or specifying spoken talk. For instance, when a person's name is typed on the screen by somebody at the other end of the link, this name on the screen can be interpreted as a summons.

During the introductions, the Finnish contributions could not be heard properly. When the second student was introducing herself and the people in Sweden could not hear her accurately, during her presenting of herself, the teacher typed *talk into the mic* on the Swedish video picture. The next student started her instruction with the microphone. So, typing could be used to deliver overlapping requests without the disturbing effect of spoken overlaps.

iv a) Typing as a first attempt

In the following two examples, a first attempt at summons was done via typed text, then sound which is more effective because it does not require the person requested to look at the computer monitor. At the beginning of 7(21), Kerttu was in the Finnish room on her own; Phil talked to the students in Sweden with his back to the camera:

7(21)

- 1 F-Kerttu: PHIL CAN HEAR <- <- <- <- YOU HERE <- <- AR ME??
2 (14)

- 3 F-Kerttu: ((moves to the computer to which the microphone attached))
 4 **Phil**,
 5 (3)
 6 S-Phil: ((moves to a computer)) **yeah?**

In this extract, it is visible for Kerttu that Phil might not see the text on the video picture as he is not near to any of the screens in the class and his back is towards the screen with the camera. In the following extract, however, the non-availability of the text on the screen is not self-evident from the video picture:

7(22)

- 1 F-Mirja: [---] **the women were the subjects of these texts.**
 2 F-Kerttu: *phil*
 3 F-Mirja: **a:nd (.)[I also (.) thought that (.) women would be]**
 4 |
 5 F-Kerttu: [!!!!]
 6 F-Mirja: **aware of any stereotypes that might exist in the text. uhm:**
 7 ((gaze up to the approaching Kerttu))
 8 F-Kerttu: ((takes up the microphone)) **phil phi- sorry I have**
 9 **something to say to phil now because vesa is telling that**
 10 **there is no sound from sverige now so is that a**
 11 **problem phil?** ((walks out of the picture))

The decision to talk instead of writing is much more consequential in this extract, as a student has to be interrupted for Kerttu to get Phil's attention.

iv b) Multiple dialogues

Sometimes a separate dialogue from the ongoing audio connection emerged on the video picture. Usually it concerned problems about the connection. In Finland, the teacher's helper would type on the screen something that was meant to be read by the teacher in Sweden or by a person in the other site in Finland. Meanwhile, the students could go on with their seminar, e.g. explaining the results of their project. In the following extract, the Finnish typed on the video picture is directed to the other site in Finland and translates 'vesa, it's ok – vesa doesn't matter - main thing that' (line 3) 'the picture shows' (line 6).

7(23)

- 1:25.02 ((*ok, it's on.*) deleted))
 1 S-Phil: ((explains the sound situation holding the mic in the pic,
 2 walks away))
 3 F-Kerttu. *vesa, it's ok – vesa ei häiritsee – pääasia [että]*
 4 |
 5 S-student: [() **mirja**]
 6 F-Kerttu: [*kuva näkyy* ↵]

7 |
 8 F-Mirja: [ok. (.) anyway,]
 9 er [what (1) I viewed (.) of the cosmol(i)- cosmopolitan]
 10 |
 11 F-Kerttu: [it's ok - don't worry ↵]

In the following, the Swedish end is informed that their picture, which had been gone for a while, was seeable again:

7(24)

1 F-Marjo: (--)
 2 [pleasure (.) and (.) gaze .h but (.) at least their- basic- (.)
 3 actually]
 4 F-Kerttu: [we can see you now ok]

So, the resource of typing on the picture was used in the video conference sessions not only to converse when the audio link was not working or when there was uncertainty about its workings: multiple dialogues emerged. They were not similar to McIlvenny's (1990) double dialogues in Talk tasks (the same participants taking care of several threads of conversation). Instead, the communication space that the visual depiction of the room, the video picture (the mediated space), offered could be used undisturbingly to give feedback about the general situation, the technical side of the link etc.

The use of the video picture for separate written communication from the spoken one seemed to be used much more than the dialogue box in Bowers et al. (1996): in their virtual environment, the problems about poor sound quality or other defects in communication were often dealt with through audio channel. Only when a total breakdown occurred, i.e. somebody could not hear anything, was the text window used. The separate dialogue box as a separate entity on the screen seemed to require a lot of co-operative work to draw the participants' attention to what somebody had written on it, even if the whole interaction was much more screen-oriented, i.e. the blockies could move in relation to each other in the shared space of the screen (or immersed world). In contrast, typing on one's video picture in TEACHING was seldom missed if the participants were looking at the screen.

Typing on the video picture was a feature of CU-SeeMe video conferencing that gave flexibility to the encounters; though face-to-face interaction is enormously adaptable, as well, different communicative structures became possible due to the mediated nature of the video conference

iv c-1) 'Normal' communication - typed text only

7(25)

1 F-picture: [((half of a man; no sound))]

2 |
3 S-Phil: [kerttu]
4 (7)
5 F-Kerttu: ((walks to the front of camera, ready to type))
6 S-Phil: *have you res*
7 F-Kerttu: *yes, who's th [ere with you?]*
8 |
9 S-Phil: [(tartered <- <- <- d)]
10 F-Kerttu: *no. i'll start [it now]* ((moves away from the video picture))
11 |
12 S-Phil: [maven?] ((moves to a computer further back))
13 F-Kerttu: *ok, it's on.*
14 S-Phil: ((working at a computer in Sweden, testing the sound with a
15 microphone in his hand))

In line 3, Phil is summoning Kerttu by writing her name on his video picture. Kerttu notices this and comes to the computer she can use to type on her picture. Phil starts typing his question immediately after Kerttu is in the video picture, and Kerttu begins her response to the summons and a question at the same time. Kerttu's question about the other person is dropped and she answers Phil's question. The typed interaction in the extract is very much face-to-face like.

iv c-2) 'Normal' speech and typed speech alternately

The following extracts are examples of smooth turn transition from Sweden to Finland, and at the same time from speaking to typing.

7(26)

1 S-Kristo: **hello, (.) Santtu, I'm very much looking forward to (.) er read**
2 **your work and: I'm thinking about working with those films**
3 **myself. I've seen them a few years ago and I remember I liked**
4 **them.**
5 (4)
6 F-Kerttu. *good*

7(27)

1 S-Silvia: (--) **see you next week.**
2 F-Kerttu: *see you*

In the following, the interaction slows down and hence typing is not just used to reply to spoken questions or other turns from Sweden, but also to make the turn boundary clear.

7(28)

1 S-Marion: **over.**

2 (21)
 3 F-Kerttu: *NO COMMENT*
 4 (90) ((during which teacher & Swedish students discussing))
 5 F-Kerttu: ↵ *YOURTURN*
 6 (29)
 7 F-Kerttu: ☹ ☹ *WE ARE WAITING...*
 8 (127)
 9 S-Phil: **hullo?** ((gives a task to both student groups))
 10 (2)
 11 F-Kerttu: ↵ *OK*

In Extract 7(28), the silence after Marion's turn allocation (*over*) is clarified by Kerttu's typing *NO COMMENT* on the picture (line 3). There is a long pause in the connection, though the Swedish end is clearly engaged in a discussion. After one and a half minutes, Kerttu reminds textually the Swedish end about them being the next speakers. When nothing happens, the expectation and frustration at the Finnish end is verbalised as *WE ARE WAITING....* Phil summons (cf. Extract 7(10) above) the Finns in line 9, and gives a task for both ends the line. This is confirmed by *OK* (line 11) in Finland. Confirmations of the type 'yes' and 'ok' could be efficiently and non-disturbingly given by typing when the teacher at the other end of the line was speaking.

The following transcript is a combination of a recording in the Finnish room and the third recording, i.e. from another site in Finland. Due to the slowness of the audio button in CU-SeeMe, the participants in Finland (and in Sweden) missed some of the talk that the other party produced because they had pressed the 'send' button in the audio and thus were not capable of hearing what the other end said. The extract comes from the first seminar in which the audio link was one-way, i.e. one party could speak at a time, typing was used at the other end to communicate short comments to the other end, not to interrupt the flow of audio data. However, the third site had a duplex audio and could thus hear everything that was said. The data thus gives information about what was happening at the Swedish end of the line (as only the Finnish site was videotaped). Arrow heads (<>) are used to indicate which part of what was said in Sweden or in Finland could not be heard at the other end.

7(29)a

1 F-Kerttu: **ok so that was (.) all from this end, and Santtu: will push the**
 2 **magic-**
 3 (11)
 4 S-Phil: **can you hear?**
 5 F-Pentti: <**yes** ((waves right arm))> ((waves right arm))
 6 F-Kerttu: *YES*
 7 (1)
 8 S-Phil: **can you hear us?**
 9 F-Kerttu: ↵ *YES* ↵
 10 S-Phil: **hullo?**
 11 F-Kerttu: <((to Santtu:)) **pu- painappa nyt** (.) / sa- please press now

12 sano[(kaa) että jee:]/say [that yee:]>
 13 S-Phil: [could Harri]
 14 F-Kerttu: aa:h::!
 15 S-Phil: <(.)and (.) molla do theirs again [(.) because they weren't]>
 16 |
 17 F-Santtu: [yes we can hear you,]
 18 S-Phil: ((speaking into the microphone)) (6) hallo?
 19 F-Kerttu: TAL[K NOW!!!!!!!!!!]
 20 S-Phil: [HELLO?] <(3) we are talking,>
 21 F-Kerttu: PHIL
 22 (10)

Kerttu's turn in lines 1 and 2 is an unfinished sentence: she never says the expected 'button' at the end of her turn. This might give the impression of the sound being cut off in Finland and therefore gives Phil a reason to check the line in line 4. This trajectory of discourse turns out to be a complicated one: the Finns have a hard time ensuring the Swedish end that the audio link is working. Thus, a missing word, a 'butterfly's wing movement', in this mediated communication context could give rise to interpretations that can lead to unnecessary complications in the communicative event.

The two positive answers (YES) are typed onto the screen in Finland after Phil's queries about the audio line (lines 4 and 8). However, they appear and disappear in a quick fashion. Phil is at the further end of the room in Sweden, and maybe misses the typed answers (the students in Sweden were sitting in a half circle around the Mac with the camera and microphone in Sweden, partly blocking Phil's view; it is interesting that they would not at this early stage of the first conference convey what they saw on the screen to Phil, if the typing appeared on their screen at all). So, the slow video line meant that emulating speech by typing an answer and making it disappear by pressing the 'enter' key (↵) right after would cause problems for the communication. Phil starts talking in line 13, but unfortunately the 'send' button for audio had been pressed before his turn (as the typing did not seem to work to ensure Phil that the audio connection was working), and thus the end of it is missed in Finland. Kerttu's cry in line 14 (aa:h::!) indicates her frustration of not hearing what Phil was going to say. Santtu's confirmation (line 17) overlaps at the Swedish end with Phil's contribution, making him perplexed about the status of the line again (hallo? in line 18).

At this point, Kerttu urges Phil to continue: she types TALK NOW!!!!!!!!!! PHIL on the video picture. Though not hearable in Finland, Phil answers the request (line 20: *we are talking*). Thus, Kerttu's written turn is interactionally integrated into the sequence of action and talk (both from the 'sender's' and from 'receiver's' point of view: TALK NOW!!!!!!!!!! PHIL is produced as a next turn and understood as a turn in the sequence in which talk and writing are used interchangeably). Phil's HELLO? in line 20 can be heard in Finland, and seems to be interpreted as an answer to the request on the screen (Phil said something). Typing on the screen is interactionally motivated and successful.

Also, Kerttu's contribution was clearly produced to a particular receiver, Phil (to emphasise the fact that his voice could be heard already earlier). It had to be added as in group-to-group meetings, eye gaze cannot work as a selector of the addressee: if the camera is looked at and hence the idea of direct eye gaze conveyed, everybody at the other end watching the video picture will receive it as looking at them! According to Blakar, having a specific addressee is "the crucial characteristics distinguishing communication from the general flow of information" (Blakar 1992, 236). This in contrast to when at the beginning of the extract, *YES* was used as a general response, giving information requested by the query about hearing us. More clearly, at the beginning of the session, the names typed on the screen were not interactional in the sense that they were not produced in a sequence of communication between the two sites, but were additional information to the face and the spoken introduction. Thus, the typing on the video picture can help think about the distinction between information and communication, or between knowledge *that* and knowledge *how* (see Chapter 6).

The attempt at securing the intersubjective understanding of a working audio connection was not very successful, however, because of the choice of wording. *TALK NOW* was interpreted as 'why don't you start talking' instead of 'go on talking' as it was intended in the situation. As often happens with typing on the screen, the text was left there until further text was typed in. This had interactional consequences, as the 'moment of the turn' with its meaning passed and the theoretical ambiguity (cf. Schegloff 1984) of a disembodied written sentence could play a part in the unfolding interaction subsequently (see below). Later in the interaction, the teacher requested for a typed confirmation about the status of reception at the other site. He formulated his spoken words as a request for a special wording: *Kerttu can you type that that's ok, you can hear us fine*. Kerttu's reply on the screen then shows her orientation to the request as a request of typing certain words: *WE CAN HEAR YOU FINE*. This way, intersubjectivity about the meaning of the words on the screen was achieved and there was no danger of a 'wrong' reading at the Swedish end. Thus, increasing degrees of specificity were needed to resolve repair sequences and requests for clarification.

v) Out of sequence

Video conferencing, even if high quality, tends to weaken the mediated gestures (e.g. Heath et al. 1997, 330). However, text on video picture has the reverse effect: because it is persistent, it can 'participate' for a longer period time in the ongoing interaction. Extract 7(29)b continues from where 7(29)a ended. In lines 45 and 47, at the same time as Kerttu types her answer to Phil's question in line 43 a student in Sweden reacts to Kerttu's earlier contribution (*TALK NOW!!!!!!!! PHIL*) 'out of' sequence in the unfolding turn-taking from

the 'sender's' point of view. At the same time, the visual message conveys to Finland the orienting to and interpretation of the encouragement to speak (as Phil's *we are talking*, line 20 in the previous example, could not be heard in Finland). The student interprets the turn on the screen as a next speaker selection⁵⁵.

7(29)b

- 23 S-Phil: **hello:?** <(7) **there's something wrong: at your en:d**>
 24 F-all: <((laughter))
 25 (4)
 26 F-Kerttu: ((patting at the armchair)) **istu Mervi tuohon, istu! ha-/**
 27 sit Mervi on that one, sit! do- >
 28 S-Phil: **hello:?**
 29 F-Kerttu: **<Harri varmaan joutuu seuraavaks tekemään () (.) okei**
 30 **sano että we can hear you please continue what you are**
 31 **going to say/Harri most probably has to do next () (.) okay**
 32 **say that we can hear you please continue what you are going**
 33 **to say>**
 34 F-Santtu: **we can hear you, you can continue (.) now,**
 35 (2)
 36 F-Kerttu: **<ooksää varma että se- ()/are you sure that it- ()**
 37 F-Santtu: **se on (.) pois päältä/it is turned off>**
 38 S-Phil: **[can you hear us?]**
 39 |
 40 S-student: **[Can y]ou hear us??**
 41 F-Kerttu: **<a:::r:::gh! ((to the front of camera; deep nod, deep nod))>**
 42 **((deep nod))**
 43 S-Phil: **can you hear us?**
 44 F-Molla: **<ye:s. yes, yes. yes, yes. ((nodding))>**
 45 F-Kerttu: **[PHIL YES YES YES]**
 46 |
 47 S-student: **[Phil is takhing] all the time)**
 48 F-Santtu: **we- we can hear you phil?**
 49 S-Phil: **<ours is set fine!> can you hear us now?**
 50 F-?: **'talking all the time'**
 51 (7)
 52 F-Kerttu: **IT'S CRACKING...**
 53 (5)
 54 S-Phil: **can you hear us now?**
 55 F-Kerttu: **YES, FINE**
 56 S-Phil: **good. (2) so what was the problem before?**

⁵⁵ At this point, a contrary problem to the turn transition in virtual 'blockie' environments (Bowers et al. 1996) occurs: instead of not being able to see who is self-selecting or has started their turn, in this instance, the Finns were interpreted as insisting the current speaker, though seeable in Finland, to start speaking!

57 (4)
 58 F-Kerttu ↵ *YOU WERE THE PROBLEM!*
 59 S-Phil: **thank you!**

In line 40, a student in Sweden resorts for the first time to the video screen as a communication site about the audio connection. The Finnish video screen is now oriented to as a possible site of an answer to the query. The Finnish video picture still has the text *TALK NOW!!!!!!!! PHIL* on it. It is impossible to know whether the Swedish student had seen the message before, but it is clear that she now is oriented to the textual content of the Finnish video picture and thus produced her answer in line 47. It's noteworthy that, unlike the delayed sound of Tarja in Example 7(7), there is nothing funny about seeing the 'voice of the past' as part of the ongoing interaction. The 'NOW' of the appeal seems to have its power as 'this instance', as the student replies to it *all the time*, contrasting the 'appeal for the moment' and the 'activity that has taken place for a while'. Strate's cybertime seems to be evidenced here: "we see the distinction between past, present, and future begin to breakdown as the three tenses become indistinguishable and essentially interchangeable" (Strate 1997, 368). The spelling mistake gives a rise to a speculation that the words on the Finnish screen were answerable as a 'voice' rather than as typed words as they were there all the time for the student as a model.

So, in line 47, the response and accompanying interpretation to Kerttu's earlier turn is done out of the context of its situated production. Indeed, the writing on the Finnish video picture at that point (*TALK NOW!!!!!!!! PHIL*) is not 'replacing Kerttu's voice', i.e. she did not utter it in that place in the interactional sequence. Therefore, even if the contribution from Sweden in no way disturbed what was going on, interactionally, the connection between the production of Kerttu's contribution and the response was abstracted away and artificial, though not for the Swedish party. In transient face-to-face turns-at-talk, the second pair part of an adjacency pair can come later, but the intervening talk must be a side sequence related to the first pair part. Any turn-at-talk can be addressed later in talk, but it cannot be understood as being about a previous turn-at-talk unless it is marked to be so (cf. Schegloff 1988, 131).

The two moments of interpreting the same piece of text on the video picture seem to fit with Anderson and Meyer's classification of singifying (cf. information) as a system and meaning (cf. communication) as local (1988, 17): in the first instance, the meaning was mutual and local, but in the second instance, the piece of text belonged to the system of language that always signifies. When the Swedish student embedded it into the interaction later, the force of signifying (as always indexical meaning making) could be detected. The process of meaning making was not mutual at this point, but the user-reader of the piece of text on the video picture interacted with the text rather than its author by choosing 'this next'. As interaction was that with

text, the progression can be compared with Aarseth's classification of discourse planes (1997, 127), also discussed in Chapter 6. Only now the mediated 'other' was not a representation of human intelligence, but a real site with human beings. The text on video picture was related to the event in the scene, and in the first instance, the *TALK NOW* sentence was negotiating the progression in the event. The second instance of reacting to the piece of text on the screen was more like a mixture of hypertextual progression and cybertextual negotiation: the student made the words on the screen part of the negotiation by metaphorically clicking on them, making them part of the progress of the event. This activity again resembled the situated reading of instructions from a manual to do things in the world: disembodied text was made to work in the unfolding sequence of actions (see Chapter 6).

7.6. Conclusion

CU-SeeMe provides interesting data for interaction research because two or more people have to succeed in their communication (secure their mutual understanding) when the visual, spatial and aural resources and the emergent participation frameworks are very different from their everyday face-to-face contacts. In CU-SeeMe video conferencing seminars, securing mutual understanding becomes a joint achievement. In the case study *TEACHING*, both groups were mediated via one camera picture from each location, and each group had to orient both to their own group and to the mediated 'others'. The incoming and sending video pictures affected the organisation of the seminars: the participants' talk and other actions were oriented to the artefacts and representations in the room. Though the equivalence of the sending picture with the 'others' incoming picture was not straightforward at the beginning, that and the effects of gaze to the camera and on the screen were soon learnt. The use of 'telephone conversation' lexicon (*this is, hullo?*) emphasised the nature of the communication as taking place in two sites. However, this behaviour seemed to emerge at junctures in talk or at speaker change without visual confirmation: more work had to be done in establishing a connection. Also, the questioning intonation of *hullo?* as a "try-marker" (Schegloff 1979, 50) made it work as if self-initiated other-repair: uncertainty about the connection was a delicate issue that was oriented to as a repairable, and the lexical choice constructed the potential repairers as those at the 'other end of the line'. That the summoner continued talking confirmed that the general hearing check (that was replied to) had been re-employed as an attention getter. Though *hello?/hullo?* never served to establish identity, *this is* was used for that purpose only, and the reason was the uncertain visual connection. Elsewhere than in the introduction sequence, it marked self-selection, or, self-selection had to be marked due to the non-shared status of the seminar. The name of the speaker in a seminar with a teacher is important

also as a log of an active participant in an educational forum: if the video picture (and sound) is not enough to carry the information of the identity, then it has to be announced. The format of the announcement reveals the orientation to the visual: the participants would not suffice with, for example, 'ok', or 'here', but used the formulation of the non-visible telephone conversations.

Unlike asynchronous computer conferencing systems which store the ongoing discussion in retrievable files, CU-SeeMe texts are used indexically and are shared in virtual copresence via video pictures, audio sound and computer graphics, adding to the complexity of the meeting. The video-mediated images of the other and of oneself objectify the interactional situation and may result in observation rather than participation, from direct to indirect social experience.

The participation framework in this 'virtual class' was not unified as there was a tension between the seminar as temporally shared and the technology-mediated audiovisual 'reality'. However, at the same time, knowledge of the other's view (literally) of oneself gives a stronger feeling of shared experience. Orienting to shared spaces and texts becomes a specific achievement when compared to routine everyday copresent tasks. By analysing desktop video conferencing the importance of the material setting of human interaction becomes clear: communication is not just about exchanging verbal messages, rather, it is very much undertaken by bodies in time and space.

In this case study, the participants were not only non-native speakers of English, but also attending a university seminar. The sometimes unclear sound connection became a site in which the expertise in the language and/or subject became threatened because not only was the availability of sound connection, but also competence, interactionally constituted via language use (cf. Goodwin 1986b, 292-293). Thus, the mediated language, both spoken and textual, together with the visual information, i.e. the technologised encounter, could result in potential tensions in and worries about the social relations: the distorted time, space and sound have to be made to work in doing appropriate time, space and sound.

The live video windows mediating the bodies on the screen could also be interpreted as metonyms, as parts standing for a whole, both in the sense that you can only see the other room or the other participants partially, and because the movements, colour, and voices are distorted because of the technology. Above, I have given an account of how the metonym of the mediated 'other' and of ourselves affects our doings and sayings in the interactional situation. The aim with video conferencing is to create a feeling of copresence but the mediated nature of the interaction could increase the chances of objectifying ourselves and the 'other': the participants in video conferencing can become event-viewers rather than event-participants (Zetzl 1990), also of their own actions.

In a video conference situation, the material circumstances make it possible that later a piece of typed communication transfers to information because of its non-transient character. In this instance, the occasioned use of *TALK NOW!!!!!!!!!! PHIL* changed into asynchronous communication that for the respondent, though, later became part of the ongoing interaction. Responding to the written order 'out of joint' from the 'sender's' point of view did not disturb the ongoing interaction, and thus the typed sentence in both cases was part of a successful interaction: in the first case synchronous communication, and in the second case semi-asynchronous (because for the Swedish student, the order was part of her ongoing synchronous interaction with Finland). The temporal 'now' (literally, *NOW*) became textual 'past' that was reincorporated into the interaction. The typed answer on the Swedish picture conveyed the interpretation of the Finnish text to Finland, though the answer did not influence the trajectory of the interaction, as, simultaneously, the Finnish picture was filled with words to reply to another query by the teacher, a reply which finally convinced Sweden that the audio connection worked.

Typing on a video picture is qualitatively different from asynchronous or synchronous text-only interaction, or video conferencing with a Chat/Talk box for written interaction. In the first two cases, apart from the appearing texts on the screen there is no idea of what is going on at the other end(s), and in video conference with a separate talk box, the visual image and the text-based interaction are separate from each other. And, though all of the produced text (with its sender) is kept in the Chat/Talk box, the disembodied nature of it from the video picture and the play script character of it (the name of the sender first, then colon and the message), means that the connection between the visual and the textual interaction is looser. Further research will have to investigate the interactional consequences of the separate Talk/Chat box for the ongoing interaction vs. text on the video picture.

Textual communication on a video picture does not necessarily degrade the sharedness of the situation. Rather, it can add to the mutual space which is created with the video image (icon) with the textual (symbolic) (cf. Bonk and King forthcoming). And though the text on the video picture can stay, or it can be deleted by backspace or by enter keys, unlike conferencing systems, the video conference as an event is not stored anywhere (unless separately taped). But the text on the picture can stay there for a while (or long): the past that stays in the cybertime longer, but eventually fades (cf. Strate 1995, 85: "the past does not fade in cybertime"). The text on the video picture is left there as an 'indexical monument' of a prior interaction, but, unlike Streeck's (1996) boxes, continue actively contributing to the interaction later. A participant's motivated 'transcription of speech' can therefore become part of abstract asynchronous communication in which the production and interpretation of text are separated: we get a glimpse of the synchronous/asynchronous division in the making. In this case study, a

transition from mediated human-human interaction towards mediated human-text interaction could be detected.

8. DISCUSSION

The basis of this dissertation has been how people interact with and through language technology. There was a special focus on how the language, especially instructions, of the 'other' was oriented to, in order to find out how different material manifestations and differing amounts of interactivity shaped the interactions. The methodology used to analyse these semiotically complex encounters was conversation analysis (CA), which was apposite, even though it was originally developed to study talk-in-interaction (between human participants). Though interacting with representations of interactive language use is different from that between real interactants, the activity the human participants are involved in takes place in a sequential manner similar to talk-in-interaction: the language is sequentially interpreted, even if the interpretative work falls only on one party. The case studies concerned different constellations of users and language technologies, providing an array of encounters in which varying aspects of communicative resources used became accentuated. Thus, the data analysis was also a contribution to expanding the scope of CA research to study encounters with language artefacts as events of active sense making, and not just receiving, of mediated messages. In video conferencing, a message becoming mediated could be witnessed.

In four case studies, the human participants were involved in a practical assignment, the achievement of which was partially dependent on their use of computer technologies. The case studies were as follows: 1) TELEPHONE: leaving and listening to messages; 2) TUTORIAL: learning how to use a word processor; 3) TASK: making mailing labels; and 4) TEACHING: participating in a virtual seminar. 'Others' were not understood as the alienated marginalised persons of cultural studies, but as a set of language-using entities (virtual 'others') with which the observed human participants engaged in interaction to accomplish their activities (and assigned goals).

In the case studies, interpretation was looked at from the point of view of (social) semiotics and ethnomethodology/conversation analysis. Situated meaning making was accepted as the basis on which conversational inferencing is taking place. The process is subtle, and the knowledge of the participants of each other and of the task at hand was also negotiated in the

interaction. The data analysis addressed various aspects of the differing communicative situations: 1) how successful can 'others' be in communication, especially instruction giving (social event; cf. Streeck (1980), and 2) how the 'other' (and, also, the user-reader) is constructed in the encounters (human relations; cf. Streeck (1980)). For both research questions, the answers were sought for and detected in actual encounters, and when possible, from the human-human interactions, the specific features of which also were attended.

8.1. Introduction

Computers are the most common form of language technology in present (Western) societies. With their increasing interactivity and language use, the question of their humanness has also been brought up in recent years (e.g. Nass & Steuer 1996). This dissertation contributed in part to that debate, as well, but not through studying the language artefacts, nor their users, separately from actual encounters. The pitfalls of questionnaire research and interviews which cannot grasp the "Intersubjective Time" (Bowers 1991, 554) could thus be avoided, as people's moment-by-moment language use and interpretation was researched, and not separate constructions of their language artefact use (cf. Burman & Parker, 1993). Though three (TELEPHONE, TUTORIAL, TASK) out of the four studies were experiments, set up only for data elicitation, the level of analysis of understanding language was so detailed that the status of the encounter as an experimentation did not really matter. In fact, instruction sheets, for example, were not 'hidden away' but shown to contribute to the ongoing semiosis as an actively used communicative resource.

To make it easier to compare the different cases, in the following, the participation frameworks that emerged are listed:

- TELEPHONE: human-computer
- TUTORIAL: human-computer
human-human
- TASK: human-computer
human-(manual) text
human-human
- TEACHING: human-human (across line and at both ends separately)
human-(video picture) text

All of the cases were examples of activities, the experiments as clearly defined tasks, but also the virtual classroom as a set of seminars between Sweden and Finland. In the pairwork cases (TUTORIAL and TASK), the activities that otherwise would have been solitary ones (a user and a

computer/manual) were now done cooperatively thus bringing in another other, the coworker. Except for TELEPHONE, the case studies were settings in which a minimum of two people instead of one participant were interacting with or through language technology. Thus, the interpretations of the mediated 'other' were not only sometimes coconstructed (and contested), but also individual actions on the basis of the actor's interpretation of the mediated language would be accounted for to the other participant(s). In a cooperative situation, the participants would naturally display to each other their understanding of what was going on, and thus bring the 'intra' action with the mediated 'other' fore in their interaction with the 'real' other. This was especially true in the two pairwork situations of TUTORIAL and TASK, in which the interpretations of the (computer or manual based) texts were enacted as actions at the computer. The pair had to manage the situation through talk, but also through gesture and body posture, gaze etc..

By examining situated interpretation, the present work contributes to the research on indexicality in language interpretation. Heath has noted:

Research concerned with the contextualisation of language will make little progress through the creation of premature theoretical distinctions or conceptual models whatever their intuitive appeal; rather it is necessary to continue to undertake painstaking empirical work, across a broad range of naturalistic (perhaps even experimental) materials, and further chart the interpretive procedures and interactional organisations which inform the contextual production and intelligibility of social actions and activities.

(Heath 1992, 125)

As increasingly, the contexts of encounters are those of language technology, the question of intelligibility of activities must also concern interactions between humans and language technologies. Conversation analysis as a method could help reveal what can be problematic, and why, in the interpretation of the mediated language.

The unavoidably indexical interpretation of instructions was shown to be an enormously complex phenomenon, especially when the direction giver was not present. The data analysis showed how the indexicality of sign (whether symbol, icon, or index) interpretation is dependent on the sequential and material features of the setting. Therefore, the threefold division into sign types becomes irrelevant: also symbols and icons, and not just the 'indexical' *I*, *you*, *this* etc. get their meaning in the unfolding situation. Texts have also been classified into symbolic, iconic and indexical on the basis of the 'sender's intention'. According to Nöth (1995, 46-47), the otherwise symbolic language becomes indexical in commands, instructions and other conative functions, and iconic in the order of narrative depicting events. However, the division does not seem to work at the receiver's end: either iconic narrative can be interpreted as indexical commands (when information is acted upon), or, the indexical requests are produced in the style of iconic narrative which does not cater for the indexicality of situated interpretation.

Though, for example, TUTORIAL and TASK were superficially similar, two users at a computer system, the directions that were given to the users were found on different surfaces. In TUTORIAL, the user was given guidance by the system to work on the same surface (i.e. the computer screen), whereas in TASK, the manual text gave directions about a different exterior (the screen). This had the consequence that in TUTORIAL, there were problems of interpretation due to the fact that the instruction resided on the same material surface: for instance, the two-phased command of *Move the mouse so that the mouse pointer is on any letter in the phrase Next Screen* and *Press Next Screen*, would result in the user clicking with the mouse the material expression *Next screen* of the first direction. This was not a problem in TASK; on the contrary, the (materially) divorced instructions and warnings about the computer (e.g. *Do not use the Toolbar...*) resulted in perplexity. However, the problem was not the divorced surfaces as such, but the fact that the sequence of interacting with the instructions was separated from the object of the activity. As discussed in Chapter 6, this was mainly due to the narrative nature of the manual text, even in places of instruction giving. When language technology guides actions (that often concern the use of the same artefact), the success of the activity depends on how the text produced in the past can break its inherent 'textual time' for the service of the human participant's situated, sequential, interaction. Thus, Streeck's claim that "speakers use sequential resources to convey intended illocutionary force; sentence semantics plays a less important role in this process" (Streeck 1980, 142) can be extended to when acting on the basis of text, even if the intricacies of copresent sense making, such as 'on-line' recipient design, cannot take place.

Coherence of indexical actions could not be easily achieved by the use of indexical language features. For instance, the textually coherent (orienting to the text-internal sequence) and indexical marker *the* in the *Do not use the Toolbar* sentence actually caused confusion and the sentence for a while had the effect of an other-correction. Nor does a visual depiction (an arrow) necessarily help. For instance, the *Next Screen* in the directions to press the same words was highlighted, together with a long arrow which pointed at the words to press (see Extract 5(4)). However, the highlighted words foregrounded the same element to become the object of the action. These were examples of 'real' products (and interpretations); how indexicals are used in predesigned directions (intended to appear interactively or not). Thus, though "many of the spatial, temporal, and objectural divisions signalled by indexicals are predefined by things like architecture, activity spaces, calendars, work rhythms, and the sociocultural values of objects" (Hanks 1996, 222), sometimes these 'external forces' can affect the interpretation of the indexicals in ways not intended.

Compared with the nameless producers of computer tutorial (in TUTORIAL) and manual texts (in TASK), the pieces of text appearing (and staying) and disappearing on the video picture in TEACHING were very

organically rooted in the contingencies of use. The interactive situation differed because the producer of the text would be present to monitor the text being interacted with; in the case of predesigned texts, the user has the sole responsibility to endure the consequences of 'choosing this now'. In TEACHING, the situation was predominantly one between two groups, and the responsibility of a text producer was not so much about his or her interaction with somebody else, but a contribution to the ongoing seminar and to managing its 'technology' (sound and picture). It was shown that a piece of text can start having a 'life of its own' (*a posteriori*) and shape the interaction trajectory after its occasioned production (and interpretation). Maybe due to the 'quest for the ubiquitous video', the latest versions of the CU-SeeMe video conferencing program do not allow participants to type on their video picture. Instead, all written communication takes place in a Chat box. Thus, the architecture of the space now hides the process of 'turn-taking', and only shows the ready-made product sent by the participants to the Chat window. Typing on the video picture meant that taking a turn, and editing it, would be seen from the start. It seems that while the text stayed on the video picture, it was prone to become interpreted as an active contribution by the sender, whereas the Chat box visually 'hides' the turns (which are still accessible) when new turns are sent in: it is not the 'speaker' of the turn, but the technology of the program that gets rid of turns. So, typing on the video picture, even after its production, was intimately connected with the 'now' of what was going on in the site.

The mediated 'others' of the technologised Information Society are often (sought to be) those of service encounters, one commodity of which nowadays is seen to be teaching. The anonymous 'other' of a computer tutorial or manual is as such a harmless creation that aims at conveying some facts about the computer program use to the user. However, when more complicated systems, such as airport information (e.g. Wooffitt et al. 1996) and social welfare information (e.g. Frohlich & Luff 1990) are computerised, there is a danger, especially if the role of technology is not contested at all in the society, that these systems become knowledge *that*, storage places of knowledge to be inquired about, but not to be discussed (and criticised). The one who seeks for knowledge is necessarily in an asymmetrical position in relation to the one that has the knowledge. However, at the same time as the conduit metaphor of interaction and teaching is under critique, the role of technology in the life long learning project in modern society should not be only to convey information but to offer a channel for discussing and contesting the knowledge and ideas of, say, a welfare system. The 'intelligent' systems that are built for database queries cannot be the sites of negotiation and challenge, it can only happen through discussion groups and bulletin boards etc., i.e. through mediating, not 'interacting', technology.

In the light of the empowerment of the technology user, the contesting of the system, usually by laughing and therefore interpreting the computer's

turn as a laughable (cf. Goodwin 1986b, 300), should be a welcoming feature rather than a design failure. When in the case study TUTORIAL, for instance, the machine-initiated other-repairs or praising comments were laughed at, the conversation analytic observation of a critical audience of a 'computerised speaker' was witnessed: "recipients through their interaction with each other can offer competing frameworks for both interpretation and alignment which undercut those of the speaker" (Goodwin 1986b, 283). Though the 'other' of the system would be laughed at, the real other would be defended against the *impolite thing* (TASK, Chapter 6). Laughing at the system (also in TELEPHONE) was the result of not being able to challenge its ways of doing otherwise. Not making fun of the other human participant was also a sign of the delicacy of the status of expertise in new technology use.

That the users laughed at the system's responses shows that the sequentially produced language can cause strongly copresent feedback (even if laughter is, at the same time, displaying one's reaction to the other participant(s)): "Some types of expression, such as laughter or displays of emotion, obviously lend themselves to copresence" (Boden & Molotch, 1994, 268). The reason for laughter sometimes lies in the difficulty of designing for indexicality of action: a praise changes meaning to an almost ironical comment when it is displayed after a series of repair initiators. Whether the laughter appears after misplaced or successful interactive 'turns' by the computer, there is an element of negotiating the 'other', not just of the sign meaning: at these points, the program orients to the user as a 'learner who just now succeeded'; by laughing, the users show that the 'intersubjectivity' or 'awareness' that the comments require in human-human interaction is not quite appropriate when expressed by the program. It is noteworthy that there was no laughing at the noninteractive manual.

8.2. The interactionally salient features of the four activity systems

This dissertation investigated how and to what extent the user's sense making of text, sound and (video) picture in a (spatially and/or temporally) distanced communication environment varied when linguistic and other signs were incorporated (by user and the technological artefact(s)) in an activity. When the artefact activated text, the user was forced or invited, depending on the implementation, to accept the same piece of text as a relevant part of the interaction. The work studied the interpretation of signs in language technology surroundings in all its contextual complexity. Of the four different case studies of interaction in technology-mediated communication environments, one had two groups of students who were linked together in an on-line video conference (TEACHING). The rest (TELEPHONE, TUTORIAL, TASK) were environments in which the 'other' was functioning on the basis of a predesigned computer program, or a printed manual.

A common word processor (Word 2.0) and its manual were used in TASK, representing an everyday use of an utility program. The printed manual provided a useful measure for comparing traditional media for helping the user-reader with new interactive media for doing similar activities. In TUTORIAL, the Learning Word program of MS-DOS Word 5.0 was used. The desktop video conferences in the TEACHING case study were arranged via a freeware program (CU-SeeMe). TELEPHONE was a completely experimental system which was designed for the case study only.

Severinson-Eklundh concludes her dissertation on computer conferencing systems: "the consequences of different system design's discourse needs extensive study" (Severinson-Eklundh 1986, 153). This dissertation concentrated on how various types of language technology engaged the user into activity on the basis of instructions: TELEPHONE (spoken 'electronic'), TUTORIAL (written electronic), TASK (written electronic and paper), and TEACHING (spoken and written). The systems as spoken or written interaction sites provided different potential for the users to do turn-taking: the transition relevance places were spacious in TUTORIAL and TASK, whereas in TELEPHONE the users tried not to overlap to hear the whole of the system's 'turn', and in TEACHING, the mediating sound connection was such that overlapping caused echoing which was avoidable in the situation. However, the video picture provided a communication space for nondisturbing overlaps and even completely separate dialogues.

To get a more comprehensive picture of all the four encounters, the following summary will deal with the aspects of what the users wanted to do with the systems (to look up information or to do something) and what part instructions played in communicating the options or possible choices to the user. Each activity system is also compared with hypertextual or cybertextual systems (cf. Aarseth 1997) as a way of describing their use potential. The 'no time out' nature of copresent interaction which involves continuous production (Boden & Molotch 1994, 264) was true for TELEPHONE and (however, more loosely) for TEACHING. In TASK and TUTORIAL though, there was some 'time out' (e.g. Boden & Molotch 1994, 264, referring to Garfinkel) in encountering the textually mediated 'other'.

Also, the question of control in the situation will be addressed.

8.2.1. TELEPHONE

The time of the human-computer interaction was such that the program forced to proceed in 'talk time', resulting in fast, spoken turn-taking. The interaction was direct between the system and the user(s). On the part of the user, language was both interpreted and produced as talk.

The electronic and other texts that were the basic communicative resource in the interaction:

i) The system. At the computer, a mediator read aloud the computer's turns and typed in the user's turns, resulting in an extreme case of computer 'driven' talk as the mediator repeated word-for-word what the system exhibited on the screen. As the role of the reader was that of an in-between, the system was the 'real' participant in the interaction, not a 'third party'. The computer's turns were informative, directive, requesting and repairing; they were predesigned (programmed) to be enunciated in any encounter with a user (three users tried the system out in the study). The predesigned 'aim' of the system was to ask information from the user about what they wanted to do with the system. So, the first request for all the users was *What do you want to do?* which did not tell the user what the options were at all, i.e. the system did not exhibit the possibilities or scope for action like, for instance, a screen in the TUTORIAL did. The telephone answering system also provided a so-called natural conversation end vs. e.g. 'spoken menus' that are fairly common nowadays in which case the system verbally tells the user which 'button to press' ('To leave a message press 1, to listen to a message press 2', etc.). The system had a very simple pattern matching mechanism for detecting 'a possible (right) answer', which it checked with a self-initiated other-repair (*So you want to – is this right?*) and confirmed: *Allright, the message 'Jackie is going to a meeting 7 pm' is to be given to Fred.* If neither of the two options was the result of the parse, the system initiated a repair (the format of which was chosen randomly). However, in the actual encounters, these repair initiators were also to be interpreted as further requests for information, and therefore as implicit confirmations of the user's previous speech acts or turns.

ii) The instruction sheet. The users had agreed to do a task which was given to them on a piece of paper (Appendix 4-1). Thus, the instruction sheet, typed well before the encounter, was a kind of third party, representing the initiative why the secretaries were asked to do the experiment in the first place. The task was about doing something with a system, but the sheet of paper only described the subtasks to do (what requests to pose to the system), and only gave some general directions about using the system (e.g. asking not to use personal pronouns; the user's ability to initiate repair was informed about but all instructions were given as knowledge *that* (*It is possible to indicate mishearing and to some extent misunderstanding*), but not as knowledge *how* (e.g. 'for example, by saying 'What?''). Thus, the instruction sheet presented 'loose' requests to act, to direct an 'other' to leave or play back messages.

iii) The human participants. The secretaries had time to read the sheet of paper before they phoned up the 'system'. Thus, the users could 'translate' the directions on the sheet of paper into action, and in this case, interaction (into speech acts) with a system. They had to make the general directions to work in a specific, turn-by-turn situation, realising thus the basic conversation

analytic idea of context sensitive and context-renewing turns: their requests to act (verbally and directly) to the system were at the same time their analyses of the system's previous turn. The overall goal, action-in-the-world, was to make the telephone system record and play back messages, and to manage any problems in this verbal activity. Thus, each user was 'pushing the buttons' of a hypothetical telephone answering system, but doing it verbally; they were trying to make the system do something and in that work was the use of the linguistic resources available. In a way, each user was 'clicking hypertextual links' to go on as they 'chose' the words used by the system and the instruction sheet. The user's requests were verbally confirmed specifically, e.g. *Allright, the message...* or implicitly, e.g. *Leave a message for who?* (as every turn is an analysis of the previous turn, taking the discussion further). The user's turn also could be disconfirmed by repair initiators (e.g. *Leave what?*) by the system, thus helping the user to search for a proper wording or 'hyperlink'.

The users were in charge of doing the task or action (trying to use the system), they had agreed to have an 'intention' described in the sheet of directions. The cases showed how the users would try to 'guess' the format of the relevant next, reformulate their request into a 'system friendly' configuration. Or, as shown by one case, the user could 'override' the system's 'intention' i.e. the user treated the system's repair initiators as another request to take the interaction forward, and not as an initiator of a side sequence. As the dialogue proceeded by the system's turns to which the user had to align to achieve her goals, the interactive constitution of the encounter was system-driven, though mutually accomplished.

8.2.2. TUTORIAL

The time of the encounter was in 'slow motion', as the program would only alert the user if he or she did nothing for a minute (which never occurred in the case study). The interaction was direct between the system and the user(s); there were no instruction sheets or manuals, for example, in the situation. The material setting made the use of the system different from that in TELEPHONE: the two users would 'mediate' the next action by negotiating the meanings. Their interpretation of what to do next was based on the (material manifestation of) language on the screen. A user would also produce language by typing when the program instructed him to do so. The interpretation of the screen language was otherwise realised as mouse clicks. There was no textual 'third party' in the form of a separate sheet of directions, for example; only the screen and the two users. However, a human tutor was present ready to give advice if needed.

The electronic and other texts that were the basic communicative resource in the interaction:

i) The Learning Word 5.0 program. The system exhibited predesigned requests and repairs, and thus was the initiator of the exchange. However, there was a mixture of directions and information on the screen. The system's requests for action were 'implicitly confirmed' by the user's action (and sometimes explicitly confirmed by human-human interaction in which interpretation tended to be offered as 'trials' to the other participant). The system gave evaluative feedback if the user carried out the requested action correctly: e.g. *EXCELLENT!* (a 'follow-up', e.g. Severinson Eklundh 1986). If the user's action was not correct, i.e. if the user chose a wrong one in the possible scope of activities offered by the system (mouse to click on screen, keyboard to type on the screen, and sometimes directions what to do next), then a repair sequence was initiated by the system. As was shown in Chapter 5, though the system could actively repair the user, the visual semiotics of the screen and the user's orientation to the system as consistent in its 'ways of doing' (e.g. repairing) resulted in 'delearning' the earlier explanation of the troublesource.

ii) Human participants. Out of the two participants, one was always in charge of using the computer gadgets (mouse and keyboard), though in repair sequences (between the two) the assistant would click a key on the keyboard, for instance. The actions-in-the-world for the user were to 'do what the teacher says', i.e., to type, move the mouse and click the mouse. The users' 'general goal' was to request information, to learn how to use the Microsoft Word 5.0 program. The user could choose from a menu what subsection to do (e.g. *Entering text*), and at the very beginning, the user had to choose whether he wanted to learn how to use the mouse. The 'hypertextual link' to do the next was sometimes open: *Next Screen* did not tell anything about the content of the screen to follow; it was just a link to go on. In fact, the scope for the user of Learning Word 5.0 was actually larger than in the 'conversational' TELEPHONE, because the material situation, the encounter with the system, gave the user more options to act directly than to try and tell the system to do something verbally. For instance, *Next Screen*, *End Practice*, *Course Options* were available most of the time as options to click next, as were clicking, typing, or attempting to do something else out of the scope with the system (which caused repair initiation).

The human participant was in charge of the task/action/'intention', which was detectable in the options available for the user to bypass the parts of the learning program that he did not need. Also, the human participant was in control of the system in selecting when to exhibit new information (knowledge *that*) by pressing *Next Screen* or *End Practice*. But the system was in control of repairs and praise. There was a sense of game playing in the way

the participants talked about 'trying' a chosen next action in case of uncertainty.

8.2.3. TASK

The human-computer interaction time was 'frozen' in TASK: there were no demands on the part of the computer on the pace of the user's interaction with it, though the menus, options etc. could be analysed as exhibiting passive requests for action (cf. Smith and Whalen 1995). Thus, it was possible for the human participants to take 'time out' and monitor the textual object rather than interact with it. The participants were engaged in direct interaction with the computer, the 'requests' of which got confirmed implicitly by the users' actions. Their interaction with the word processor was also always a request for the system to do something, and in this work they reacted to or interpreted the linguistic or other signs on the screen as mediating their request: the word or sign was clicked. Thus the signs on the screen worked in two ways: as invitations to click and as promises to carry out a command. The participants made use of the manual for technical advice on what to do next, to select which screen 'request' to reply to. The requests and directions of the manual were mediated into actions with the computer by the user.

The electronic and other texts that were the basic communicative resource in the interaction:

i) The human participants were the user (interacting with the computer) and assistant, though, like in TUTORIAL, the roles were not agreed beforehand but were also negotiated throughout the session, mostly through self-selection to be in charge of the mouse. The action-in-the-world was to use Word 2.0 (by typing or using the mouse) in order to finish an assignment. The users 'mediated' each other's expertise and that of the manual by negotiating next steps. So the manual was turned to for help or advice, and thus pieces of text in the manual became interpreted as advice (i.e. requests for action) which were then confirmed by carrying out the action. When the manual was used for general information, the request was 'confirmed' by reading and exhibiting to the other participant that what was read was understood (*mm-m*).

ii) The assignment text. The participants were to produce mailing labels, i.e. to do something with (the help of) the computer. The task was described on a separate piece of paper which the participants could read beforehand and which they had available throughout the session. The names and addresses for the mailing labels were given on a separate sheet, as was a model sheet for what the labels should look like.

iii) The word processor Word 2.0. The icons and words on the screen could be analysed as exhibiting passive 'requests' (cf. Smith and Whalen 1995 in which they treated electronic forms as slots of requests). These requests were implicitly confirmed by the user's action (using mouse or keyboard). The slots were 'hypertextual' options to choose something to do from (knowledge *how*). These options were closed (vs. open in TUTORIAL) because they already described what sort of activity they would do (e.g. *Format*).

iv) The manual presented technical information (in English) about how to produce mailing labels in general (general information). It also exhibited 'requests' for action (the steps or knowledge *how* of a general action), this time much more verbosely than the slots of the word processor. But, at the same time, the invitations for action were 'descriptions' (narrative), a sort of hypertextual knowledge *that* to expand one's information about how to use the word processor: these 'links' could be entered at any point, there were no constraints on the book format, though the step-by-step instructions were given often in a numerical order. The requests to do were 'tight' and indexical which tended to be followed by the users 'on-line' (and not studied beforehand). The requests were deciphered often through negotiation by the users, and also confirmed by user action.

The human participants were in control for moving on using the word processor on the basis of the task description sheets, the model sheet, and the manual. They were in charge of the task/action, i.e. the 'intention'. The separateness of the manual and the word processor (which, unlike Learning Word 5.0 in TUTORIAL, did not narrow down the user's options of actions) meant that as long as the user did not violate the word processor's 'system rules' (e.g. that some dialogue windows could not be open while another function was aimed for), the user could no way tell whether he or she was proceeding correctly in the task of producing mailing labels. Thus, uncertainty about a relevant next could only be resolved by proceeding one way and backtracking if the path taken proved to be the wrong one. Guessing took place, but the trajectory of the guess as a wrong one could be much longer than in TUTORIAL (which gave an interactive error message).

8.2.4. TEACHING

The time of the encounter was mediated copresent 'talk time' between human participants. However, the mediating language technology at times distorted the copresence, forcing into textual interaction (typing on the video picture) which then could influence the human-human interaction due to its permanent nature. So, the media might not be the whole message, but it does shape the message interpretation. All meaningful communication between the two sites was based on language (the attempts at for e.g. head nods were not successful interactionally).

The electronic and other texts that were the basic communicative resource in the interaction:

i) The two groups of participants were in separate places, one in Sweden, one in Finland. Their turns-at-action were not done to manipulate the screen, but the person(s) behind the screen. The two ends of the line often mediated or negotiated the next turn before delivering it, giving their interpretation of the computer mediated turn from the other end, whether spoken or typed. When they engaged in interaction with each other, the participants were doing the requests for action/information/help/advice etc. to each other through the system. The turns were aimed usually at the whole seminar group, though sometimes discussions between individuals took place (or a mixture: from the teacher to the group in Finland). There was always, however, an origo, a 'sender' of the message. Although when something was typed on the video picture, the 'sender' might not be detectable, and the visual message could become an 'authorless' piece of text. Requests were confirmed linguistically (e.g. *it's on* (typed on the screen) or *oh but it's on!* (said to the microphone)). *Ok* was used before action as a confirmation which made it a promise to do the action; *ok* was metalanguage or metaaction in the same way as the *Allright, the message ' ' is to be given to Fred* was in TELEPHONE. This type of linguistic activity did not take place between the mediated 'other' and the participants in TUTORIAL and TASK. The duty of the participants in TEACHING was to have or manage a seminar between two places. The constraints to the content of the assignment were those negotiated between the teacher and the students.

ii) The language technology was just a mediator of the interaction between Finland and Sweden. There were technical problems in this mediation, which were repaired or drawn attention to by either saying or typing on the video picture, e.g. *can you hear us?* When the request or attention seeking was in written form, it could be interpreted later as a 'now' contribution, giving a practical example of the enduring force of text vs. speech. Both spoken and textual interaction took place from both parties of the encounter on a turn-by-turn basis (cf. TELEPHONE).

The human participants were in charge of the task/action/'intention'. The interaction between the mediated 'others' was different from the other three cases in that guessing was not a major way of resolving ambiguities. However, ambiguities arose with turns typed on the video picture.

8.2.5. *Situated interpretation*

Conversation analytic research stresses the importance of the position of an utterance for its interpretation. Coulter argues:

Semantic values cannot be detached from 'pragmatic' functions in speaking, especially the *positioning* of what was said within a *sequence*, and syntactical categorization, if needed (which is not often), is contingent upon whatever pragmatic function a locution may be heard to be performing in its context of utterance.

(Coulter 1991, 333)

And Heritage says: "speakers can rely on the *positioning* of what they say to contribute to the *sense* of what they say as an action" (Heritage 1984b, 261).

The spoken interaction in case studies TELEPHONE and TEACHING showed the importance of the position of a sign in a sequence for its interpretation. In speech, the sequence is transient, but in writing, the position holds longer, allowing its integration (repositioning) into the flow of interaction again. In the case study TEACHING, typing on the video picture showed the use of written language 'as if' speech, which, if left on the picture, crossed the boundary to 'as if' written interaction. This is where the materiality of language as an important feature for meaning making became concrete: the spatial metaphor of a position in a sequence of speech turned into spatiotemporal, into how long a sign can hold a place (or return to one) in an unfolding sequence. With written language, the position can be taken up by a materially stable, fixed sign which can fill a position in the sequence of action for longer (or more often) than a transient spoken turn. When technology mediated spoken or written language conveys instructions, as in the case studies in this dissertation, they are 'voices' that are turned to for guidance in actions. The case study TEACHING showed how not only writing, but also voice could become (unintentionally) a trace of the past that occupied its position in the interpretative sequence. The consequences were interactionally different from when 'intentionally' past traces, i.e. written interaction (in TUTORIAL and TASK) were incorporated in the ongoing task. In TELEPHONE, the past traces were concretely voiced, and the pace of interaction was fast in the dyadic, voice-only interaction.

In all the cases, the language technology and its requests/directions were designed to be encountered by countless others, and the ones with predesigned language (TELEPHONE, TUTORIAL, TASK), the language used was meant to make sense in various settings by different participants. Thus, the interactive role they filled was an active one, giving instructions, even if the realisations of that role were interactionally different in each case study. The user-reader's interpretation of the mediated 'other's' turns were limited (by the predesign): for instance, in TELEPHONE, the user's answer (*yes* or *no*) to the system's prechecks (or self-initiated other-repairs) of the format *so you want to* tended to be expanded on from the simple confirmation (or correction). It seems that in the spoken medium, one syllable answers are difficult to give (unless the systems specifically requires one). Sometimes (in TUTORIAL) the system could give evaluating feedback to the user (*Great!*), a feature which was always received as amusing, but laughed at after an

'other'-initiated repair sequence. In TASK, the user-reader's actions were implicit confirmations of the computer's or manual's 'demands'. Only in TEACHING was a mutuality of understandings an issue. However, compared with face-to-face discussions, the mediated nature of the interaction demanded much more specificity in the negotiation of repair sequences and requests for clarification. As regards the shape of request-reply sequences due to the interactive slowness/fastness of the system (cf. Severinson Eklundh 1986, 73), only TELEPHONE demanded the user's rapid reply to the requests put forward. Even though in TEACHING, human-human turn-by-turn basis (voice) interaction was going on, the questions asked and requests put forward often concerned the whole group, i.e. there was no specific addressee. Also when the recipient was one person, there still was a possibility for him or her to negotiate the reply with the other people in the room. Typing on the video picture gave the interaction an openness towards 'no closure' as the written request was available for another reply. In TUTORIAL, the design of the hypertextual, partly interactive message sending program was such that though the user could go through the same requests (parts of the tutorial) time again, there was no real possibility (for the user) to open up again a request-reply sequence during one encounter, either by leaving the answering (because the system was programmed to alert the user if nothing happened for one minute), or doing the same thing again (because the system would give the turn-taking repair message *Just read now* (etc.)). But though the use of the word processor in TASK was channelling down the user once they had chosen a 'request' from the menu bar, the manual could be returned to, its requests replied to at any time. In practice, the word processor screen might limit the replies as action, but the manual was a piece of language technology least closed for reopenings.

In the case study TELEPHONE, both the system and the users posed requests: the system by laying out the possibilities of action and by repairing the user; the users by requesting the system to do something (e.g. to leave a message for somebody). In fact, the system's turns always included first parts of an adjacency pair: the user's task was to provide a reply such that the next turn by the system would show an 'analysis' that confirmed for the user that her answer was correctly interpreted (i.e. the turn was context sensitive), and, at the same time, would pose another request (i.e. the turn was context-renewing). When the system initiated repair, the focus was specifically on the content of what the user had just said. As the system could only initiate a next turn repair, the user's possible misunderstanding of the repair initiator as another request could not be handled specifically, i.e. there was no repair initiation at the third position. Though another repair often followed the user's turn, it was not designed to correct her understanding of the first repair (which would have shown in the formulation 'No, I mean'), but the repair

initiators were always reactions to the user's immediate previous turn⁵⁶. In both initiating and rephrasing (after a repair) their turns in 'talk-time', the users utilised the instruction sheet and the program's turns as resources for shaping their utterance. The interaction took place in a sequential manner, as if that of talk-in-interaction: the technology of turn-taking during the encounter was, therefore, similar to that between humans.

In TUTORIAL, the program's requests to repair resulted in a 'rephrased' action, and in TASK, the user had an opportunity to reexamine a previous request (or previous set of requests) in the manual if they came to the conclusion that their previous interpretation (as action) was wrong. In both cases, the text that was directing their actions was interpreted in a sequential fashion: in TUTORIAL, the program was designed to support sequential interpretation and to give an idea of 'turn-taking'; in TASK, the user's sequential interpretation of directives activated certain words as 'this now', though the responsibility for the first choice in a set of directions was the user's. In both cases, the user was able to restart the interaction from the beginning, in TUTORIAL by backtracking in the program, and in TASK by going back to the first directive again. Thus, the permanence of the text allowed for reiterating the interaction. In TEACHING, however, a typed request (or an utterance so interpreted) on the video picture, was treated as an active repetition of the request, rather than a passive possibility of repeating the event of 'request taking'. Thus, the availability of the authors of the text meant that the visible request, once extended, was active as long as it was present: the text did not pass completely to the realm of written interaction.

8.3. Knowledge *that* vs. knowledge *how*

As discussed in Chapter 6, Harris typifies action sequences that can result from communication into enactive (action meant to ensue) and assimilative (for information only). This echoed the distinction between knowledge *how* and knowledge *that*, the former concerning the practicalities of doing things, the latter implicating an awareness of state of affairs. In all the case studies, both types of language use could be detected. In TELEPHONE (Chapter 4), the enacting and assimilative language had to be listened to and the possible enactments done verbally. In fact, the user was forced to perform verbally all the time: even when the telephone system's turn included knowledge *that* information (e.g. *You can leave a message or listen to messages*), each turn would ultimately push the user to choose a relevant next action (*select either*). In the case study TUTORIAL (Chapter 5), the actions enacted included not only language use (this time typing on the screen instead of talking), but also physical actions such as moving the mouse and clicking it, or clicking the

⁵⁶ In TUTORIAL, the lack of the specific formulation ('No, I mean') of a third position repair caused problems in detecting the repair initiation.

return button on the keyboard, etc.. When in TUTORIAL the instructions were read on the computer screen, and were made seeable by using a simple hypertextual system, in the case study TASK (Chapter 6), the instructions of use were also written, but this time resided in a book format, i.e. a manual was helping in the use of word processing program. In TASK, the manual included a fair amount of knowledge *that* text, meant to be assimilated, but of course also instructions intended to be transformed into action. The problem for the user was that the two text types were, in a way, in the same format, i.e. they were both ultimately written in a style of narrative, which is not geared towards concrete actions in the world. The case TEACHING went back to the spoken mode again, and this time with real people, who however became representations ('others') through the mediated video picture and sound. The requests and instructions that could be given by speaking or by typing on the video picture resulted in enactive sequences of providing the second pair part, be that saying something or doing something. The saying/doing could also be done by typing on the video picture. Typing was a hybrid between spoken and written language, the endurance of which could result in an informative (knowledge *that*) language meant for assimilative action being interpreted as a request (knowledge *how*) and therefore enacted as one.

Therefore, my claim is that one basic difficulty in interpreting mediated instructions is whether a piece of direction is meant for information only or to be acted upon immediately. This distinction is visible also in telephone-mediated human-human interaction, of which an example can be found in Whalen and Vinkhuyzen (forthcoming, 37):

- | | | |
|-------|-------|---|
| 49 | CSSR: | it's so simple here's what you need to do |
| 50 | Cust: | okay |
| 51 | CSSR: | the access panel, |
| 52 | | (1.0) |
| 53 | CSSR: | where the dry ink cartridge is? |
| 54 | Cust: | huhuh |
| 55 -> | CSSR: | open it |
| 56 | | (2.0) |
| 57 -> | Cust: | you want me to through it right now? |

In line 55, the customer service and support representative (CSSR) instructs the customer to open the dry ink cartridge. The customer's question in line 57, *you want me to through it right now?*, orients to the inherent ambiguity of the CSSR's instruction as an enactive or an assimilative one. Whalen and Vinkhuyzen make the observation of the CSSR's talk that: "the CSSR is physically manipulating an imaginary machine in front of him, "opening" the "door," "turning" the "auger," and such like" (Whalen & Vinkhuyzen (forthcoming), 39). If in synchronous advice giving the tendency is to create an illusion of handling an imaginary machine, in manual writing the actual instruction giving sequence is the imagined event.

8.4. Results concerning uncertainty and repair work

As all the case studies were interactions in which the task done and/or technology used were new for the participants, troubles were to be expected. The actions, whether spoken or not, took place in a sequential manner. The less material and semiotic experience the participants had of the interactional situation, the higher the uncertainty. Designers had prepared for problems with actions in differing ways: in TELEPHONE, the system would actively repair the user (by voice), and the same happened in TUTORIAL (by written pop-up messages), and in TEACHING, the 'others' were present to repair and be repaired. However, unlike the active repairing in the other case studies, in TASK, the repair work by the 'other' of the manual was passive: the user-readers had to find a remedy for a problematic situation by (re)reading the instructions. More importantly, the next piece of information could result in self-repairing the previous action. It is understandable that the manual could not actively repair users as it could not possibly keep track of what they were doing. Nevertheless, like in other predesigned instructional systems, problem sources were oriented to: unlike in systems where the users actions revealed the problem and a repair message ensued, in noninteractive written instructions the troublesource could be marked before the erroneous action, as a 'warning' in the text. As was shown in the data analysis, a warning, due to visual placement on the page and ensuing positioning in the sequence, could become 'elevated' to a correction.

All case studies were examples of how important indexicality (and, thus, 'context') of language use is for meaning interpretation. When technology was involved, the question was always not only 'what next' (in *kairos* time) but also 'when next' (in *chronos* time), i.e. problems might result in 'turn-analysis' and 'turn-taking' repairs (cf. Raudaskoski 1992). The material circumstances in the case studies would result in different trajectories of action (and repair), according to the enables and constraints posited by the mediated 'other'.

In instructional texts, there is not only a tension in whether the instructive 'turn' is analysed correctly (i.e. understood), but also whether the enactive sequence should take place next, i.e. whether the user-reader's next turn should be to act at all. In TASK (Chapter 6), where the instructions and the object of action were separated, turn-taking could not be controlled by the text. And in fact, the inherently narrative (vs. interactive) nature of not only the general information, but also the instructions, meant that the participants were able to act on the basis of their 'turn-analysis' at any time (and decide how big the 'turn of the manual' was, i.e. it could cover one action or a set of actions). Thus, the only 'active' problem oriented sentence *Do not use the Toolbar to open the new file, because you need to use the Templates option* caused distrust on the turn-analysis of the previous sentence *From the File menu, choose New (ALT, F, N)*. There was no doubt that a file should be opened, only the manner was contested for a while. Thus, the knowledge *that* resided in the

manual (both general information *that* and instructions *how*), and the knowledge *how*, the doing on the basis of the manual, took place at the computer interface. This 'division of labour' made it possible for the user-reader to check her understanding (as action) of a previous direction on the basis of whether the interface (literally) displayed a state that was needed for the next instruction to be acted out.

In the other cases, both types of knowledge (information and directives), together with the ensuing actions and their repairs would all take place within the same system. For instance, in TUTORIAL, the system would repair the user if they tried to act on the basis of an assimilative (informative) piece of text when they were not in a 'rehearse' window: a turn-taking repair initiator would ensue. Also, within the practice mode, a wrong turn-analysis (and therefore a wrong action by the user) would elicit a repair message from the system. The two types were separated in the design in that the first one (turn-taking repair) would disappear without any action by the user, whereas the latter one (turn-analysis repair) would stay until the user did the right action. However, in the interactional sequence analysed in Chapter 5, it was shown that the turn-taking type of repair initiator and the other instructions had similarities that had an effect on the interpretation of the turn-analysis repair message *Move the mouse pointer to the "d" in "due". Click-L* as another instruction rather than an initiator of a side sequence.

In TELEPHONE, the user did not have to worry about when to do and when not to do: the turn-taking went smoothly as if in any telephone conversation. The program was not designed to elicit turn-taking repairs as the user was meant to do by talking after each system contribution. Sometimes the users overlapped with the system turn, a feature of conversations that can potentially lead to turn-taking repairs (e.g. 'let me talk now'). However, in the case study, overlaps were just a source of evidence that tag questions are 'overlappable' and that first pair parts of adjacency pair tend to elicit a reply from the user. Thus, the whole encounter was about the user finishing her task by formulating her requests in words acceptable (i.e. parsable) by the system. The repair initiations by the system either led to reformulations of requests or were taken as a next question by the system (i.e. fitted into finishing the task of the caller). The latter shows again the risk of machine-initiated repairs being passed as taking the interaction a step further rather than recycling the previous turn.

In the case study TEACHING, the turn-taking was repaired a lot, but this time it was the mechanics of turn-taking that were at stake ('can you see/hear me'). The analysis showed how the problems in mediating the sound and picture not only resulted in turn-taking repairs by the addressee ('I can't hear you') but also the speakers would orient to the addressees' possible difficulties in seeing and hearing them (as shown by the use of *hello* and *this is*). However, dealing with the hearing problems turned out to be a delicate issue which, on the part of the addressee, was potentially concerning

difficulties in turn-analysis, i.e. difficulties in understanding (the English/academic) language rather than not hearing it.

The *chronos* and *kairos* times constitute the sense of interactivity: the faster the pace, the more interactive the encounter. The case TELEPHONE required fast interaction, TUTORIAL was fairly fast, and TASK had no limitations at all on the users. In the case TEACHING, then, a face-to-face *chronos/kairos* time was aimed at through mediation and representation. Thus, the cases were different from, for example, SAMi in Frohlich et al. (1994, 398-99) which had a very slow response time, resulting in totally different repair trajectories: "the user takes the unresponsiveness as evidence for trouble in her performance of the action" (Suchman 1987, 147). This was not a problem in my data because the *chronos* time did not really have pauses from the technology's side and no unnecessary repair initiators resulted from the users due to apparent nonresponse from the system.

In repair work, the initiator is responsible, he or she is orienting to a troublesome understanding. In Frohlich et al. (1994), the user was the initiator of all requests, which the system did not 'understand'. In my data, the problems were of reverse type and consequently the computer was not treated as 'responsible': its repair initiators were easily considered as a 'next' request (e.g. in TELEPHONE and TASK). Thus, to rephrase Jensen (1995, 48; see 8.5.2): others' implications what I can do describes who or what I am. The users did not treat the system as being able to take the responsibility of initiating repair.

Self-initiated other repair is different from the other repair types if takes the form of a yes/no question which asks for confirmation/rejection: it concerns what is to come rather than what (just) happened, thus providing a possibility for the users for an *a priori* understanding check, preempting an other-initiated repair (cf. Frohlich et al. 1994, 418 and Frohlich & Luff 1990, 205-6). The yes/no format of uncertainty was abundantly used between pairwork participants: expertise was interactionally constituted as a delicate matter. The gist of self-initiated other-repair was a bit different in the implemented 'uncertainty checks', in TELEPHONE (*So you want to...* and *Do you mean...*) as the system was programmed to check whether its 'interpretation' was right or wrong, 'face' issues were not a (design) concern there. In semiotic terms, self-initiated other-repairs could be analysed as a type of request, a request at a metalevel to accept or agree with the participant's interpretation of the sign in front of them (cf. Interpretative as Ponzio's sign of answering comprehension discussed in Section 2.3).

So, all in all, the mediated nature of the 'other' resulted in various possibilities and connotations for repair work. As discussed earlier, different participation frameworks emerged on the basis of the number of 'participants' in the case studies. Thus, the TELEPHONE consisted of that between the caller and the system, the TUTORIAL had the human-computer and human-human participation frameworks, and the TASK human-computer, human-

text, and human-human interactions. In TEACHING, the human-human participation frameworks were in flux, sometimes concerning individuals, sometimes groups within each site and between the two university surroundings. Also, the human-text interaction had various degrees: sometimes the text would replace the other party's speech, sometimes it would contribute to the meeting without the author. Making a summary of the uncertainty and repair work in these environments is easier if a rough division is made between human-human and human-'other' interaction, though the latter often concerned, and was intermittent in, the first participation framework.

8.4.1. Human-'other' interaction

Mulkay discusses how in written interaction (more specifically, agreements) the time of the previous 'encounter' is stated; he goes on to say that in spoken interaction this would be superfluous: "your proposal a moment ago that.. I concur with" (Mulkay 1985, 211). Mulkay's data consisted of letters exchanged, and therefore was 'text-text' interaction: the action of reading resulted in another letter being produced. However, if we think about human-text interaction and about actions taken on the basis of written text (transforming text into action), in the case of 'interactive' written text (cf. TUTORIAL) the problem was stated in other-initiated repairs. For instance, *You have pressed the Num-Lock key* was used. However, the formulation was unfortunate because the users had actually not pressed the Num-Lock key (for some reason, the starting of the Learning Word 5.0 program caused the program to infer this was the case). Instead of formulating what the user had done wrong, the program could be designed to state that the previous action was not the desired one, for example 'You did not move the mouse to "d" in "due". Please try again: Move the mouse pointer to the "d" in "due". Click-L.' (Instead of *Move the mouse pointer to "d" in "due". Click-L.*) As the repair initiators were not distinctive enough, they were often treated as a next step in the task instead of as starters of a side sequence (cf. Suchman's 'garden path' (1987, 165): the users did not realise the system was backtracking). It looks like interactive texts (whether spoken or written; cf. TELEPHONE and TUTORIAL), compared with turns-at-talk in face to face interaction, require that the turn's temporal relationship to the user's previous action has to be clarified much more: is this command/instruction referring to the user's previous action as problematic and therefore a request to repair previous action, or as accepted and therefore a request to do a next action. Thus, for predesigned interactive texts, sequential appearing was not enough for the correct interpretation of the turn as context sensitive and context-renewing.

In encounters with visual, traditional written instructions (manual in TASK), the user-reader has to decide the size of the turn the manual is

'saying'. This led to the reading of 'one step' as two due to the visual layout of the text on the page. Also, the affordances of the word processing program and the 'world of the manual' did not coincide: when the command to open a new file and the ensuing clarification/warning concerned keyboard and mouse clicks respectively, the user had actually a potential of translating the *open a new file* into action with mouse clicks but on a different place on the screen (selecting *File, New* by mouse). The directions were to do the opening of the new file from the keyboard (*ALT, F, N*), and thus the next item (about not using the Toolbar) was more readily understood as a potential correction. We know that the user did read the warning next, i.e. she oriented to the following text as a next step. Thus, the user's action potential was bigger in using the word processor than what the manual implied. A contrary situation took place in TUTORIAL: the use of backspace key, which had just been taught, was inhibited by the program. These examples show the difficulty of program designers to pay close attention to the sequential nature of 'activating' the inscribed instructions. If pieces of text are meant to be read together, as one 'turn', and not as separately (increasing the danger of them being interpreted as separate commands), either the visual layout (including the direction and the warning in the same paragraph) or framing (by adding e.g. 'NB' or 'Make sure..' to emphasise that the intended perlocutionary effect is that of a warning) have to convey the aim unequivocally.

The TUTORIAL showed the effect of visual semiotics on problematic interpretations even at the level of a letter: the several manifestations of 'l' on the screen contributed to misinterpreting the instruction *Click-L* as concerning pushing the letter *L* on the keyboard rather than the left mouse button. Actually, the misinterpreted instruction was a repair initiator by the program (about moving the cursor to the beginning of a word on the screen by moving the arrow with a mouse and clicking its left button). Thus, the visual semiotics of the screen could be shown to contribute to the repair initiator as not having been understood as one, but interpreted as another instruction from the program.

The same problem about missing a turn by the system as a repair initiation happened in TELEPHONE, the first case study. In the last case, TEACHING, the reverse took place: a direction on the screen to continue talking was interpreted as a repair initiation about the status of the sound connection. Thus, all in all, asynchronous interactive language requires that the 'speech acts' be marked much more clearly than in face-to-face discussions. From this data, Winograd's contested message exchange system for more efficient work interaction seems an understandable effort (see CSCW 3, 1995 for discussion). Nevertheless, when the responsible originators of the messages are available (like in TEACHING), unlike with manuals and tutoring programs, the political risks of 'clear talk' counter the benefits of it. However, with stand-alone systems, whether interactive (TELEPHONE, TUTORIAL) or not (TASK), it could be beneficial for the user to have the

system announce the 'intent of the speaker' at the level of speech acts, e.g. to frame the message with the label 'correction' or 'warning': the accessibility of the messages would be smoother, as sequential placing does not seem to be enough for the messages to be interpreted as corrections. Another way of rechannelling the course of action would be to use a clear format of a third position repair, e.g. 'No, I mean', when the system can detect users' wrong interpretations of a suggested next action.

As the use of voice-based systems is dependent on transient sound, there are less resources for the user to formulate their interpretation of the system's each turn: the pace of interaction is fast, there is no visual data to draw inferences from. Yet in the more complex visual data and slower pace of interaction with an interface, more complications are potential for the interpretation process, because not only sequential meaning making, but also the visual semiotics play a part. However, in both types of interaction, the users try to adopt to the interaction style of the program, and sometimes this results in 'as if' talking to a person, sometimes 'as if' using a tool, sometimes 'as if' interacting with a text. The mediated 'other' is thus not oriented to as a person, as a machine, or as a text, but there is fluctuation that shows the ability of people to adopt to communicative environments and make use of them on the basis of their experience of various types of encounters.

8.4.2. Human-human interaction

In the pairwork situations, repair work was especially abundant, as only one of the persons was interacting with the system in a basically cooperative task. There was a clear tendency to 'doing uncertainty', i.e. resorting to self-initiated other-repair (Raudaskoski 1992) in these interactions, even when the initiator was the 'onlooker' as in *do we do that now?* in Extract 5(2), line 51. As indicated by *we*, uncertainty about a referent or a next turn-at-action in a pairwork situation not only was a sign of not knowing (really) but also an orientation to the task as a combined effort, and sharing the responsibility of the interpretations with the other participant. So, the human way of dealing with problems of reference or of 'what next' was done very vaguely. Though uncertainty and self-initiated other-repair are not linked in Schegloff et al.'s work (1977), an orientation to repair work can be seen in uncertainty marking as it prevents the less preferred other-initiation (Schegloff et al. 1977): the 'challengable' (Jefferson 1972) can be avoided. Uncertainty from this point of view is then an invitation to other-repair or confirmation by the other. Being a self-initiated invitation, it could be assumed that, in the same vein as ambiguity of meaning could be to do with politeness, in that the other is given the freedom to choose the interpretation (Raudaskoski 1995, 117), self-initiated other-repair avoids face loss by letting the other avoid repairing, and by giving the other the right to decide whether something is acceptable or not:

the other participant can decide the interpretant in the negotiation of the sign meaning. If a participant is doing self-initiated other-repair, then he or she is (strategically or not) assuming that the event is a so called AB event (Labov & Fanshel 1977), that the other might share the knowledge; or even that the other might be the expert (i.e. it being a B-event). Uncertainty is an other-oriented, possibly shared token (cf. Cheepen 1988, 62); self-initiated self-repair in contrast is underlining the A or self-oriented nature of the event, and other-initiated repairs the B or other-oriented nature of the action (the initiator 'knowing better' and imposing this knowledge on the other). However, though self-initiated hesitation (i.e. treating one's own knowledge as disputable) was abundant in the data, examples can be found (e.g. 6(3), lines 2 to 17) in which a sequence ensues in which the original questioner treats the following explanation as an underestimation of her understanding (cf. Auer 1984, 7). This demonstrates how delicate an issue the state of knowledge can be between coparticipants.

In his paper on referential problems in conversation, Auer (1984) deals with the 'middle ground' of the polarised other-initiation and self-repair. Many of his remarks concerning referring to an (absent) referent are valid for the data of the present work, in which the referent can be visible or the next action inferred. Auer finds hesitancy and try-marking as useful devices for testing out whether the other knows what the reference is, and other-initiation can then be invited for the speaker to self-repair, which I call self-initiated other-repair (as the speaker orients to the possibility of the other to disagree with his or her putative interpretation).

In my data, the potential 'challengable' and thus the uncertainty marking was not so much about testing the other's knowledge, but about probing whether he or she agreed with what the speaker's interpretation of the referent/next action was. This became obvious in Example 5(2), for instance, in which the *do we do that now?* questioning of the nonactive participant was replied with *we wanna add the sometimes?* (line 52), i.e. with a suggested interpretation of 'this next'. The extract showed how the person at the keyboard was trying to elicit a reply from the other participant, and how the other's reading aloud the screen (to himself, almost) was taken as a comment to the still unresolved 'this next'. In the end, the formulation with *we* changed into a suggestion *shall I shall I figure out sometimes on there do you think?* which in the end got a go-ahead from the partner (whose reluctance to take responsibility of the next action was made clear in his response *give it a try*). Thus, a continuer (cf. Auer 1984, 14) was actively sought for by the coparticipant, and therefore pushed the interactional device of try marking (the continuers of which are optional) towards an adjacency pair (offer/acceptance or question/answer): the uncertainty of one's actions would change a try marking to genuine agreement seeking.

Self-initiated other-repair seems to have stayed a fairly empty class of phenomena (word/name search) since Schegloff et al.'s first treatment of

repair work (Schegloff et al. 1977). If the self-initiated other-repair would cover Auer's 'intermediary' class (Auer 1984, 20), there would not be the need for an elaboration for the self/other initiation/repair dichotomy as such. For instance, formulations⁵⁷ such as *So you want to...* could sometimes be treated as self-initiated (potential) other-repairs.

8.4.3. Summary

The analysis of repair work that emerged in the different participation frameworks showed that in cooperative work and problem solving, the shared status of the effort and responsibility of the next action was reflected in the abundant use of uncertainty or self-initiated other-repair for moving on in the task. Thus, between humans, the status of the other person's knowledge was not to be underestimated but the other participant was treated as a potentially more knowledgeable. It was a task of equal 'we' solving a puzzle.

But, unlike in cooperative action, in which problems can be negotiated away by checking the other's interpretation, with predesigned systems, the erroneous actions are either repaired by the system (TELEPHONE, TUTORIAL), or discovered by the user (TASK). Also, the mediated nature of TEACHING resulted in turn-taking repairs that were qualitatively different from the cooperative preventive repairs (in the human-human interaction). Thus, in human-human interactions, doing uncertainty was a way of avoiding claiming expertise over the coparticipating other (cf. Goodwin 1986b, footnote 9), be it taking one's own knowledge or the other's inexpertness for granted. In all the cases, it became apparent that, in comparison with the cooperative repair (self-initiated other-repair), increasing degrees of specificity would be needed for repair initiators to work efficiently. The self-initiated other-repair of the type *so you want to...* from the system could preempt the user's possible repair after next turn.⁵⁸

With stand-alone language technology, asymmetries can be found that have an impact in the emerging repair work. A programmed system does not 'know' anything (it is programmed to make 'inferences' and give an output on the basis of keyed in sentence, mouse click or other inserted data). Thus, in TELEPHONE and TUTORIAL, the repair initiations by the program were a result of a simple pattern matching. However, this same asymmetry gave the systems the power of 'dictating the other' because the human participant had to change their behaviour to fit the program. Though the repair initiations by

⁵⁷ According to Heritage and Watson, "the primary business of formulations is to demonstrate understanding and, presumptively, to have that understanding attended to and, as a first preference, endorsed" (Heritage & Watson 1979, 138).

⁵⁸ In Luff and Frohlich (1990), repair initiation by the system was limited to two types of 'meta-questions' *What?* and *OK?*. They regard both types leading to other-initiated self-repair, though the initiator in the case *OK?* is not an other but the self.

the programs were in a way responding to the user's 'mistakes', there was no chance for the user to challenge the programs (other than laughing at them or skipping some exercises, for instance). The programs were regulating the human 'other' who had to try to make sense of them (cf. talking with a young child) on their terms (see e.g. Rommetveit 1985).

Thus, in both human-human and human-'other' interaction, expertise was interactionally constituted: in human-human participation frameworks through doing uncertainty, and in human-machine participation frameworks through the human party aligning to the system's 'needs'. Interpretation was shaped to be a shared phenomenon between equal human participants, whereas the human-'other' interaction became a site of one-way adapting to an 'other' in charge: the users would repeat the language use both at the level of lexicon and at the level of sequential functioning (e.g. orienting to repair as being done similarly as 'the last time'). Also, even with 'interactive' systems, but especially with manuals, the responsibility of interpretation was on the human participant who could check his or her earlier interpretation by matching it with the next instruction by the 'other': if nothing in the next request pointed towards the earlier one being misunderstood, then there was no need to repair one's understanding of a first position advice in the fourth position.

8.5. The interactionally constituted 'other' and assumptions of intentionality and sharing as prerequisites of meaning negotiation

In this section, the concept of 'other' is discussed. One way of pinning down the notion is to see how the user is as an other to the system, i.e. how the creators of the textual object, though not necessarily intentionally, positioned the user.

In TELEPHONE, the user had to work out how to manage the medium. Though there was a human mediator reading out the program's responses and keying in the user's turns, the program never referred to itself as 'I', even if the user was talked to as 'you'. Thus, the user was individualised at the level of dialogue: the program and the user were in a dialogic relationship.

In TUTORIAL, at the beginning of the encounter, the user's name was inquired (as an option). In a way, the name created a stronger 'contract' (e.g. Blakar 1992) for the praising comments, for instance, *Good!*. Thus, the user was individualised at the level of rapport. And, as the receiver was particularised, the feeling of 'communication' (vs. general information) was sought to be established. Through taking a position of giving evaluative comments, the program also exhibited a status of (a faceless) authority over the user. When an assessment is made of the human participant in human-text interaction, the dialogicity of the piece of text is at its highest. However, the 'sender', the 'other' of the message has no origo: is it maybe the company Microsoft that

values the client's skills? According to the first screen, the copyright of the program was that of Microsoft. No authors are mentioned, maybe to strengthen the illusion of the program as a participatory 'other' rather than reminding the reader of the original origo(s) of the words. When the mediated 'other' remains vague, not just absent, the responsibility for the words produced also becomes obscure. If the 'other' is an animated cartoon character (e.g. the help wizard 'clip' in Microsoft Word 97), or a figure of the past (wizard 'Shakespeare' in Microsoft Word 97), instead of, say, an animated picture of Bill Gates, maybe the user of the program will not see whose 'voice' he or she is in 'dialogue' with. It is hard to regard a funny paper clip as truly responsible for whatever it is 'saying' on the screen: the anonymity of the 'other' prevails, this time masked behind a cartoon figure.

In TASK, the human other was oriented to as a reader rather than user. Though the *you* used in the manual was dialogic as such (i.e. 'talking to the user'), it also had features of a more generic, i.e. narrational, 'you': the 'other perspective' taken in the manual was not as strong as, for instance, in giving directions with the help of a map "there *you* are in the upper-left corner of a rectangle" (Blakar 1992, 246). The difficulty of communicating through the manual might be due to lack of the user's perspective: "Only by the communicants taking the perspective of the other(s) into account, so that they (sender and receiver) may establish commonality, is communication rendered possible" (Blakar 1992, 246). The manual writers have a special difficulty of orienting to the sequential use and interpretation of the manual user. It was impossible for the manual to provide for the "last structurally provided defence of intersubjectivity in conversation" (Schegloff 1992), namely repair after next turn (i.e. after what the user did on the basis of manual's directive). However, the intersubjectivity between the manual writer(s) and the user-reader could be maintained by 'self-repair after the 'other's' next turn', i.e. the next directive in the set of directions also provided the reader clues about the status of her interpretation of the previous direction. Thus, when acting on the basis of written directions, the human interactant's interpretation is crucial, not only in regard to the present direction's 'action import' (Schegloff 1995, 204), but also in regard to what the present direction indirectly reveals about the meaning of the previous one.

The 'other's' perspective was clearly oriented to in the last case study, TEACHING. The mediated image of oneself (as the others would see it) and one's voice provided a possibility to orient to the receivers' perspective: the visual contact would shape the interaction, and the echo of one's voice at the other end would be a guarantee that the sound did carry over the line. Typing on the video picture was a representation of voice on a representation of visual image; after its production, the typed message was not a concern of the sender, but the receiver could make it part of the ongoing interaction later.

The power of the mediated 'others' as 'regulatory' texts (cf. Smith and Whalen 1995) was at its strongest in the case TELEPHONE, in which the user

'translated' the system's previous turn into her interpretation of it, repeating the lexicon and syntactic structures available for her not only in the system's turns, but also in the task sheet. The fast pace of the interaction meant that the user did not have time to ponder about 'what next', but had to try and adopt to the system's requests.

8.5.1. *Creating mutuality*

Normally, encounters have an interactionally marked beginning and an end. This also happened in the case studies of the telephone answering system (TELEPHONE) and in a way in the Learning Word 5.0 situation (TUTORIAL) as there was a greeting (*Welcome*) at the beginning, and there was a hypertextual menu from which a lesson was chosen (which was returned to automatically after the lesson had been finished). Also the seminar sessions in TEACHING were started and ended with greetings and goodbyes. However, the task of creating mailing labels (TASK), as far as the computer/manual interaction is concerned, did not really have a 'beginning' and an 'end', though the task itself had to be started and finished. Thus, TASK was most clearly, in the 'behaviour' of the mediated 'other', a transactional rather than interactional encounter. The text and the computer program were there as tools rather than active helpers, in contrast with a salutation at the beginning of a program that constructed it as a societal 'other'. However, for instance bank tellers clearly are used for transactional purposes, and yet they often display a greeting to the user. The aim seems to be that of a 'personalised' service: the user is treated as an individual and therefore the machine does better than many transactions (e.g. buying a ticket in a railway station) in which greetings do not always occur. Though the user cannot do what normally is done with a greeting: it is returned (which some people do jokingly even with a bank teller). Thus, the user is inhibited to something which "indicates that B is able and willing to share A's politeness by reciprocating it" (Graumann 1995, 1).

In the case study TELEPHONE the system greeted the user, the prerequisites of the participants sharing a common culture in which *hello* is an informal greeting that can be used towards the other. However, the informality of the greeting creates an even stronger requirement to return the greeting than would, for instance, a more formal 'good day'. When the system is an interactive written one, there are no consequences for the interaction, whereas with spoken systems that greet without expecting the second pair part from the user (as in TELEPHONE), problems can arise as in the fast pace *chronos* and *kairos* times of speech systems, every word uttered might matter in the evolving of the interaction.

Though the role of an initial greeting cannot be given too much weight without further investigation, it seems in TELEPHONE, the users show an

expectancy of interacting with an entity that behaves logically: the encounter can proceed if the user provides logical turns, i.e. repeats the words that the system exhibits or orients to the interactive style of the system as a logical one. Therefore, dialogical reciprocity, if not cognitive mutuality, seems to take place. There are, on the users' side, "expectations, however vague, of how the other partner will respond or otherwise contribute to the developing dialogue" (Graumann 1995, 19). How Graumann then describes the effect of anticipation, seems to fit with how the human participants in TELEPHONE oriented to the repair initiators as taking the encounter forward rather than initiating a side sequence: "In extreme cases, this may lead to hearing what one expects to hear rather than what our interlocutor actually utters and, hence, to responding to what one believes one has heard. Although this kind of 'exchange' is rather one-sided, it is still a case of assumed or apparent mutuality." (Graumann 1995, 19). He continues: "If we accept the cognitive usage of 'mutuality', we must concede that occasionally mutuality can exist only in the mind of a single person" (Graumann 1995, 19). Thus, interaction as a spatiotemporally shared, sequential, turn-by-turn phenomenon is contingent and highly indexical, especially if the interaction is about guiding the other's actions. The 'commonality', "everything that human (or infrahuman) beings have to share in order to communicate by means of symbols" (Graumann 1995, 19) is different when the symbols are not produced in shared time, but are 'voices of the past' trying to guide their receiver in a practical task. The 'other', whether understood as the original writer(s) of a text or a program or the text/program using language, can assume that the language used as a code is shared (though sometimes troublesources are oriented to, e.g. when computer terminology is used). However, 'grounding' one's instructions by using indexical language (the dialogic *you*, the cohesive *the*, the 'pointing' *this*, the interactive repair messages, etc.) cannot guarantee that the receiver shares the intention of the sender of these messages, i.e. that he or she interprets the repair initiator, for example, as indicating trouble in his or her previous interpretation.

The participation frameworks human-human and human-'other' (computer, text) differ in occupancy (Hanks 1996, 208): the occupation of the speaker and addressee roles is a process that, due to the mediated nature of the 'other', is nearer to the "entities whose characteristics are determined solely by the facts of language". However, whereas in corporeal occupancy "posture, gaze, gestures, physical proximity, tone of voice and timing" matter (Hanks 1996, 208), in mediated attendance, the body of the 'other' might not matter, whereas the physical proximity and timing have an impact on how the 'other' is perceived: in TASK and TUTORIAL, the language artefacts were close by and physically interacted with. Though there was a voice in TELEPHONE, the audible 'tone of voice' was monotonous and therefore was not of importance. And even if in video conferencing (TEACHING), mediated bodies could be seen, the distance from the others would be evident in the

language use of 'pointing at oneself' (*this is Santtu*). However, unlike in traditional video studio, the group interacting with the distant 'others' could physically be close to the screens. The distance was bigger, anyway, than that between the user and the language technology in the other three case studies. The mediated sound turned out to be the crucial factor in that the participants oriented to the computer with loudspeakers, even though not as a copresent 'here', at least as conveying a feeling of being 'more there' (in Sweden). So the space came to be meaningful not only as an 'us here and you there' arrangement, but, the room in Finland had zones of proximity: the 'others' in Sweden were virtually closer in one part of the room. Timing (both in *kronos* and in *chronos* time) was crucial in all the interactions which were considered successful if the task-at-hand was to be finished.

Not only is the 'other' interpreted through interactions: also 'self' becomes defined through the constraints and enables of technology-mediated communication. For instance, in TELEPHONE and TEACHING, oneself became objectified because of the 'forbidden pronoun' 'I' or the representation of oneself on the video picture. In both cases, the origo from which the human spoke was either prohibited to be mentioned or was displayed as a representation on a screen.

To grasp the construction of 'self' and 'other' in computer-mediated communication, in the following section, the concepts are approached from a semiotic point of view.

8.5.2. Aspects of 'other' as a sign foregrounded/made relevant in CMC interaction

Jensen's ideas of social semiotics were discussed in Chapters 2 and 6. He differentiates communication from interaction, considering the former as a negotiation of intersubjectivity about signs (i.e. meaning making), and the latter as one in which also other social agents (interpretant(s)) are negotiable signs: "social agents may redescribe each other — and their purposes and contexts — as both subjects and objects of action, ends and means of society. Others' description of who or what I am, in which contexts of action, implies what I can do." (Jensen 1995, 48).

Jensen's description of interaction is at a very general (societal) level, concerning the status of signs. Carbaugh (1996) gives a good overview of research concerning individual, self, or person: the individual as a neuropsychological being; self as symbolic interaction, self as social construction, person as culturally configured, and self in communication and rhetoric. After discussing the problems of these approaches, he proposes his own research agenda, namely Cultural Pragmatics, which consists of analyses of cultural dimensions, communication practice, cultural scene, social identity and identification. For the present study, his notions of communication practice and social identification are of specific interest, and the latter actually

comprises the former: "I use the terms *social identity* and *identification* as ways of drawing attention to two aspects of communicative practices. One is the use in communicative practices of particular *symbols of identity*, those that identify human agents. [--] The other aspect of practice of concern to us here are the communicative *forms* being used to enact social identities." (Carbaugh 1996, 16). In contrast with the (popular) idea of identity as concerning individuals, persons, the self as a distinguishable material entity (body), Carbaugh emphasises social identification, which he summarises as follows: (i) I/We/You/They (ii) know/show/constitute who I am/we are/you are/they are in part, by the way (i) I/we/you/they (iii) symbolise/perform/participate in situated social scenes.

Though Jensen and Carbaugh do not go into the details of semiotic action (semiosis), it is possible to use their ideas of social identification in summarising the encounters with(in) the mass media of instructional texts. In the present data, the 'others' address the user-readers in ways that sometimes elicit reactions that can be interpreted as reactions to the 'others'' description of the addressee(s). The interactive writing-as-speaking in TUTORIAL (e.g. *Just read now*) and in TASK (e.g. *Do not use the Toolbar*) were repairing the user's past action as an erroneous or potentially erroneous one. The language used was straightforward, directed to eliminating a wrong action. However, the turn was not spoken, but was a product that appeared in the process of making sense of the instructions. Though in copresent instructional situation both formulations might have worked well in the process of giving guidance, in these cases they prompted reactions of 'draining' (Goffman 1972). When the tutorial program tells the user to *Just read now*, after the user's keystroke, in spite of the *now* in the written sentence, it was not produced directly (synchronously), and therefore the implications of *now* as referring to not only the wrong timing of the user's keystroke, but also anything else that was going on that resulted in the keystroke, resulted in the system being called an *impolite thing* (Chapter 5). And, when the user finds the *Do not use the Toolbar* sentence as the relevant next to her just finished action of the first step of opening a new file, the asynchronous nature of the piece of text is even stronger (it does not pop up after the reader's action). The reaction is that of a surprise (*mitä?* 'what?') and astonishment, rather than a calm checking of the previous action as not being the one the user was warned against.

The conversation analytic method of investigating interactions at the minutiae of actual encounters and the interactional semiosis within the resources and restrictions of the communicative situation can be sketched diagrammatically as follows (as was done in Section 2.4).

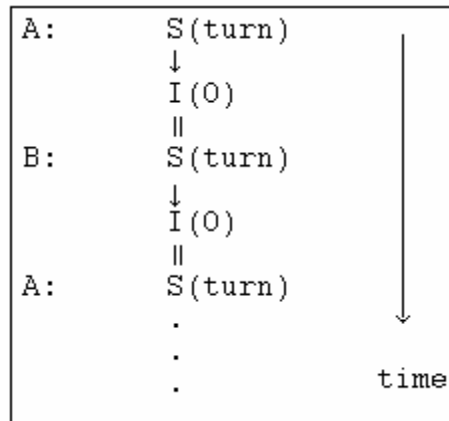


Figure 8-1

Sign S, which is paralleled with a turn in the diagram above, also covers the material (visual, audio, etc.) aspects of the encounter which are shared by human participants. However, in the case of predesigned computer programs or printed instructional texts, the intersubjective status of the sign/turn is challenged: in fact, even the medium has not concept of what 'this turn means now'. I stands for interpretant and O for object, as in the Peircean interpretation of semiosis, sign (S) relates to something (O) for someone (I) in some respect or capacity.

When Carbaugh's list of (i) I/We/You/They (ii) know/show/constitute who I am/we are/you are/they are in part, by the way (i) I/we/you/they (iii) symbolise/perform/participate in situated social scenes is contrasted with the process of semiosis sketched in Figure 8-1, the following equations can be made: (i) = Interpretant; (ii) = Object changing in time; (iii) = Sign (symbolic: language; performance: (language as) action; participation: S as a timed material manifestation of I(O)). If the constitution of a mediated 'other' is researched, it is not enough just to look at the language as a (referring) code as exhibited by the 'other', but also how the language is made to participate in the ongoing situation with a human participant. In the case of instructive language use, it is not possible just to constitute an object of interest, i.e. directions, warnings, etc.; at the same time, attitude to the 'other' is revealed (cf. Streeck 1980).

The interplay between the material and the semiotic in semiosis

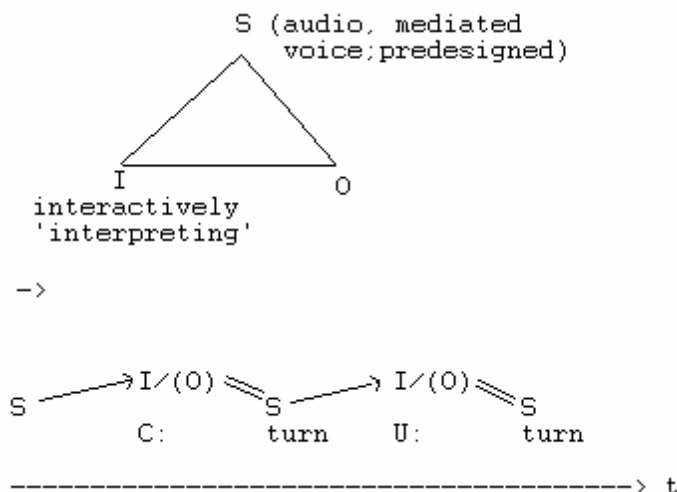
The sign could be regarded as an activity in time and space; its parts constituting that activity and, at the same time, the participation framework for each participant. S(ign) is the behaviour, the substance that is tangible to both participants; I(nterpretant) the inference that S causes in the receiver, and O(bject) the action that the receiver/producer of the next S is constructing. The placement of S in the sequence, and the foregrounded other S's (for

example the other participant(s); the screen), effect what the I/O will be in the next step. The participation framework is restricted and constructed by the interplay of the semiotic and the interaction of the participants.

I(nterpretant) is the receiver's interpretation (as intent; cf. Schiffrin 1990), the receiver's participation in interpreting language or other signs. It is the perlocutionary effect (Austin 1962, 110), the effect the sign causes in the receiver. O is the object of meaning negotiation, the actor's information (as his or her intent). O can be paralleled with an illocutionary act that produces the agent's turn (cf. Austin 1962, 110). Sign is then a locutionary act which is produced by the material sign vehicle. When a participant is interacting with a mediated 'other', the participation frameworks differ according to the activity of the 'other' as producing its turns-as-objects and interpretants in the ongoing activity. The sequential organisation of the activity in question is dependent on how the directives of the 'other' become inserted as turns in the interaction. Sometimes they are provided (TELEPHONE), sometimes the human participant interprets a (chosen) piece of text as a next directive.

The following pictorial representations describe roughly the restrictions and resources that the user/reader/interpreter had in each case study vis-à-vis the mediated 'other' as a communicative sign. As was concluded in Chapter 5, depictions of this sort can only describe the general conditions of semiosis, not the detailed moment-by-moment sense making. Thus the representations cannot depict situated activity systems, only the general semiotic resources of each case as an activity system. In the following picture, the lower part depicts the sequence of interpretation in a horizontal, rather than vertical fashion (cf. Figure 8-1).

TELEPHONE:



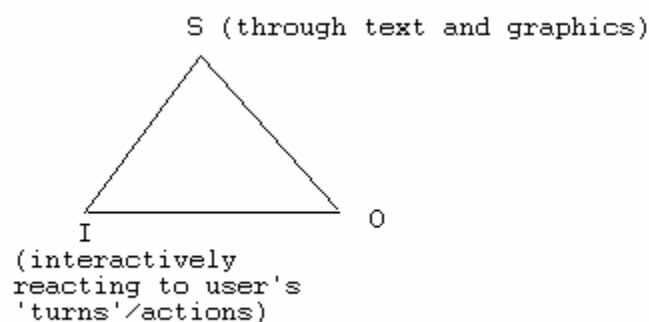
The activity was a task of leaving and listening to messages (as an experiment). The system produced next turns 'as if' interactively interpreting the user's turns. With the voice system especially, a sense of an individual

'other' with an origo was created, with a result of apparent interaction (interpretant/object creation), participation by the system. However, the computer's 'interpretation' did not often match the user's intent (object), as evident in retrials and repairs. Also, the system's repair initiations could be interpreted as another request, with the result of a type of 'garden path' (Suchman 1987, 165). In Suchman's study, the user's actions fitted with the system's 'repair initiation', i.e. the system had backtracked to a previous stage, but the user thought she was going forward in her task. In my study, the system's repair initiations were sometimes taken as a next instruction by the user, and, similarly to Suchman's data, the user thought she was proceeding in the interaction while the system was initiating one repair after another. The difference between the two cases is that with the photocopier, something was done (other than what the user thought), whereas with the telephone system, the smooth interaction would not have resulted in completing the task.

When the users did understand the system's turns as exhibiting problems with what the user had done before, they repeated and reformulated their requests, trying to produce a turn that would fit the 'other's' restrictions of formulating a request. Thus, the whole interaction could be labelled as 'how to do things with words', but at a much more concrete level than the Austinian idea of speech acts.

Through sequential organisation, participation frameworks, and encompassing activity structures, the mediated 'other' was constituted either as a puzzle to be solved or as an entity that was proceeding in the user's task, and thus for instance repair initiators would be 'read' as a next step in the interaction. So, either the right words had to be found or any words by the system would do.

TUTORIAL



The 'interaction' between the computer and the users could be sketched semiotically as in Figure 5-3, reproduced here as Figure 8-2.

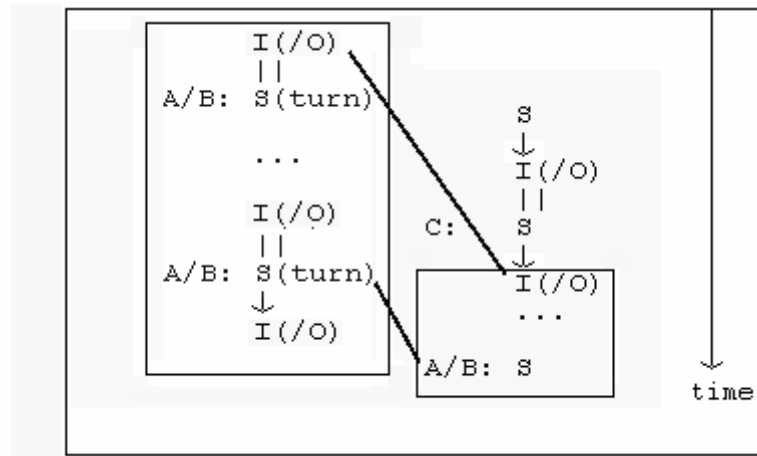
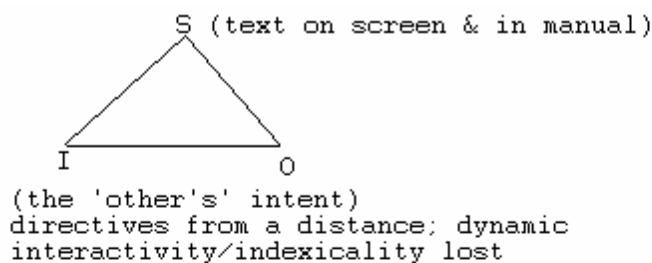


Figure 8-2

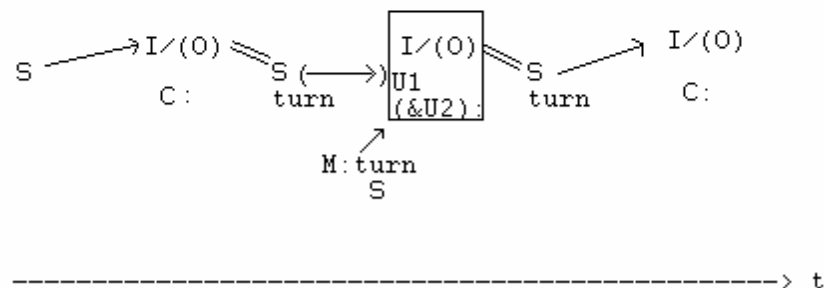
The mediated 'other' had an origo in the sense that it was individualising the user by asking his or her first name at the beginning of the session and then using the name in the next 'turn': *Thanks, Tim*. However, the Learning Word 5.0 was a hypertextual system that only reacted to the user's actions if they succeeded in practice or if the user's action was out of place.

Through sequential organisation, participation frameworks (human-human and human-computer), and encompassing activity structures, the mediated 'other' of the tutoring program was constituted as a logical entity.

TASK



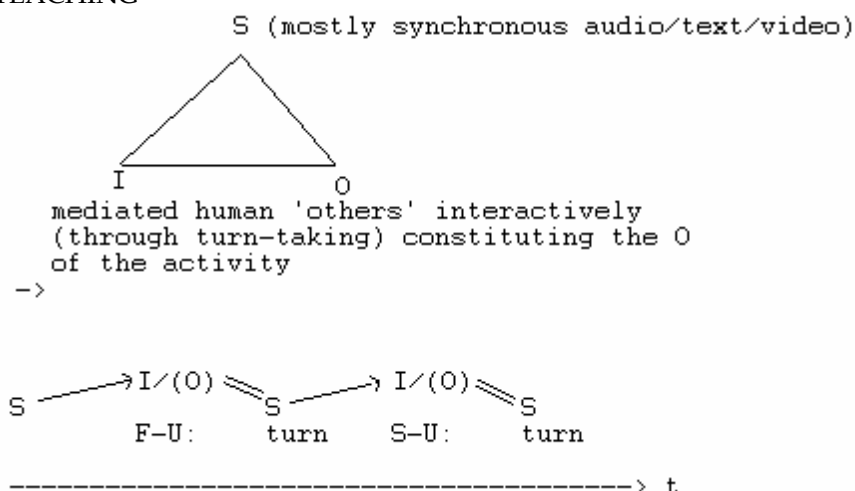
->



The graphic depiction tries to capture the interdependency of the user-manual and user-computer interactions: the computer's 'turns' are a result of a (mechanical) interpretation of the user's 'turn' (e.g. a mouse click) and the resulting sign is crucial in the user's interpretation of the next manual instruction. In TASK, both the word processing program and the manual were fairly generalised 'others': though the manual was talking to the user as a *you*, the user's name, for instance, was never brought up. The anonymous (chorus of?) 'others' were not individualised; even the word processing program was minimally interactive, it only gave options for a next step in the program, and only seldom was there a warning about what the user was doing or about to do. Thus, the role of the human participants (pairwork) was to constitute the O(bject) of the program; their role was most important in this case study as the manual and the program did not actively participate in the meaning making.

The manual's set of directives, however, provided the user a chance of analysing whether her previous action, i.e. the Interpretant of the manual's previous Sign had been correct: if the present directive was such that the state of the program (e.g. a dialogue box opened) allowed the user to do what was required to do, her previous action was indirectly confirmed as a correct one. Though the interpretants were thus produced asymmetrically, the user was in the position of a questioner or a learner: the program was a puzzle to be solved, and this time there was a manual from which the answers were to be found, with the human participants solely responsible for finding the answers.

TEACHING



The signs/turns of video mediated interaction were different from face-to-face interaction, because of the differences in the material circumstances of the encounter. The extended awareness of oneself as a metonym which at the

same time was the ‘others’ reception was one feature of the context that shaped the verbal interaction. Also, like in TUTORIAL and TASK, the groups at each end of the line often negotiated the next turn (sign). Thus, the encounter was much of how to maintain an activity, construct and develop an O(bject) through mediated signs that emphasise the ‘otherness’ of ‘others’, and of oneself.

With typed turns on the video picture, the turn could stay available longer, and would be oriented to as much more insistent than the written format of absent (anonymous) authors: it was a demand, not an offer artefact.

As the cases were different as instances of semiosis, uncertainties also differed; they concerned different levels of the interaction. In Section 2.3, the ‘other’ as a sign was discussed: the ‘other’ could, in Wiley’s categorisation, fulfil the three roles you-present-sign, you-future-interpretant, and you-past-object. The dialogic would emerge in interaction: the turn-at-talk (or action) being addressed to the present ‘other’, and that turn-at-talk being formulated on the basis of the past interaction with the ‘other’ (recipient design), with the future ‘other’ in mind (e.g. uncertainty). Turn-at-talk could be replaced by the more general turn-at-action, because not all relevant nexts are spoken.

In relation to sign as action or request/demand, and the orienting to the ‘other’ as a sign the following observations can be made.

- You-as-present sign: the S is addressed to the ‘other’, creating a participation framework; e.g. the role of the ‘other’ is to answer the request by action or by words. The technology of turn-taking is instantiated.
- You-as-past object: (the sign as) object formulated on the basis of what happened so far, resulting in recipient design through language use and other activities; the formulation of the request addressing the ‘other’ as an agent-on-the-basis-of-this/past interaction (cf. repetition/reformulation especially: referring to the ‘other’ as immediate past; extensive repetition ‘playing safe’, i.e. adopting to the machine (even if ‘interactive’ like in TELEPHONE, TUTORIAL); ways of referring: *it*, *impolite thing*). Assessments and categories are invoked in interaction: the ‘what’ and ‘who’ (the content and the humanness) are intertwined in the activity.
- You-as-future interpretant: the ‘other’s’ (next/future sign as) interpretant; uncertainty about what that interpretant will be in relation to the object put forward by the user.

The origo whose ‘you’ the other is, can be seen as the human participant (how s/he orients to the requests by the computer as signs of ‘you’); or, the system (how the system’s requests etc. position the user). The more separated the material sign is from the author/origo/sender, the less interactive the encounter (i.e. it might position the ‘other’ as words ‘only’ or as words

representing a concert of voices); the closer (interactive) the production of the material sign, the more individual the 'other'.

Also, the mediated 'other' was constituted through language use, especially by using various indexical formulations of requests. At the most easily detectable level, the occasioned usage of indexical expressions referring to the 'other' as an entity or as occupying a space or time: *it, you, here, there, now, then*. The mediated 'other' could be, through language use, established as absent or present, as near or far away in space. The use of *it* or *thing* (e.g. in TUTORIAL and TASK) about a computer that was being used, distanced the human participants from the interaction, and foregrounded the asynchronous, nonsharing aspect of the 'other'. In the case TELEPHONE, both *I* and *you* were used by the callers, constituting the mediated 'other' as a synchronous, sharing participant. And in TEACHING, the spatial *here* and *there* foregrounded the encounter as synchronous, but spatially nonshared.

If uncertain requests are explained as a semiotic phenomenon, uncertainty can be detected at the level of suggesting rather than imposing objects. Thus, self-initiated other-repairs can be analysed as requests for approval or disapproval (yes/no questions) of one's interpretation: uncertainty emerges at the level of negotiating the object. This happens in cooperative situations (when the interpretation of the 'other' is negotiated) between the human participants, revealing their orientation to the 'other' as an audience to which the 'other' performs. At these points, the users were not interacting with the system but rather about the system. The three computer systems in TELEPHONE, TUTORIAL and TASK were 'performers', i.e. repeating the author/designer's lines. In TEACHING, the mediated sound would sometimes 'perform' rather than participate (when the original speaker could hear herself 'answering' to a later turn), and the words on the video picture were 'performing' in losing the connection to the original typist. The oscillation between the human-computer and human-human participation frameworks meant shifting the attention to one or the other interaction (and, in TASK, there were three interactive spaces).

8.6. Summary

In all the cases examined in the present work, the 'other' was instructing the human participant(s) in their activities: 1) the answering machine in TELEPHONE by (re)directing the trajectory of the encounter, 2) the instructive program in TUTORIAL by advising the user-learner, 3) the Word 2.0 program in TASK by creating visual spaces that enable and limit the user's possibilities of action, and the manual by offering advice on how to proceed to achieve certain types of document. In 4) TEACHING, the disembodied text on the video picture, though having become asynchronous like the texts with anonymous origos in the three other cases, still had distinguishable 'others' as

the 'senders' of the message. The 'other' of the case studies was 'talking' to the users all the time, giving directions or making requests. The users were inclined to 'listen to' the 'other', as they needed the information to proceed in what they were doing.

In a way, all the three programs of the case studies (hypothetical telephone answering system in TELEPHONE, Learning Word 5.0 in TUTORIAL, and Word 2.0 in TASK) were like the electronic form of the call-taker in Smith and Whalen's study: "even before it is filled in, its formatting into various fields or slots projects an a priori sense of what that work could involve" (1995, 17). Also, the interactively different interfaces (in the present research 'talk-time', 'slow motion' and 'frozen') all limit the user in what they can do with the system, ranging from the mediated 'other's' forceful "this next" to the reader/user's possibility to ask "what next?".⁵⁹ However, unlike with the electronic form of Smith's and Whalen's research, the user did not know the scope of the requests at one glance, but had to go through the program bit by bit and 'answer' to its directions request by request (be they about a new action or repeating/repairing the previous one).

The four case studies showed how mediated, sometimes interactional, requests and other language use were interpreted by human users in various human-language technology encounters. Visuality and text were intimately related in some text types, e.g. instructions of all sorts, the inherent nature of the task being not just understanding certain words or commands in the abstract but doing something in the world according to those instructions. Also, the users' depiction of what was going on did not appear only in their turns-at-talk, but in their gaze and pointing, for instance, in which the effect of material phenomena over the semiotic was at its most concrete: the talk could not be fully understood without the bodily gesture, and vice versa.

Directions usually have a referent, an object that the direction is about: in TELEPHONE, the target was metacommunicative, instructing the user what to say and how to say it to the telephone answering system. In TUTORIAL, the user was trained both about how to use certain functions of a word processor, and how to use the Learning Word 5.0 program in question. In TASK, the word processor used was real (Word for Windows 2.0), the use of which was guided by a separate printed manual. And, in the last case study TEACHING, requests, directions and instructions were mediated through video picture, sound and text from others that shared the time (though often with a lag) but not the place of the situation. As the circumstances differed in each case study, different solutions were used and possible to use in each encounter. For instance in TASK, the manual designers had used pictures and icons, together with the layout possibilities of text to help the process of identifying what the text is talking about. Pictorial depictions of the visual environment can be paralleled with the typed names on the video picture in

⁵⁹ Thus, also human-'other' 'turn-taking systems' adapt to "the properties of the sorts of activities in which they operate" (Sacks et al. 1978, 8).

the video conference introductions: there is a one-to-one connection with a stable, easily recognised meaning.

However, in printed manuals, pictorial representations cannot be used throughout a text to guarantee the right interpretation (in electronic hypertextual format, linking every potential troublesource to an explanatory node is easy enough; e.g. the dilemma over *Toolbar* would have been easily solved with an extension to an iconic or verbal depiction). But, as was shown in the case studies, it is not only the ease of denotation that can guarantee an interactional success of a request: whether synchronous or asynchronous, a direction is produced to create a perlocutionary, visible or hearable effect. The interactional sequence preceding the reading or hearing and acting upon a text always brings new elements to interpretation. This is maybe why increasingly more manuals are not going to be printed on paper, but implemented in the actual program to be adhered to from a detectable sequence which the computer can then give instructions about. However, as was shown in case studies TUTORIAL, TELEPHONE and TEACHING, too much interpretative weight should not be put on the strength of the placement of a piece of language in a sequence to bring about its intended interpretation. Language technologies provide environments for encounters with mediated asynchronous or synchronous 'others' in which requests, due to the limits of meaning negotiation, need to be specified much more carefully than in face-to-face interactions.

Interpreting and acting upon instructions is a local phenomenon. Also, the interpretative freedom is very narrow (if the aim of the 'directee' is to 'obey the intention of the other'). In synchronous language use, it is possible to check and negotiate meanings, and to repair wrong interpretations 'on the fly'. The sequence of interpretation is a crucial factor, highlighting the importance of time in language phenomena. Printed text in a manual is a trace of a past activity, instruction giving that makes mutual negotiation of meaning impossible, whereas in copresent activity, language is moulded in and by the unfolding situation, as an indexical and integral part of shaping the activities, and uncertainties are resolved cooperatively.

Written texts for action cannot change as such, they have to be made to work by new interpretations-in-time: the situation unfolds around written language. Uncertainty has to be resolved by the reader, the user, the receiver of the text on his or her own. One piece of text can thus become a recurring resource for new structures or interpretations, as happened in Example 7(28) (*TALK NOW!!!!!!!!!! PHIL*). The piece of text on the video picture in the second interpretation was activated, brought back to present by the human participants who made it part of the ongoing interaction, just like with the texts in TELEPHONE, TUTORIAL and TASK.

Both synchronous and asynchronous texts (as verbal and visual, e.g. pictures and graphs) have symbolic, iconic and indexical dimensions: they can be descriptive (of an event), they can take the interaction forward

(technically), and they can offer sites for (meaning) negotiation. However, though the descriptive stance was shown to cause difficulties in printed texts-for-action (Chapter 6), separating the three levels is also hard in interactive language, as indicated in the complexities of the (wrong) interpretation of *Click-L* in Chapter 5. The meaning of the acronym was 'negotiated' by the program; *Click-L* was a metaphor of clicking the left button of the mouse, and the interpretation of it was indexical, depending on the changing visual and interactional context. When the writers of a manual tend to depict, in a text internally coherent way, the 'context of use' as a fictitious event on paper, the designers of computer tutorials transfer the fiction into partly interactive interfaces designed for fictitious action, as was shown by the program's repair initiator *Please check your spelling...* (Extract 5(5), lines 167-169) after a completely wrong word had been typed in: the programmers of the tutorial had designed a repair initiator that only oriented to typos rather than misunderstanding/misreading the direction being a potential problem at this point of the tutorial.

As the encounters in the case studies often involved typed or written language, reading research was also relevant, especially studies on reading and computers. However, there is a distinct difference between the reading of fiction (whether in paper or interactive electronic format) and the present case studies in which the 'readers' or 'users' had committed themselves to act on the basis of the requester's instructions. As in all human interaction, what the 'other' demanded was not enough to secure action on the basis of the interpretation: each 'saying' was part of an unfolding interactional situation in which not only the linguistic context, but the material environment, as well as the individual 'layers of experience', played a role. Although literary research (e.g. Aarseth (1997)) could not explain the success of actions-in-the-world with which a user of a guiding text was occupied, it offered insights to the troubles exhibited by the traditional format of these texts when 'activated' (Smith & Whalen 1995, Watson 1997).

In the search for efficient interactivity in new language technology environments, there has to be an understanding of how 'making texts talk' can be accomplished best. The four case studies examined the resources that people make use of when they encounter 'as if' spoken texts in telephone, computer tutorial, manual and word processor, and video conferencing environments. In each environment, the participants used a variety of communicative resources to manage the interaction with the mediated 'other'. In the TELEPHONE case study, the user adapted to the language of the system in the 'on-line' unfolding of the encounter; in TUTORIAL, they adapted to the interactive style of the tutorial program; in TASK, they overcame the fixed narrative of the manual and objectified the so-called interactive messages of the word processor; and in TEACHING, they made use of the same text in both synchronous and asynchronous modes.

The 'other' of the language technology encounters can be 1) nonproblematically assumed, i.e. the user-readers engage in an activity that is constructed with another 'participant', and, at the same time the 'other' (and the user-reader) are constructed during the encounter (the 'language use' aspect); or, 2) the 'other' is marked specially through reference (*you/it* etc.) (the 'language used' aspect). Streeck's notion of social events and human relations being interactively constituted by speech acts could be seen as another way of describing the intertwined nature of human communication (Streeck 1980). The level of participation reveals how the participants position themselves (e.g. as audience), and through this the positioning of the 'other' can be found out (e.g. as a tutorial rather than a tutor). In the minutiae of the interaction different aspects of the 'other' come to be foregrounded: when in TUTORIAL one of the participants said *it hasn't told us anything*, the 'other' was an animated tutor with an origo; whereas in saying *go to next screen*, the 'other' was treated as a tutorial screen. Thus, 1) and 2) resemble the difference between traditional linguistic problematics and (speech) action theories of language use and referring. In the end, they are inseparable, as the referent gets its situated, i.e. 'real' meaning in each occasion of use. If the users are referring to the system as *it says/does*, they are combining the tool and symbolic aspects of the artefact. And, when language use is encountered, it is difficult not to describe a machine as 'saying' or 'telling' something. Not too much attention should be paid to this phenomenon in terms of personifying the computer. Of much more concerning is, if the users' 'turns' become abstracted away by mediation such that the language is oriented to as an object rather than a means of coconstructing meaning. Thus, rather than classifying the user-reader or the 'other' according to what participatory roles they seem to be occupying, we should examine what the interactional consequences of the language technologies are for the encounter as a language using occasion.

The case studies showed that language as a system (linguistics), human cognition (cognitive science) and activities in the world (social sciences) cannot be separated if a meaningful explanation for human-language technology interaction is sought for. When human-computer or human-computer-mediated-human interaction is researched as a sense making activity, language and other signs that are encountered come to mean different things not only on the basis of their sequential positioning in the unfolding interaction, but, also, according to the pace of that interaction, and due to the visual space and its meaning potential in the complex continuity of the encounter.

9. CONCLUSION

This study demonstrated that Streeck's observation about the 'meaning' of speech acts holds true for all signification: "What counts in the realm of human communication is not the acts as they are intended by the speakers, but the consequences they have; that is, how they are interpreted and subsequently responded to by other participants" (Streeck 1980, 145). Any sign-using occasion is constructed by sequential interpretation, the 'technology' of which brings about the social event, but also contributes to human relations (*ibid.*). When the encounter takes place with or through language technology, the shape of the social event differs from that between copresent humans: the semiotic fields activated are special to the material circumstances, and so is the process of interaction, also due to the various levels of interactivity of the technologies. Thus, communication in language technology environments is a complex process which can be described and analysed from several aspects, as indicated by the discussion about the present study in the summarising Chapter 8.

Interpretations are constructed between participants, and the situated, sequential character of face-to-face interaction has been successfully demonstrated by conversation analytic research. Research on written communication concerns mainly the text-internal coherence as a device to ease the interpretative process of the reader. What I have demonstrated in this dissertation is that user-text interaction becomes sequential when the text occupies the reader-in-action. The interpretative work is dissimilar to face-to-face interaction, as the user-reader is the only participant that can actively construct the meaning of the encounter. Consequently, the shape of repair work also becomes different. In face-to-face encounters, the machinery of interaction allows 'on-line' checking of interpretations. To understand possible and actual difficulties in user-text interaction, instead of concentrating only on text-internal characteristics, it is important to find out about the material and interactional enablements and constraints for the checking of interpretations.

Thus, sequentiality of action resembles that of talk-in-interaction which is context-sensitive and context-renewing. In language technology environments, the communicative resources available are different. For

instance, when mediated language is guiding action, context-sensitivity is easily accomplished as the user-reader's activity brings forth a piece of language, either by the program or by the user's active selection of it (e.g. by reading a manual). A word processor provides its users with an array of options (in the form of icons, for example) of what to do. When the user then clicks an icon, another, more restricted set of options is offered: there are no 'wrong' interpretations from the program's point of view at each dialogue box, and it is totally up to the user to deem whether his or her trajectory of action is as desired. When a manual is used to help decide 'what next', each new screen becomes a resource to understand the manual's directives, and *vice versa*.

The present work shows that it is much harder for mediated language to be context-renewing, taking the event forward, if a change in the direction of the encounter is aimed at for instance by repairing. A recommendation was made that the status of the repair should be clearly stated, either by formulating what the problem was, and/or by clearly marking the 'turn' as a corrective (by a 'label' or by a recognisable repair initiation formulation). The problem with written manuals as instructions is that the context-shaping and renewing of 'turns' tends to be text-internal, as part of the iconic narrative, i.e. a description of a hypothetical course of action. The burden of 'externalising' the interpretation is solely on the reader: not only has she to decide which piece of text is relevant for the moment (to activate it and thus make it context-sensitive), but also the user's interpretation can be confirmed – either directly or indirectly – by the following directive that the action undertaken is in accordance with the previous directive (i.e. the context-renewing aspect of the previous one was as intended). For instance, a click of an icon on a word processor screen might create an action space with a dialogue box, another icon in which the user is then asked click for the next piece of manual instruction. Thus, although texts cannot offer a next turn repair initiator on the basis of the user's action in the world, the reader can use the resource of directions to check her previous understandings. Human-(directive) text interaction has therefore a different shape of repair work in which one's interpretations are not checked by the 'other', but by oneself, though through 'interacting' with the text.

Thus, the analysis of the case studies demonstrated that conversation analysis is an effective method to find out how exactly the communicative resources affect the interpretation process. The research did not rely on observations made on the basis of the researcher's (predefined) categories, but the "active agent perspective" was taken (cf. Ginsburg 1985, 267). When the interaction took place through telephone, the detectable impact of the visual surroundings was restricted to the instruction sheet. But when the use of a language artefact itself (whether computerised or printed) relied on reading and seeing, and the actions of the participants were 'silent' mouse or keyboard clicks, then the coparticipants became important as people to whom

verbal and other accounts of understanding were offered. Although the interactive constitution of action as such is interesting, in this work it also supplied a window to how the 'other' was oriented to. Thus, all the cases provided good data for studying "significant use of signs and symbols" (Jensen 1998, 4) in new technology contexts.

The 'other' in face-to-face interaction is perceived utterance-by-utterance; however, in printed texts, the sequentiality is distorted because the manual, for instance, can be read from any point (cf. hypertext). This means that the reader has to achieve sequentiality, they have to decide what the 'other' is going to say to them in each concrete moment of the interaction. Heritage and Watson point out how the sequence so far affects what the next can be: "There is thus a sense in which some preceding utterance may be said to provide a constraint on the production of some next utterance: indeed if this were not grossly the case the sequential analysis of conversation would prove impossible and it would be hard to find a rationale for referring to some sequentially ordered collection of utterances as "a conversation"" (1979, 139-140). All in all, the complexity of context boundedness became clear in the present study because the computerised surroundings force people to include in the interaction the 'other' (re)presented in the spatiotemporal unfolding of the situation.

The communicative resources in the various case studies differed, but in all the encounters, formulations of existing linguistic materials (instruction sheets, text on the screen or in the manual) were abundant, and were used to take the situation forward or to remedy troubles. In formulations, repair work, and other speech events, preferred vs. dispreferred status is an important conversation analytical finding. However, in the case of self-initiated other-repair, it is not so much the preferred status of agreeing or self-initiation that is at stake but the fact that there is a choice, a possibility for the other to choose, however insincere this manifestation of uncertainty on the part of the speaker would be. When a nonhuman 'other' (computer, book) is encountered, uncertainty is geared to the other human participant who usually comments, agrees or disagrees, or even takes the part of the 'other' in reading aloud what he or she understands to be the problematic formulation. This way, the meanings of the 'other' of the text surface as negotiable – not so much the person behind those meanings. But, at the same time, they bring the 'other' to the unfolding situation as a contributor.

Another aspect of instructing is the inherent asymmetry that instructions/requests can bring about by initiating first pair parts of adjacency pairs. Though power as realised in communication is a very complex and even evasive phenomenon, I would be willing to argue that, from the perspective of empowerment, that even more important than the possibility to initiate a first pair part of an adjacency pair is the chance to initiate repair. Repair work can be regarded as a technology to help build social events, but it also has interactional consequences in human relations. In language

technology environments, the possibility to negotiate meanings through repair would increase 'the ease of use', but also the power to take responsibility for meanings would be enhanced. For instance, in desktop video conferencing, the sometimes unclear sound connection became a site in which the expertise in the language and/or subject became threatened because not only was the availability of sound connection, but also competence, interactionally constituted via language use (cf. Goodwin 1986b, 292-293). If repair orients to a past phenomenon, then in a technology-mediated classroom it can be about understanding the 'other's' language or topic, not just the availability of signs as such. Thus, the mediated language, both spoken and textual, together with the visual information, comprising the technologised encounter, can result in potential tensions in and worries about the social relations. The distorted time, space and sound has to be made to work in doing appropriate time, space and sound.

In self-initiated other-repair by a system, technically, other-initiation is prevented (as the system self-initiates), but, at the level of interaction, the other (the user) is responsible for her meaning, and also for repair (other-repair). By orienting to the user as the expert to judge the 'interpretation', a computerised system can exhibit interactionally that the information is not authoritative but available for negotiation by knowledgeable users, and not just receivers, of the information: self-initiated (by the system) other-repair aligns to the indexical uncertainty of the encounter as arising from the system, not the user. Checking the system's 'interpretation' is user-friendly as well, since the interactionally least wanted and, as shown by the present work, not always successful other-initiation (by the system) of repair is avoided.

Self-initiated other-repair (by the system) is also 'system-friendly'. The task of building fully capable dialogue systems is a very difficult one as, in semiotic terms, the machine's Interpretant of the user's Interpretant/(Object) always has to be limited: possible troublesources are hard to detect. However, if the aim is to build working systems for people to accomplish actions in the world (with so-called public computer systems, e.g. Grönlund (1994)), an important question is who is in charge of the activity. A computerised system can manage its work if the system can control what is happening in the encounter, i.e. if it is a 'demand' artefact, providing the first pair part of adjacency pairs, and controlling henceforth the process of the encounter. But an 'offer' artefact can allow the user abundant responsibility within a clearly defined domain. This can be done if the system is checking via self-initiated other-repair what its 'understanding' of the encounter so far is: the system offers a possibility to accept or reject the system's current interpretation of what is going on in this specific interaction, thus avoiding repair which orients to the past interaction, and if initiated by the human 'other', can be very complex since anything in interaction can be a repairable.

If a system can evade initiating repair of a past phenomenon, the problem of 'nonconstructive' repair initiators, i.e. "the error is not manifest in

such a way as to be visible" (Suchman 1987, 184) can be avoided. If the checking question of a self-initiated other-repair is answered with a 'no', then the system has a better possibility of 'guessing' the trouble source. In noninteractive, text-based instructions about another entity, the way of approaching the problem of accessibility of singular terms or words is to express them in an electronic hypertext format as links which can be clicked for more information. Hypertext links free the instruction taker from a linear learning process that traditional instructive text and talk support by explaining (teaching) in a narrative first the (terms of the) parts in the object of action, then the actions that can be undertaken (see the telephone instructions found in Whalen & Vinkhuyzen quoted in Chapter 8). With hypertext, the instruction can start from the action as the terms can be given a potential explanation via links. This way, the user is given an opportunity (and encouraged) to ask: what does X mean? The hypertext format supports 'interactivity' in this sense. Also, users with differing levels of expertise (layers of interpretants) can be advised with one textual format, and the extra information does not have to be found elsewhere.

The 'next step' — doing the action instead of explaining it — has been taken, for example, by the designers of Word 7.0 *Microsoft Word Help Topics*. Sometimes the program guides the user through the chain of actions by actually performing them. For instance, the question *How do I add page numbers?* is answered by the 'invisible hand' choosing *Insert* and *Page Numbers* on the screen, and providing the user with some information about the dialogue box opened. However, within a limited (self-referring) domain like a text processor, it should be possible to create an interface in which the *Help Topics* feature would become a 'What do you want to do?' interface: the user types in (or says) to the program what they want to do instead of asking how they can do something. The step-by-step doing (and learning) process can be skipped altogether. The system could then interactively 'ask' the user about the specific wishes (e.g. the size of mailing labels). The wizard and agents of Word 97 already exhibit some of this type of interaction, but the bulk of the instructions are given in a hypertext narrative format. If the user of a word processor, say, could start the program by interacting with a separate dialogue entity (be that in a format of a graphic agent or just a 'faceless' interactive text), the object of action (the document to be formatted) could be kept separate, and the results of the hypothetical formatting could be shown (in the way they are displayed in the 'dialogue boxes' of the present programs). By directing the interpretation work from the users' ability to act upon instructions to their ability to use language interactionally, the system resembles that of the telephone answering system in Chapter 4. However, interactivity should not be sought for at any cost if the material form of the instructions allows for less ambiguous and interactionally preferred ways of dealing with uncertainty: instead of being guided with other-initiated other-repairs, the user can explore the definitions of words at his or her own

initiative: the pitfalls of on-line spoken language could be avoided if the questions displayed by the system had hypertext links to possible troublesources for the user. As written instructions (of use) sometimes do not succeed in conveying a monolithic reading of them, the most effective way for guiding action would be a visual video depiction of the process. In this way, the 'reader' is forced to have a monolithic understanding of the directions and ensuing actions. Thus, video clips could be inserted in the written general information via links, also making a clear distinction between when to 'read' and when to do. This way, the instruction writers present tendency to produce a somewhat misleading written guidance for action in the form of a description of a scenario could be useful for a script producing a video of the enacted action.

According to Lombard and Ditton, on the basis of our traditional experiences we may "be more likely to feel that we are dealing with a social entity when we use (interact with) an automatic teller machine or an educational computer program than when we use a database or word processing program" (1997, 32). Indeed, the results of Chapter 6 indicate that the answers given to the queries posed by the word processor were indicative of treating the interaction as more like 'pushing a button' than 'answering a question'. However, the objectifying of language might not be due to it appearing on the traditional computer tool but because it is represented on another representation, a button. If the user had to type in the answer, to produce the language by herself, then the interaction might be constructed as dealing with a social entity, even if the 'other' was a word processor.

This dissertation explores from a 'microanalytic' perspective various user-reader encounters with language technology. Individuals, pairs, and groups make sense of 'textual voices' in their 'positioned-practices' (Thibault 1991, 241). I think that it is only in these concrete encounters that the 'discursive formations' (Thibault 1991, 241) of the Information Society come alive and are maintained (or changed): whatever the hypothesis of the impact of new media on communicative practices in the activity systems of work life, education and leisure, the only way of getting to the heart of the matter is to see what precisely is happening in the actual use of technologies. My case studies show how the material and semiotic are intertwined in meaning making. Indeed, extrasign semiotic materiality proves to be influential in sequential interpretation. Interactive activities in and through language technology are shaped by the material circumstances.

One conclusion already well-known in the humanistically oriented human-computer interaction research community is that there should be more cooperation between designers of technology, especially for public consumption, and researchers into human practice. By so doing, language technologies that can orient to the time and space of the language used might emerge, resulting in appropriate timing and forms of 'turns at action'. There is

a difference between reading and encountering language technology for action: the latter is a semiotically complex situation in which sequentiality of action and the materiality of language (text in space, voice) matter. Therefore, theories of reading and visual layout are not enough to explain the success of a language technology as an instructor for action.

It is important to recognise that language technologies support active, inquisitive social agency: the nonauthoritarian 'other' preferably has identifiable authors, whatever characters are designed to perform the interaction with the user. The aim for the designers of language technology should be to provide a robust means for 'participation' by the user. However, the decision makers of the Information Society should seek for technologies that provide a channel for real participation, i.e. giving feedback to those that design what is said. Although language artefacts can make some encounters quicker and easier, citizens should not surrender to the 'miracle' and the authority of 'computers that talk like you and me'. Public computer systems are on the increase, and they should be researched not only as sites of smooth communicative events, but also as enabling or constraining the possibilities of social and political participatory action.

References

- Aarseth, E.J. (1997) *Cybertext: Perspectives on ergodic literature*. Baltimore: John Hopkins University Press.
- Agre, P. (1988a) *The dynamic structure of everyday life*. Unpublished PhD thesis. Dept. of Electrical Engineering and Computer Science. Univ. of Oxford.
- Agre, P. (1988b) *Writing and representation*. MIT AI Laboratory, Working paper 315.
- Andersen, P.B. (1990) *Computer semiotics*. Cambridge: Cambridge University Press.
- Anderson, J.A. & Meyer, T.P. (1988) *Mediated communication. A social action perspective*. Newbury Park: Sage.
- Angiolillo, J.S., Blanchard, H.E., Israelski, E.W. & Mané, A. (1997) Technology constraints of video-mediated communication. Finn, K.E., Sellen, A.J. & Wilbur, S.B. (eds) *Video-mediated communication*. New Jersey: Lawrence Erlbaum Associates. 51 – 73.
- Auer, J.C.P. (1984) *Referential problems in conversation*. Department of Sociology, University of Manchester. Occasional paper no. 13.
- Austin, J.L. (1962) *How to do things with words*. Oxford: Oxford University Press.
- Baecker, R.M. & Buxton, W. (1987) Cognition and human information processing. Baecker, R.M. & Buxton, W. (eds) *Readings in human-computer interaction*. Los Altos: Morgan Kaufman Publishers.
- Bakhtin, M.M. (1986) *Speech genres and other late essays*. Austin: University of Texas Press.
- Ball, M.S. & Smith, G.W.H. (1992) *Analyzing visual data*. Qualitative research methods series 24. Newbury Park: Sage.
- Barfield, L., van Burgsteden, W., Lanfermeijer, R., Mulder, B., Ossewold, J., Rijken, D. & Wegner P. (1994) Interaction design at the Utrecht School of the Arts. *SIGHI bulletin* 26, 49-86.
- Barnes, S. (1997) Cyberspace: Creating paradoxes for the ecology of self. Strate, L. Jacobson, R. & Gibson, S. (eds.) *Communication in cyberspace: Social interaction in an electronic environment*. Cresskill: Hampton Press. 193 – 216.
- Barthes, R. (1977) *Image, music, text*. London: Fontana.
- Beach, W.A. (1990) Language as and in technology: Facilitating topic organization in a videotex focus group meeting. Medhurst, M.J., Gonzales, A. & Peterson, T. R. (eds) *Communication & the culture of technology*. Pullman: Washington State University Press.
- Biocca, F. (1997) The cyborg's dilemma: Progressive embodiment in virtual environments. *JCMC* 3 (2). [Online]. Available at: <<http://www.ascusc.org/jcmc/vol3/issue2/biocca2.html>> [Date accessed 22 October, 1997].
- Blakar, R.M. (1992) Towards an identification of preconditions for communication. Wold, A.H. (ed) *Dialogical alternative: Towards a theory of language and mind*. Oslo: Scandinavian University Press.
- Boden, D. (1990) The world as it happens: Ethnomethodology and conversation analysis. Ritzer, G. (ed) *Frontiers of social theory*. New York: Columbia University Press. 185 – 213.
- Boden, D. and Molotch, H.L. (1994) The compulsion of proximity. Friedland R. & Boden D. (eds) *NowHere. Space, time and modernity*. Berkeley: University of California Press, Berkeley. 257 – 286.
- Bolter, J.D. (1997) Virtual reality and the redefinition of self. Strate, L., Jacobson, R. & Gibson, S. (eds) *Communication in cyberspace: social interaction in an electronic environment*. Cresskill: Hampton Press. 105 – 119.
- Bonk, C.J. & Cunningham, D.J. (forthcoming), Searching for learner-centered, constructivist, and sociocultural components of collaborative educational learning

- tools. Bonk & King (eds.) *Electronic collaborators: Research the discourse of learner-centered technologies*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bonk, C.J. & King, K.S. (forthcoming) Electronic conferencing and collaborative writing tools: starting a dialogue about student dialogue. Bonk C.J. & King, K.S. (eds.) *Electronic collaborators: Research the discourse of learner-centered technologies*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bordieu, P. (1977) *Outline of a theory of practice*. Cambridge: Cambridge University Press.
- Borning, A. & Travers, M. (1991) Two approaches to casual interaction over computer and video networks. Robertson, S.P., Olson, G.M., and Olson J.S., (eds) *Reaching through technology. CHI '91 Conference proceedings*. New Orleans, Louisiana: Addison-Wesley. 13–19.
- Bowers, J.M. (1991) Time, representation, and power/knowledge. *Theory & psychology* 1(4): 543–569.
- Bowers, J.M. & Churcher J. (1988) Local and global structuring of computer mediated communication: Developing linguistic perspectives on CSCW in COSMOS. CSCW 88. *Proceedings of the Conference on Computer-Supported Cooperative Work*. Baltimore: ACM, p. 125-139.
- Bowers, J., Pycock, J. & O'Brien J. (1996) Talk and embodiment in collaborative virtual environments. *CHI'96 Electronic Proceedings*. [Online]. Available at: <http://www.acm.org/sigchi/chi96/proceedings/papers/Bowers/jb_txt.htm> [Date accessed 11 November, 1997]
- Boyd, J. (1992) The act in question. Parret, H. & Verschueren, J. (eds) *(On) Searle on conversation*. By John Searle et al. John Benjamins, Amsterdam. 31–34.
- Bruce, V. (1996) The role of the face in communication: Implications for videophone design. *Interacting with Computers* 8(2), 166–176.
- Burman, E. & Parker, I. (1993) *Discourse analytic research. Repertoires and readings of texts in action*. London: Routledge.
- Button, G. (1993) (ed) *Technology in working order. Studies of work, interaction, and technology*. London: Routledge.
- Button, G. & Sharrock, W. (1995) On simulacrum of conversation: Toward a clarification of conversation analysis for human-computer interaction. Thomas, J. (ed) *The social and interactional dimensions of human-computer interfaces*. New York: Cambridge University Press. 107–125.
- Buxton, W.A.S. (1997) Living in augmented reality: Ubiquitous media and reactive environments. Finn, K.E., Sellen, A.J. & Wilbur, S.B. (eds) *Video-mediated communication*. New Jersey: Lawrence Erlbaum Associates. 363–384.
- Campbell, J. (1990) Observer and object, reader and text: Some parallel themes in modern science and literature. Slade, J.W. & Yaross L.J. (eds) *Beyond the two cultures. Essays on science, technology and literature*. Ames: Iowa State University Press.
- Canter, D. (1985) Intention, meaning, and structure: Social action in its physical context. Ginsburg, G.P., Brenner, M., von Cranach, M. (eds) *Discovery strategies in the psychology of action*. London: Academic Press, London. 171–186.
- Carbaugh, D. (1996) *Situating selves: the communication of social identities in American scenes*. Albany: State University of New York Press.
- Cheepen, C. (1988) *The predictability of informal conversation*. London: Pinter Publishers.
- Cole, R., Hirschman, L., Atlas, L., Beckman, M., Biermann, A., Bush M., Clements M., Cohen, J., Garcia O., Hanson B., Hermansky H., Levinson S., McKeown K., Morgan N., Novick D.G., Ostendorf S. O., Price P., Silverman H., Spitz J., Waibel A., Weinstein C., Zahorian S. & Zue V. (1995) The challenge of spoken language systems: Research directions for the nineties. *IEEE (The Institute of Electrical and Electronics Engineers) Transactions on Speech and Audio Processing* 3(1), 1–21.
- Cook, G. (1995) Principles for research in text accessibility. Nyysönen, H. & L. Kuure (eds) *Principles of accessibility and design in English texts. Research in progress*.

- Publications of the Department of English, University of Oulu. A12 Text and Discourse Studies. 9–18.
- Coulter, J. (1991) Is the 'new sentence problem' a genuine problem? *Theory & Psychology* 1(3), 317–336.
- Coulter, J. (1994) Is contextualising necessarily interpretive? *Journal of Pragmatics* 21(6), 689–698.
- Crystal, D. (1995) *The Cambridge encyclopedia of the English language*. Cambridge: Cambridge University Press.
- Davis, B.H. & Brewer, J.P. (1997) *Electronic discourse. Linguistic individuals in virtual space*. Albany: State University of New York Press.
- De Fornel, M. 1992, The return gesture: Some remarks on context, inference, and iconic gesture. Auer, P. & Di Luzio, A. (eds) *The contextualization of language*. Amsterdam: John Benjamins. 159–176.
- Denzin, N.K. & Lincoln, Y.S. 1994, Introduction: Entering the field of qualitative research. Denzin, N.K. & Lincoln, Y.S. (eds) *Handbook of qualitative research*. Thousand Oaks, CA: Sage. 1–17.
- Dourish, P. & Bly, S. (1992) Portholes: supporting awareness in a distributed work group. Bauersfeld, P. Bennett, J & Lynch, G. (eds) *CHI '92 conference proceedings. ACM Conference on human factors in computing systems*. Monterey: ACM, p. 541-547.
- Eco, U. (1979) *The role of the reader*. Indiana University Press. Bloomington.
- Ehlers, T.S. & Steinfield, C.W. (eds) (1992) *ITCA teleconferencing yearbook 1992*. Washington, D.C.: International Teleconferencing Association Research Committee.
- Eiser, J. R. (1985) Social cognition and social action. Ginsburg, G.P, Brenner, M., von Cranach, M. (eds). *Discovery strategies in the psychology of action*. London: Academic Press. 187–206.
- Engeström, Y. & Middleton, D. (1996) Introduction: Studying work as mindful practice. Engeström, Y. & Middleton D. (eds) *Cognition and communication at work*. New York: Cambridge University Press. 1–14.
- Erickson, F. & Shultz, J. (1982) *The counselor as gatekeeper. Social interaction in interviews*. New York: Academic Press.
- Fairclough, N. & Wodak, R. (1997) Critical discourse analysis. van Dijk, T. (ed) *Discourse as social interaction*. London: Sage. 258–284.
- Finn, K.E., Sellen, A.J. & Wilbur, S.B. (eds) (1997) *Video-mediated communication*. New Jersey: Lawrence Erlbaum Associates.
- Firth, A. (1998) *Exploring the intersections of talk, texts, and (work) tasks*. Paper presented at the AAAL conference, Seattle, USA, 14th March, 1998.
- Fish, R.S, Kraut, R.E. & Root, R.W. (1992) Evaluating video as technology for informal communication. Bauersfeld, P. Bennett, J & Lynch, G. (eds) *CHI '92 conference proceedings. ACM Conference on human factors in computing systems*. Monterey: ACM, p. 37–48.
- Fiske, J. (1990) *Introduction to communication studies*. London: Routledge.
- Flynn, P.J. (1991) *The ethnomethodological movement. Sociosemiotic interpretations*. Berlin: Mouton.
- Freire, M.M. (1995) A socio-cultural/semiotic interpretation of intercommunication mediated by computers. *The Proceedings of International Conference L.S. Vygotsky and the Contemporary Human Sciences*. September 5.-8. 1994, Moscow. [Online]. Available at: <<http://www.glas.apc.org/~vega/vygotsky/freire.html>> [Date accessed 12 November, 1996].
- Friedland, R. & Boden D. (1994) NowHere: An Introduction to Space, Time and Modernity. Friedland R. & Boden D. (eds) *NowHere. Space, time and modernity*. Berkeley: University of California Press, Berkeley. 1–60.
- Frohlich D., Drew P. & Monk A. (1994) Management of repair in human-computer interaction. *Human-computer interaction* 9(3/4), 385–425.

- Frohlich, D. & Luff, P. (1990) Applying the technology of conversation to the technology for conversation. Luff, P., Gilbert, N. & Frohlich, D. (eds) *Computers and conversation*. London: Academic Press. 187–220.
- Fulk J., Schmitz J.A. & Schwarz D. (1992) The dynamics of context-behaviour interactions in computer-mediated communication. Lea, M. (ed) *Contexts of computer-mediated communication*. London: Harvester/Wheatsheaf. 7–29.
- Gaver, W, Moran, T, MacLean, A, Lövsstrand, L, Doursh, P, Carter, K. & Buxton, W. (1992) Realizing a video environment: Europarc's Rave system. Bauersfeld, P. Bennett, J & Lynch, G. (eds) *CHI '92 conference proceedings. ACM Conference on human factors in computing systems*. Monterey: ACM, p. 27–35.
- Ginsburg, G.P. (1985) The analysis of human action: Current status and future potential. Ginsburg, G.P, Brenner, M., von Cranach, M. (eds). *Discovery strategies in the psychology of action*. London: Academic Press. 255–279.
- Goffman, E. (1971) *Relations in public: Microstudies of the public order*. New York: Harper & Row.
- Goffman, E. (1972) *Encounters*. Harmondsworth: Penguin.
- Goffman, E. (1979) Footing. *Semiotica* 25(1/2): 1–29.
- Goodwin, C. (1979) The interactive construction of a sentence in natural conversation. Psathas, G. (ed) *Everyday language: studies in ethnomethodology*. New York: Irvington. 97–121.
- Goodwin, C. (1981) *Conversational organization*. New York: Academic Press.
- Goodwin, C. (1986a) Gestures as a resource for the organization of mutual orientation. *Semiotica* 62(1/2), 29–49.
- Goodwin, C. (1986b) Audience diversity, participation and interpretation. *Text* 6(3), 283–316.
- Goodwin, C. (1994) Professional vision. *American Anthropologist* 96/3, 606–633.
- Goodwin, C. (1995a) Co-constructing meaning in conversations with an aphasic man. *Research on language and social interaction* 28(3), 233–260.
- Goodwin, C. (1995b) Seeing in depth. *Social studies of science* 24(2), 237–274.
- Goodwin, C. (1996) Transparent vision. Ochs, E., Schegloff, E.A. & Thompson, S.A. (eds) *Interaction and grammar*. Studies in interactional sociolinguistics 13. New York: Cambridge University Press. 370–404.
- Goodwin, C. (1998) *Embodiment within talk-in-interaction*. Course materials for the Summer Linguistics course 'Embodiment within talk-in-interaction' in the International Graduate School in Language and Communication, Odense, 22-26.6.1998.
- Goodwin, C. (in press) Pointing as situated practice. Kita, S. (ed) *Pointing: Proceedings of the Nijmegen Max Planck Conference*. Cambridge: Cambridge University Press.
- Goodwin, C. & Goodwin, M.H. (1987) Concurrent operations in talk: Notes on the interactive organization of assessments. *IprA Papers in Pragmatics* 1(1): 1–55.
- Goodwin, C. & Goodwin, M.H. (1992) Context, activity and participation. Auer, P. & Di Luzio, A. (eds) *The contextualization of language*. Amsterdam: John Benjamins. 77–99.
- Goodwin, C. & Goodwin, M.H. (1996) Seeing as situated activity: Formulating planes. Engeström, Y. & Middleton, D. (eds) *Cognition and communication at work*. New York: Cambridge University Press. 61–95.
- Goodwin, M.H. (1980) Process of mutual monitoring implicated in the production of description sequences. *Sociological Inquiry* 50(3/4), 303–317.
- Goodwin, M.H. (1990) *He-said-she-said: Talk as social organisation among black children*. Bloomington: Indiana University Press.
- Goodwin, M.H. (1995) Assembling a response: Setting and collaboratively constructed work talk. ten Have, P. & Psathas, G. (eds) *Situated order. Studies in the social organization of talk and embodied activities*. Washington, D.C: International Institute

- for Ethnomethodology and Conversation Analysis & University Press of America. 173 – 186.
- Graumann, C.F. (1995) Commonality, mutuality, reciprocity: a conceptual introduction. Marková, I., Graumann, C.F. & Foppa, K. (eds) *Mutualities in dialogue*. Cambridge: Cambridge University Press. 1 – 24.
- Greenlee, D. (1973) *Peirce's concept of sign*. The Hague: Mouton.
- Grice, H.P. (1975) Logic and conversation. Cole, P. & Morgan, J.L. (eds) *Syntax and semantics 3: Speech acts*. New York: Academic Press.
- Grönlund, Å. (1994) *Public computer systems, the client-organization encounter, and the societal dialogue*. Research reports in information processing and computer science 18:94. Department of Informatics, Umeå University.
- Gumperz, J.J. (1995) Mutual inferencing in conversation. Marková, I., Graumann, C.F. & Foppa, K. (eds) *Mutualities in dialogue*. Cambridge: Cambridge University Press. 101 – 123.
- Gurvitch, G. (1990) The problem of time. (From Gurvitch, G. 'The Spectrum of Social Time'. Dordrecht: Reidel 1964) Hassard, J. (ed) *The Sociology of time*. London: MacMillan.
- Hagège, C. (1990) *The dialogic species: A linguistic contribution to the Social Sciences*. New York: Columbia University Press.
- Halliday, M.A.K. (1973) Linguistic function and literary style: An inquiry into the language of William Golding's "The Inheritors". Halliday, M.A.K. (ed) *Explorations in the functions of language*. London: Arnold. 103 – 143.
- Halliday, M.A.K. (1984) Language as social semiotic. Fawcett, R.P, Halliday, M.A.K, Lamb, S. & Makkai, A. (eds) *The Semiotics of Culture and Language. Volume 1. Language as Social Semiotic*. London: Frances Printer.
- Halliday, M.A.K. (1993) Language and the order of nature. Halliday, M.A.K. & J.R. Martin (eds) *Writing science: literacy and discursive power*. London: Falmer Press. 106 – 123.
- Halliday, M.A.K. (1994a) Spoken and written modes of meaning. Graddol D. & Boyd-Barrett O. (eds) *Media texts: Authors and readers*. Clevedon: Multilingual Matters/The Open University.
- Halliday, M.A.K. (1994b) *An introduction to functional grammar*. London: Arnold.
- Halliday, M.A.K. & Hasan, Ruqaiya (1976) *Cohesion in English*. London: Longman.
- Halliday, M.A.K. & Martin, J.R. (eds) (1981) *Readings in systemic linguistics*. London: Batsford.
- Hanks, W.F. (1991) Series Foreword. In Lave, J. & Wenger, E. *Situated learning. Legitimate Peripheral Participation*. Cambridge University Press, New York.
- Hanks, W.F. (1992) The indexical ground of deictic reference. Goodwin, C. & Duranti, A. (eds) *Rethinking context: Language as an interactive phenomenon*. Cambridge: Cambridge University Press. 43 – 76.
- Hanks, W.F. (1996) *Language and communicative practices*. Boulder: Westview Press.
- Harré, R. (1991) The discursive production of selves. *Theory & Psychology* 1(1), 51 – 63.
- Harris, R. (1996) *Signs, language and communication: Integrational and segregational approaches*. London: Routledge.
- Hassard, J. (ed) (1990) *The Sociology of time*. London: MacMillan.
- Hauptmann, A.G. & Rudnicki, A.I. (1988) Talking to computers: An empirical investigation. *The International Journal of Man-Machine Studies* 28, 583 – 604.
- Heath, C. (1986) *Body movement and speech in medical interaction*. Paris: Cambridge University Press.
- Heath, C. (1992) Gesture's discreet tasks: Multiple relevancies in visual conduct and in the contextualisation of language. Auer, P. & Di Luzio, A. (eds) *The contextualization of language*. Pragmatics and Beyond 22. Amsterdam: John Benjamins. 101 – 127.

- Heath, C. (1997) The analysis of activities in face to face interaction using video. Silverman, D. (ed) *Qualitative research. Theory, method and practice*. London: Sage. 183–200.
- Heath, C. & Luff, P. (1991) Disembodied conduct: Communication through video in a multi-media office environment. Robertson, S.P., Olson, G.M. & Olson, J.S. (eds) *Human factors in computing systems. Reaching through technology. CHI '91 conference proceedings*. New Orleans: ACM, p. 99–103.
- Heath, C., Luff, P. & Sellen, A. (1995) Reconsidering the virtual workplace: Flexible support to collaborative activity. Marmolin, H., Sundblad, Y. & Schmidt, K. (eds) *Proceedings of the fourth European conference on computer-supported cooperative work*. Dordrecht: Kluwer Academic Publishers. 83 – 99.
- Heath, C., Luff, P. & Sellen, A. (1997) Reconfiguring media space: Supporting collaborative work. Finn, K.E., Sellen, A.J. & Wilbur, S.B. (eds) *Video-mediated communication*. New Jersey: Lawrence Erlbaum Associates. 323-347.
- Heritage, J. (1984a) A change-of-state token and aspects of its sequential placement. Atkinson, J.M. & Heritage, J. (eds) *Structures of social action*. Cambridge: Cambridge University Press. 299–337.
- Heritage, J. (1984b) *Garfinkel and ethnomethodology*. Cambridge: Polity Press.
- Heritage, J. (1989) Current developments in conversation analysis. Roger, D. & Bull, P. (eds) *Conversation: An interdisciplinary perspective*. Clevedon: Multilingual Matters Ltd. 21–47.
- Heritage, J. & Watson, D. (1979) Formulations as conversational objects. Psathas, G. (ed) *Everyday language. Studies in Ethnomethodology*. New York: Irvington Publishers. 123–162.
- Herring, S.C. (ed) (1996) *Computer-mediated communication. Linguistic, social and cross-cultural perspectives*. Amsterdam: John Benjamins.
- Hodge, R. (1988) Discourse in Time. Torode, B. (ed) *Text and talk as social practice*. Dordrecht: Foris.
- Hockett, C.F. (1977) *The view from language*. Athens: University of Georgia Press.
- Hodge, R. & Kress, G. (1988) *Social semiotics*. New York: Cornell University Press.
- Hollan, J. & Stornetta, S. (1992) Beyond Being There. Bauersfeld, P., Bennett, J. & Lynch, G. (eds) *CHI '92 Conference Proceedings. ACM Conference on Human Factors in Computing Systems. Striking a Balance*. May 3-7, 1992, Monterey, California. New York: ACM.
- Holquist, M. (1990) *Dialogism. Bakhtin and his world*. London: Routledge.
- Hopper, R. (1990) Telephone speaking and the discovery of conversation. Medhurst, M.J., Gonzalez, A. & Peterson, T.R. (eds) *Communication and the culture of technology*. Pullman: WSU Press.
- Hopper, R. (1991) Hold the phone. Boden, D. & Zimmerman, D.H. (eds) *Talk and social structure. Studies in Ethnomethodology and Conversation Analysis*. Cambridge: Polity Press. 217–231.
- Hutchins, E. (1995) *Cognition in the wild*. Cambridge: The MIT Press
- Isaacs, E.A., & Tang, J.C. (1997). Studying video-based collaboration in context: From small workgroups to large organizations. Finn, K.E., Sellen, A.J. & Wilbur, S.B. (eds) *Video-mediated communication*. New Jersey: Lawrence Erlbaum Associates. 173–197.
- Ishii, H. & Kobayashi, M. (1992) ClearBoard: A semaless medium for shared drawing and conversation with eye contact. Bauersfeld, P., Bennett, J. & Lynch, G. (eds) *CHI '92 Conference Proceedings. ACM Conference on Human Factors in Computing Systems. Striking a Balance*. May 3-7, 1992, Monterey, California. New York: ACM, p. 525–532.
- Isomursu, T. (1997) 'In a way it's a problem you don't know that data terminology' – On the relationship between vocabulary and text accessibility. Nyysönen, H. & Kuure, L. (eds) *Principles of accessibility and design in English texts. Research in progress*

2. Publications of the Department of English, University of Oulu. A14 Text and Discourse Studies.
- Jaques, E. (1982) *The form of time*. New York: Crane Russak.
- Jefferson, G. (1972) Side sequences. Sudnow, D. (ed) *Studies in social interaction*. New York: The Free Press. 294–338.
- Jefferson, G. (1983) *Issues in the transcription of naturally-occurring talk: Caricature versus capturing pronunciation particulars*. Tilburg University. Tilburg papers in language and literature 820.08.203.
- Jefferson, G. (1989) Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation. Roger, D. & Bull, P. (eds) *Conversation: An interdisciplinary perspective*. Clevedon: Multilingual Matters. 166–196.
- Jensen, K.B. (1995). *The Social Semiotics of mass communication*. London: Sage.
- Jensen, S.S. (1998) *Human actors and software factors. "Netviews" - qualitative method of networking interviews*. Paper presented in Netværk for Læring og Interaktive Multimedier seminar in Silkeborg, Denmark, 4.-6. May, 1998.
- Jordan, M.P. (1994) Toward plain language: a guide to paraphrasing complex noun phrases. *Journal of Technical Writing and Communication* 24/1, 77–96.
- Kendon, A. (1985) Behavioural foundations for the process of frame attunement in face-to-face interaction. Ginsburg, G.P, Brenner, M., von Cranach, M. (eds). *Discovery strategies in the psychology of action*. London: Academic Press. 229–253.
- Kendon, A. (1992) The negotiation of context in face-to-face interaction. Goodwin C. & Duranti, A. (eds) *Rethinking context: Language as an interactive phenomenon*. Cambridge: Cambridge University Press. 326–334.
- Kitzinger, C. & Wilkinson, S. (1996) Theorizing representing the Other. Wilkinson, S. & Kitzinger, C. (eds) *Representing the Other. A Feminism and Psychology reader*. London: Sage. 1–32.
- Kleifen, J.A. & Frenz-Belkin P. (1997) Assembling knowledge. *Research on Language and Social Interaction* 30/2, 157–192.
- Kress, G. (1993) Against arbitrariness: The social production of the sign as foundational issue in critical discourse analysis. *Discourse & Society* 4:2, 169–191.
- Kress, G. & van Leeuwen T. (1996) *Reading images: The grammar of visual design*. London: Routledge.
- Kuzuoka, H. (1992) Spatial workspace lollaboration: A sharedview video support system for remote collaboration capability. Bauersfeld, P., Bennett, J. & Lynch, G. (eds) *CHI '92 Conference Proceedings. ACM Conference on Human Factors in Computing Systems. Striking a Balance*. May 3-7, 1992, Monterey, California. New York: ACM, p. 533–539.
- Labov, W. & Fanshel, D. (1977) *Therapeutic Discourse. Psychotherapy as Conversation*. New York: Academic Press.
- Langham, D. (1994) The common place MOO: Orality and literacy in virtual reality. *Computer-mediated communication magazine* 1(3), 7. [Online]. Available at: <<http://sunsite.unc.edu/cmc/mag/1994/jul/moo.html>> [Date accessed
- Lash, S. & Urry, J. (1994) *Economies of Signs & Space*. London: Sage.
- Laurel, B. (ed) (1990) *The art of human-computer interface design*. Reading, Massachusetts: Addison-Wesley.
- Laurel, B. (1991) *Computers as theatre*. Reading, Massachusetts: Addison-Wesley.
- Lave, J. & Wenger, E. (1991) *Situated learning. Legitimate Peripheral Participation*. New York: Cambridge University Press.
- Laver, J. (1975) Communicative functions of phatic communion. Kendon, A. (ed) *Organization of behavior in face-to-face interaction*. The Hague: Mouton. 215–238.
- Leary, T. (1990) The interpersonal, interactive, interdimensional interface. Laurel, B. (ed) *The art of human-computer interface design*. Reading, Massachusetts: Addison-Wesley. 229–233.

- Lemke, J.L. (1993) Discourse, dynamics and social change. *Cultural Dynamics* 6(1): 243 – 275.
- Lemke, J.L. (1995) *Textual politics. Discourse and social dynamics*. London: Taylor & Francis.
- Leppänen, S. (1995) Analysis and interpretation: Theoretical positions, subjective interpretations and the persuasion of language. Nyysönen, H. & Kuure, L. (eds) *Principles of accessibility and design in English texts. Research in progress*. Publications of the Department of English, University of Oulu A12, Oulu. 161 – 206.
- Levine, J. M. (1990) PRAGMA – a flexible bidirectional dialogue system. *Proc. The eighth national conference on artificial intelligence, AAAI-90*, Boston, USA, vol 2, p. 964 – 969.
- Levinson, S. C. (1988) Putting linguistics on a proper footing: Explorations in Goffman's concepts of participation. Drew, P. & Wootton, A. (eds) *Erving Goffman. Exploring the interaction order*. Cambridge: Polity Press.
- Lewis, D.J. & Weigart, A.J. (1988) The structures and meanings of social-time. (From Lewis, J.D. & Weigart, A.J. 'The Structures and Meanings of Social Time'. *Social Forces*, 60, 1981. 432 – 62). Hassard, John (ed) *The Sociology of time*. London: MacMillan.
- Linell, P. (1998) Discourse across boundaries: On recontextualization and blending of voices in professional discourse. *Text* 18:2, 143 – 157.
- Lombard, M. & Ditton, T. (1997) At the heart of it all: The concept of presence. *JCMC* 3(2). [Online]. Available at: <http://www.ascusc.org/jm/vol3/issue2/lombard.html> [Date accessed 22 October 1997].
- Luff, P., Gilbert, N. & Frohlich, D. (eds) (1990) *Computers and Conversation Analysis*. London: Academic Press.
- Manning, P.K. (1987) *Semiotics and fieldwork*. Qualitative research methods series 7. Newbury Park: Sage.
- Marková, I. (1994) The mutual construction of asymmetries. van Geert, P., Leendert P.M. & Baker, W.J. (eds) *Annals of Theoretical Psychology* 10. New York: Plenum Press. 325 – 341.
- Martin, J. (1992) *English text, system and structure*. John Benjamins, Philadelphia.
- McHoul, A. (1997) Cyberbeing and ~space. V/I/D. [Online]. Available at: <http://kali.murdoch.edu.au/cntinuum/index1.html> [Date accessed].
- McIlvenny, P.B. (1990) *Dialogue and the machine: An interactional perspective on computer dialogue models, mediation and artifacts*. Unpublished PhD thesis, Univ. of Edinburgh, Depts. of Linguistics and Artificial Intelligence.
- McIlvenny, P.B. (1995) Virtuaalinen luokkahuone mahdollistaa opiskelun eri paikoissa samanaikaisesti (Virtual classroom makes simultaneous studying possible in different places). *Tietopisto. Oulun yliopistoväen tiedotuslehti* 11, 14 – 17.
- Medway, P. (1996) Virtual and material buildings. Construction and constructivism in architecture and writing. *Written Communication* 13:4, 473 – 514.
- Meier, C. (1998) *In search of the virtual interaction order: investigating conduct in video-mediated work meetings*. Institute für Soziologie, Universität Giessen Arbeitspapier Nr. 3.
- Minsky, M. (1975) Frame-system analysis. Johnson-Laird, P.N. & Wason, P.C. (eds) *Thinking: Readings in Cognitive Science*. London: Cambridge University Press.
- Muhlbach, L., Bocker, M., & Prussog, A. (1995) Telepresence in videocommunications: A study of stereoscopy and individual eye contact. *Human Factors* 37(2), 290 – 305.
- Mulkay, M. (1985) Agreement and disagreement in conversations and letters. *Text* 5(3), 201 – 227.
- Murray, D.E. (1988) The context of oral and written language: a framework for mode and medium switching. *Language in Society* 17, 351 – 373.

- Mårdsjö, K. (1994) Man — text — technology: technical manuals as means of communication. Steehouder, M., Jansen, C., van der Poort P. & Verheijen, R. (eds) *Quality of technical documentation*. Amsterdam: Rodopi. 185–200.
- Nass, C. & Steuer, J. (1993) Voices, boxes, and sources of messages. Computers and social actors. *Human Communication Research* 19(4), 504–527.
- Nass, C., Steuer, J. & Tauber, E.R. (1994) Computers are social actors. Adelson, B., Dumais, S. & Olson, J. (eds) *ACM conference on human factors in computing systems. CHI '94 conference proceedings*. Boston: ACM/SIGCHI, p. 72–77.
- Norman, D.A. (1988) *The psychology of everyday things*. New York: Basic Books, Inc. Publishers.
- Nystrand, M. (1992) Social interactionism versus social constructionism: Bakhtin, Rommetveit, and the semiotics of written text. Wold, A.H. (ed) *Dialogical alternative: towards a theory of language and mind*. Oslo: Scandinavian University Press. 157–173.
- Nyyssönen, H. & Kuure, L. (1995) *Principles of accessibility and design in English texts. Research in progress*. Publications of the Department of English, University of Oulu. A12 Text and Discourse Studies.
- Nyyssönen, H. & Kuure, L. (1997) *Principles of accessibility and design in English texts. Research in progress 2*. Publications of the Department of English, University of Oulu. A14 Text and Discourse Studies.
- Nöth, W. (1995) *Handbook of Semiotics*. Bloomington: Indiana University Press.
- Nöth, W. (ed) (1997) *Semiotics of the media: State of the art, projects and perspectives*. Berlin: Mouton de Gruyter.
- Ochs, E. (1979) Transcription as theory. Ochs, E. & Shieffelin, B. (eds) *Developmental Pragmatics*. New York: Academic Press.
- Ochs, E., Schegloff, E.A. & Thompson, S.A. (eds) (1996) *Interaction and grammar. Studies in interactional sociolinguistics* 13. New York: Cambridge University Press.
- Parkes, D. and Thrift, N. (1978) Putting time in its place. Carlstein, T., Parkes, D. & Thrift, N. (eds) *Making sense of time*. Bungay: Edward Arnold. 119–129.
- Perrault, C.R. & Allen, J.F. (1980) A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics* 6, 167–182.
- Peräkylä, A. (1997) Reliability and validity in research based on transcripts. Silverman, D. (ed) *Qualitative research. Theory, method and practice*. London: Sage.
- Petrilli, S. (1990) Appenix II: On the materiality of signs. Ponzio, A. *Man as a sign. Essays on the philosophy of language*. Berlin: Mouton de Gruyter. 367–401.
- Pilto, R. & Rapakko, T. (1995) Testing accessibility of utility texts - work in progress. Nyyssönen, H. & Kuure, L. (eds) *Principles of accessibility and design in English texts. Research in progress*. University of Oulu. Publications of the Department of English. Series A, Text and discourse studies nro 12. Oulu. 37–58.
- Ponzio, A. (1985) Semiotics between Peirce & Bakhtin. *Kodikas/Code* 8(1/2): 11–28.
- Ponzio, A. (1990) *Man as a sign. Essays on the philosophy of language*. Berlin: Mouton de Gruyter.
- Prawitz, D. (1990) Tacit knowledge — an impediment for AI?. Göranson, B. & Florin, M. (eds) *Artificial intelligence, culture and language: On education and work*. London: Springer-Verlag. 57–59.
- Quirk, R. & Greenbaum, S. (1988) *A University grammar of English*. Harlow: Longman.
- Rahko, K. (1991) *Comparative study of the use of videoconferencing in the Pacific area and in the Nordic countries*. Helsinki University of Technology. Laboratory of Telecommunication switching and information technology. Report 3/91. Otaniemi Finland.
- Raudaskoski, P. (1990) Repair work in human-computer interaction. A conversation analytic perspective. Luff, P., Gilbert, N. & Frohlich, D. (eds) *Computers and Conversation Analysis*. London: Academic Press. 151–171.

- Raudaskoski, P. (1992) *Repair work in the context of English language mediated computer interface use: A conversation analytic study*. Unpublished Lic.Phil. thesis, Univ. of Oulu, Dept. of English.
- Raudaskoski, P. (1995) These signs here now. Nyysönen, H. & Kuure, L. (eds) *Principles of accessibility and design in English texts. Research in progress*. Publications of the Department of English, University of Oulu A12, Oulu. 111 – 118.
- Raudaskoski, P. (1996) Copresence in Video Conference Mediated Interaction. Harakka, T. & Koskela, M. (eds) *Kieli ja tietokone (Language and computer)*. AFinLA vuosikirja. AFinLA Series 54. 133 – 148.
- Raudaskoski, P. (1997) Semiosis at computer media. : Nöth, W. (ed) *Semiotics of the media: State of the art, projects, and perspectives*. Berlin: Mouton de Gruyter.
- Reeves, B. & Nass, C. (1996) *The media equation. How people treat computers, television, and new media like real people and places*. New York: CLSI Publications & Cambridge University Press.
- Rogers, H.G. & Brown, F.W. (1993) The impact of writing style on compliance with instructions. *Journal of Technical Writing and Communication* 23/1, 53 – 71.
- Rommetveit, R. (1985) Language acquisition as increasing linguistic structuring of experience and symbolic behavior control. Wertsch, J.V. (ed) *Culture, communication and cognition: Vygotskian perspectives*. Cambridge: Cambridge University Press.
- Ryle, G. 1975 (1949). *The concept of mind*. London: Hutchinsons of London.
- Sacks, H., Schegloff, E. & Jefferson, G. (1978) A simplest systematics for the organization of turn taking for conversation. Schenkein, J. N. (ed) *Studies in the organization of conversational interaction. Volume I*. New York: Academic Press. 7 – 55.
- Sallio, P., Boudeville J.P. & Cabon, P. (1983), Visual ergonomic of a future videoconference service. Proceedings of the tenth international symposium on human factors in telecommunications. Helsinki, HFT.
- Santamäki, H. (1984) *Videoneuvottelutekniikan nykytilanne*. Technical research centre of Finland. Research notes 277. Espoo.
- Sattler, M. (1995) *Internet TV with CU-SeeMe*. Indianapolis: SamsNet.
- Schegloff, E.A. (1979) Identification and recognition in telephone conversation openings. Psathas, G. (ed) *Everyday language. Studies in Ethnomethodology*. New York: Irvington Publishers. 23 – 78.
- Schegloff, E.A. (1981) Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. Tannen, D. (ed) *Analyzing discourse*. Washington: Georgetown University Press.
- Schegloff, E.A. (1984) On some questions and ambiguities in conversation. Heritage, J. & Atkinson, M. (eds) *Structures of social action*. Cambridge: Cambridge University Press. 28 – 52.
- Schegloff, E.A. (1988) Goffman and the analysis of conversation. Drew, P. & Wootton, A. (eds) *Erving Goffman. Exploring the interaction order*. Cambridge: Polity Press. 89 – 135.
- Schegloff, E.A. (1992) Repair after next turn: The last structurally provided defence of intersubjectivity in conversation. *American Journal of Sociology* 97(5), 1295 – 1345.
- Schegloff, E.A. (1995) Discourse as an interactional achievement III: The omnirelevance of action. *Research on Language and Social Interaction* 28(3), 185 – 211.
- Schegloff, E.A., Jefferson G. & Sacks H. (1977) The preference for self-correction in the organization of repair in conversation. *Language* 53, 361 – 382.
- Schegloff, E.A., & Sacks, H. (1973) Opening up closings. *Semiotica* 7, 289 – 327.
- Schiffrin, D. (1990) The principle of intersubjectivity in communication and conversation. *Semiotica* 80(1/2), 121–151.
- Schutz, A. (1980) *The Phenomenology of the Social World*. London: Heinemann Educational Books.
- Searle, J.R. (1969) *Speech acts*. London: Cambridge University Press.

- Sellen, A. (1992) Speech patterns in video-mediated conversations. Bauersfeld, P. Bennett, J & Lynch, G. (eds) *CHI '92 conference proceedings. ACM Conference on human factors in computing systems*. Monterey: ACM, p. 49-59.
- Severinson Eklundh, K. (1986) *Dialogue processes in computer-mediated communication. A study of letters in the COM system*. Linköping Studies in Arts and Science 6.
- Short, J., Williams, E. & Christie, B. (1976) *The social psychology of communication*. London: Wiley.
- Shubert, S.K., Shpyridakis, J.H., Holmback, H.K. & Coney, M.B. (1995) The comprehensibility of simplified English in procedures. *Journal of Technical Writing and Communication* 25/4, 347-349.
- Silverman, D. (1993) *Interpreting qualitative data. Methods for analysing talk, text and interaction*. London: Sage.
- Silverman, D. & Torode, B., (1980) *The material world. Some theories of language and its limits*. London: Routledge & Kegan Paul.
- Silverstein, M. (1976) Shifters, linguistic categories, and cultural description. Basso, K.H. & Selby, H.A. (eds) *Meaning in anthropology*. Albuquerque: University of New Mexico Press. 11 – 55.
- Silverstein, M. (1985) The functional stratification of language and ontogenesis. Wertsch, J.V. (ed) *Culture, communication and cognition*. Cambridge: Cambridge University Press. 205 – 235.
- Silverstein, M. (1992) Indeterminacy of contextualization. Auer, P. & Di Luzio, A. (eds) *The contextualization of language. Pragmatics and Beyond* 22. Amsterdam: John Benjamins. 55 – 76.
- Smith, D.E. (1984) Textually mediated social organization. *International Social Science Journal* 36(1), 59 – 75.
- Smith, D.E. (1990) *The conceptual practices of power: A feminist sociology of knowledge*. Boston: Northeastern University Press.
- Smith, D.E. & Whalen, J. (1995) *Texts in Action*. A paper presented at the Symposium on Conversation, Linguistic Society of America Summer Institute, Albuquerque, July 1995.
- Sorensen, E. K. (1997) *Learning in virtual contexts. Navigation, interaction and collaboration*. Aalborg University, Department of communication.
- Strate, L. (1995) Experiencing cybertime: Computing as an activity and event. *Interpersonal computing and technology: An electronic journal for the 21st century* 3(2), 78 – 91.
- Strate, L. (1997) Cybertime. Strate, L., Jacobson, R. & Gibson, S. (eds) *Communication in cyberspace: Social interaction in an electronic environment*. Cresskill: Hampton Press. 351 – 377.
- Strate, L., Jacobson, R. & Gibson, S. (1997) Surveying the electronic landscape: An introduction to communication and cyberspace. Strate, L., Jacobson, R. & Gibson, S. (eds) *Communication in cyberspace: Social interaction in an electronic environment*. Cresskill: Hampton Press. 1 – 22.
- Streeck, J. (1980) Speech acts in interaction: a critique of Searle. *Discourse Processes* 3, 133 – 154.
- Streeck, J. (1988) The significance of gesture: How it is established. *IPrA papers in Pragmatics* 2(1/2), 60-83
- Streeck, J. (1996) How to do things with things. *Human Studies* 19, 365 – 384.
- Streeck, J. & Hartge, U. (1992) Previews: Gestures at the transition place. Auer, P. & Di Luzio, A. (eds) *The contextualization of language*. Amsterdam: John Benjamins. 135 – 157.
- Suchman, L.A. (1987) *Plans and situated actions. The problem of human machine communication*. New York: Cambridge University Press.

- Suchman, L.A. (1996) Constituting shared workplaces. In Engeström, Y. & Middleton, D. (eds.) *Cognition and communication at work*. Cambridge University Press, New York, p. 35-60.
- Sulkunen, K. (1992) Helsinki University of Technology and videoconferencing. *Euro Conferencing Winter 1992 Edition*, 55–59.
- Suoknuuti, M. (1991) *Videoneuvottelun nykytila. Tutkimusraportti 26*. Lappeenranta teknillinen korkeakoulu. Tietotekniikan osasto.
- Svenning, L.L. & Ruchinskas J.E. (1984) Organizational teleconferencing. Rice, R.E. (ed) *The new media*. Beverly Hills: Sage. 217–248.
- Talbott, S.L. (1995) *The future does not compute: Transcending the machines in our midst*. Sebastopol: O'Reilly & Associates.
- Teklinski, B. (1993) Style analysis of award winning technical manuals. *Journal of Technical Writing and Communication* 22/4, 415–423.
- Thibault, P.J. (1991) *Social semiotics as praxis: Text, social meaning making, and Nabokov's Ada*. Theory and History of Literature 74. Minneapolis: University of Minnesota Press.
- Thibault, P.J. (1997) *Re-reading Saussure: the dynamics of signs in social life*. London: Routledge.
- Tiffin, J. & Rajasingham, L. (1995) *In Search Of The Virtual Class. Education in an Information Society*. London: Routledge.
- Tuan, Y.F. (1978) Space, time, place: A humanistic frame. Carlstein, T., Parkes, D. & Thrift, N. (eds) *Making sense of time*. Bungay: Edward Arnold. 7–16.
- Volosinov, V.N. (1973) *Marxism and the philosophy of language*. New York: Seminar Press.
- Wardhaugh, R. (1976) *The contexts of language*. Rowley: Newbury House Publishers.
- Watson, R. (1997) Ethnomethodology and textual analysis. Silverman, D. (ed) *Qualitative research. Theory, method and practice*. London: Sage. 80–98.
- Werry, C.C. (1996) Linguistic and interactional features of InternetRelay Chat. Herring, S.C. (ed) *Computer-mediated communication. Linguistic, social and cross-cultural perspectives*. Amsterdam: John Benjamins. 47–63.
- Whalen, J. (1995) Expert systems versus systems for experts: Computer-aided dispatch as a support system in real-world environments. Thomas, J. (ed) *The social and interactional dimensions of human-computer interfaces*. New York: Cambridge University Press. 161–183.
- Whalen, J. & Vinkhuyzen, E. (forthcoming) Expert systems in (inter)action: Diagnosing document machine problems over the telephone. Heath, C, Hindmarsh, J. & Luff, P. (eds) *Workplace studies: Recovering work practice and informing systems design*. Cambridge: Cambridge University Press.
- Wiley, N. (1994) *The semiotic self*. Cambridge: Polity Press.
- Winograd, T. & Flores, F. (1986) *Understanding computers and cognition: A new foundation for design*. Norwood: Ablex.
- Witte, S.P. (1992) Context, text, intertext: Toward a constructivist semiotic of writing. *Written Communication* 9/2, 237–308.
- Wooffitt, R., Fraser, N.M, Gilbert, N. & McGlashan, S. (1997) *Humans, computers and Wizards. Analysing human (simulated) computer interaction*. London: Routledge.
- Wortham, S.E.F. (1996) Mapping participant deictics: A technique for discovering speakers' footing. *Journal of Pragmatics* 25(3), 331–348.
- Yates, S.J. (1996) Oral and written linguistic aspects of computer conferencing: A corpus based study. Herring, S.C. (ed) *Computer-mediated communication. Linguistic, social and cross-cultural perspectives*. Amsterdam: John Benjamins. 29–46.
- Zettle, H. (1990) *Sight, sound, motion: applied media aesthetics*. Belmont, CA: Wadsworth Publication Co.

Zimmerman, D.H. (1992) The interactional organization of calls for emergency assistance. Drew, P. & Heritage, J. (eds) *Talk at work: Interaction in institutional settings*. Cambridge: Cambridge University Press. 418–469.

Appendix 1-1

TRANSCRIPTION CONVENTIONS

General:

lower case (bold)	what was actually said
UPPER CASE	loud voice
<u>x</u>	stressed (part of) word
°word°	word delivered quieter than the surrounding talk
.word	word produced with an inbreath
>word<	speech item delivered quicker than other talk
not in bold	translation
---	part of transcript omitted
!	exclaiming tone of voice
.	falling intonation
,	flat intonation
/	strong fall-rise in intonation (teacher-like)
(N)	length of pause in seconds
(.)	pause shorter than one second
=	talk/action latches on another
''	quoting the screen
th:	hissing sound
()	analyst not sure what was said
(())	an activity or comment on the delivery of speech
((;))	simultaneous activities
[]	
[]	simultaneous speech/activity
	simultaneous speech/activity by two persons
<i>italics</i>	what is typed on the screen
▬	(pressing) the spacebar key
↓↑	mouse click
↵	pressing the 'enter' key
Abbreviations:	
kb	keyboard
sb	spacebar
twds	towards
btw	between

Telephone answering system (TELEPHONE):

<WORD >	text in the system's response not read aloud
<< word >>	what the user said but was not typed in
>>WORD<<	what was typed in but what the user never said

Learning Word 5.0 (TUTORIAL):

≡	screen black for 1 second
<i>italics in a box</i>	computer's (C) turn
<i>italics in a double line box</i>	C's turn appearing at the bottom of the screen
■	C's turn disappears from the screen
v	new screen (vs. minor changes)

Mailing labels task (TASK):

" "	quoting the manual
-----	--------------------

Video conference (TEACHING):

F-	speaker(s) in Finland
S-	speaker(s) in Sweden
⚡	a cut in the audio connection
<i>italics underlined</i>	typing left on the screen (of a contribution)

Appendix 4-1

Instruction sheet

This experiment is being carried out to test a hypothetical telephone message storing and sending system: you can call the system which is attached to your phone to send messages or check if there are messages for you from other people's systems.

As there is no speech recognition/production system attached to the program, I will be simulating it on the phone.

Now you are calling your telephone and want to do the following:

- I** Find out if Fred has sent any messages
- II** Find out if Eve has sent any messages
- III** You want to get a message through to Eve, saying that you are at home
- IV** You want Fred to know that you are going to a meeting at 7 pm

The system does not understand the pronouns like me, myself, you, he etc. so you should try and avoid using them!

It is possible for you to indicate mishearing and to some extent misunderstanding.

We assume that the system recognizes your voice and behaves accordingly (ie. no password or other type of checking is needed).

Dial Southbridge/ext. 232 and call up your system to start the session.

Thank you for participating!

Appendix 6-1

INSTRUCTIONS:

Please do the following task together.

TASK:

You are both working as student trainees in the English Department. You have been asked to send a letter to the department's four new postgraduate students, whose names and addresses are given in Appendix 1. Now you need to print mailing labels for the letters. In future the addresses of all postgraduate students will be transferred to the Word For Windows program, so it is useful to prepare the printing of the mailing labels by using this particular program.

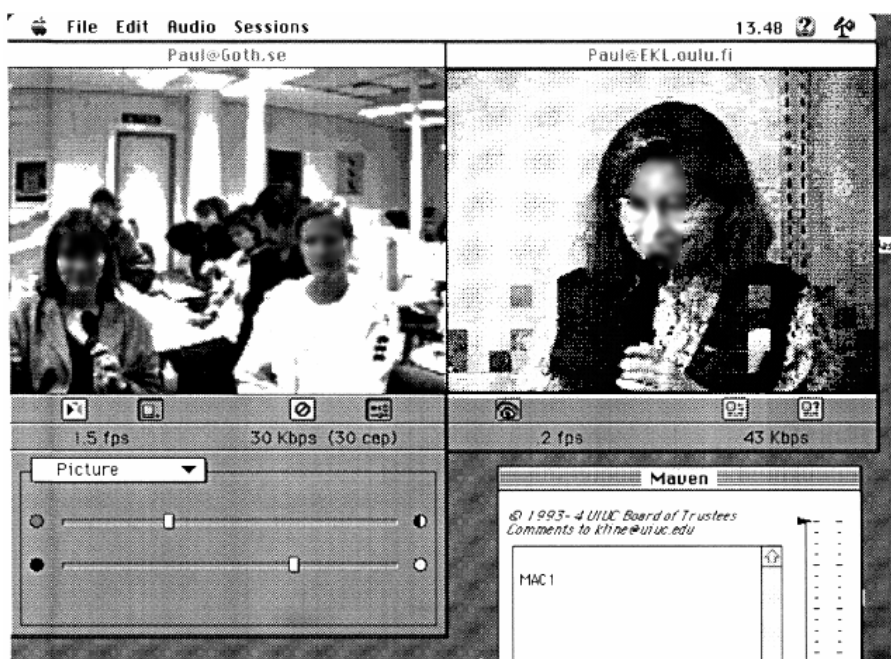
Create a setup for printing the labels into which all the postgraduate students of the department can be added later on. You can find the instructions for printing mailing labels in the User's Guide, p. 657 onwards. The labels on which the addresses will be printed come in sheets of 3 x 7 labels, 63.5 mm x 38.1 mm. The new labels should be similar to the labels printed earlier (Appendix 2). To complete the task you only need to print the labels on a plain sheet of paper. In other words, stick-on label sheets will not be used during the task.

NOTE:

You can use no more than one hour and 20 minutes (80 mins.) for the task.

Good luck!

Appendix 7-1



A snapshot from a computer screen in Sweden.

Appendix 7-2



Two double sized video pictures on a 14 inch monitor.

Appendix 7-3



A snapshot from the third video recording in Finland

ISBN 951-42-5248-9

Åbo Akademis tryckeri
Åbo 1999