The pragmatics of monologue: interaction in video blogs

Table of contents

Acknowledgements

I am deeply grateful to many people who have helped me, in one way or another, in the process of writing this study. It is to Neal Norrick that I would like to express my heartfelt gratitude, not only for his support and implicit trust in me as a scholar, but also for sharing a penchant for wordplay and etymology, and generally enjoying language in use as a practical endeavor, including a sometimes quirky sense of humor.

Furthermore, I would like to thank Elke Teich for agreeing to co-supervise this project. The insightful comments I received from her and many participants of her colloquium, most notably Sabine Bartsch, have enriched my work in various ways. I would also like to thank Wolfram Bublitz for kindly agreeing to act as external reader.

My mentor Ruth Page has been a constant source of inspiration and input on many levels. For reading and commenting on parts of this study, for inviting me into her home and giving me space and time to discuss my ideas, and for initiating and working with me on a very fruitful research project, I extend my warmest gratitude.

My work with Richard Harper has been a truly empowering experience: His generosity and his interest in advancing my skills opened up a whole new perspective on scholarly work for me – for which I am deeply grateful.

Many friends and colleagues at Saarland University were involved in this project and mastering the challenges surrounding it: Nele Gerhardt, Carrie Ankerstein, Marie-Louise Brunner, Matthias Heyne, Martina Ghosh-Schellhorn, Astrid Fellner, Stefan Diemer and Henry Rademacher, to name but a few, have signaled their interest and invested time and work in this endeavor, have generously supported me, and have helped me disentangle some knots. To all of you: thank you.

My family and friends through their unfailing belief in me, and the YouTube community through exploring new ways to communicate, paved the way for this study from the beginning to the end.

The articles listed below have been reproduced with kind permission.

Frobenius, Maximiliane. 2011. Beginning a monologue: The opening sequence of video blogs. *Journal of Pragmatics* 43 (3). 814-827. Reprinted with permission from Elsevier.

Frobenius, Maximiliane. 2013. Pointing gestures in video blogs. *Text&Talk* 33(1). 1-23. Reprinted with permission from De Gruyter.

Frobenius, Maximiliane. 2014. Audience design in monologues: How vloggers involve their viewers. *Journal of Pragmatics*, DOI: 10.1016/j.pragma.2014.02.008. available online April 12, 2014. Reprinted with permission from Elsevier.

1 Introduction

1.1 General introduction

This study is an in-depth investigation of monologues as they occur in the relatively young genre video blog or vlog. It is situated in the perspective of pragmatics (Verschueren 1995), and as such it represents a study of language in use, or language in context. The aim of this investigation is to demonstrate that speakers employ various means, or affordances, which render their monologues audience-directed acts of communication that inspire or reflect interaction.

The contributions to our understanding of human communication proposed in this study are essential in at least two major areas. For one, it advances the study of computer-mediated communication (CMC) by providing a first comprehensive description of the genre in the field of language studies. For another, it explores monologue as a mode of speaking which has received surprisingly little academic attention. Thus, the present investigation lays the foundations on which future work along two dimensions can be based – that of genre specific explorations in CMC, and that of the investigation of monologue as a particular speech type.

The relevance of doing linguistic research on video blogs can be attributed to various factors. First and foremost, any type of speech situation that humans engage in should be considered a potential object of study from the viewpoint of a discipline striving to develop a comprehensive theory of language use in general. While vlogs represent a marginal speech situation in terms of how many people in a given society regularly engage in them compared to phone calls or face-to-face interactions for example, they do represent an established, if not an essential means to communicate for some. One can assume that these people, vloggers, establish and shape the genre by adopting certain practices, thereby creating and maintaining standards and norms. A systematic investigation of these practices thus contributes to linguistic pragmatics in general.

More specifically, the linguistic study of computer-mediated communication, due to the tender age of its object of study, is an emergent discipline (Herring et al. 2013). Unlike

more established fields that investigate long-standing genres such as conversation, CMC is facing constant change: New genres are developed, with ever-changing designs, using and combining various modes that offer a multitude of affordances. At the same time, new devices allow internet users to become increasingly mobile while maintaining their level of online access (and accessibility). Clearly, research will always lag behind these rapid developments, making it particularly important to provide genre descriptions not just as a basis for further research on that same genre but, more generally, for a better understanding of the dynamics of the different types of online communication in terms of, for example, their (economic) success, longevity, or functions in the day-to-day lives of their users.

As hinted at above, some CMC genres combine various modes. Vlogs are a fairly complex example in that they originate in a setting that is similar to face-to-face conversation except for the missing interlocutor, which means that, in addition to spoken language, there are gestures, gaze, body posture, and intonation that make the text multimodal per se. The finished product, however, is placed in a two-dimensional website space, which is dominated by written language, using different fonts, colors, sizes and hyperlinks, as well as other website design features. Thus, there is a two-level multimodality inherent in the genre which represents a challenge to users in that they are faced with a multitude of choices they need to navigate, and it represents a challenge to researchers who have to trace a user's actions to understand these choices. Such a demanding task is highly relevant to our understanding of multimodality, especially with regard to such a central concept as context.

In terms of the micro-organization of interaction among several online participants, the present research contributes pioneering theoretical insights precisely because of the aforementioned multimodality. Users are challenged to create coherence across several modes, forcing us to reconsider and expand our traditional models of tying (Sacks 1995) via adjacency. Therefore, the relevance of this investigation is also instantiated in its adaptation of traditional research methods to new communicative settings, thereby

implicitly demonstrating the inadequacies of such approaches and initiating the development of more appropriate ones.

The terminology surrounding modes of speaking with regard to the number of speakers and the concepts these terms denote respectively, that is *monologue*, *soliloquy*, and *dialogue* are by no means well-defined and established. Two central strands of thought exist that represent the relationship between monologue and dialogue as follows: While the two modes can be conceived of as discrete entities, where a spoken stretch of talk is either a monologue or a dialogue, depending on how many participants are present, another approach views this relationship as a graded cline, where certain features make language progressively more monologic or more dialogic. Soliloquy has been linguistically defined as "the utterance of one's thoughts without addressing another individual" (Hasegawa 2010: 2). Literary research, especially of drama texts, makes that very same distinction between monologue and soliloquy (on stage). Here, monologue is addressed to another character, while soliloquy is not (see Pfister 1977; section on research on monologue below).

I consider it important at this early point to make salient my own stance on this specific issue of terminology: For analytic purposes I prefer to think of monologues as spoken language "where the audience isn't expected to interrupt" (Clark 1996: 5), thus I tend towards the first view of monologue and dialogue as discrete entities. In order to use the graded cline view as a helpful tool, one would have to know what precisely monologic and dialogic features are – surely they would have to be defined on the basis of their frequency of occurrence in either speech situation, making this definition a somewhat circular one. While the present thesis does not attempt to come to a conclusion as to which view is more adequate, it does contribute to solving this problem in that it provides an inventory of features that occur in monologues according to the first, setting-based definition. This could be used as a springboard for further, more extensive and systematic comparisons of other monologic and dialogic settings.

Clearly, however, the description of a genre that is based only on data from that genre runs the risk of lacking a point of comparison that allows judgment of a feature in terms

of its status as a typical or atypical feature in the context of that genre. While this study is methodologically set to work from a micro-analytic context-sensitive (e.g. Interactional Sociolinguistic) perspective, and thus is not able to provide a large scale corpus-based comparison, there are two major strands that make this study a comparative description of video blogs: (1) The previous research reviewed and applied here has, to a large extent, been developed from and for (face-to-face) conversation or other synchronous modes of interaction. Applying these findings to monologues, and pointing out the adaptations needed to do so, necessarily results in an implicit comparison of monologue and dialogue. (2) Therefore, chapter 8 will be devoted to an explicit comparison of vlogs to another monologic genre, namely the TED Talk. Selected features of vlogs and lectures will be mirrored in terms of their forms and functions.

The approach I have chosen in this treatment of an emergent web genre places the detailed analysis of data firmly at the center of attention, in other words, this study is data-driven. Consequently, this text is replete with examples taken from vlogs. Most of them appear in the form of transcribed speech, while some are represented in their visual form with the help of screen shots. A detailed description of the data corpus and a general characterization of the genre – as far as this is possible in light of the scarcity of previous research – follow below.

A note on the writing process of this thesis is in order. Five of the seven analytic chapters that form the main body of this study were originally written as stand-alone research papers in a format that would allow them to be published individually in journals. As journals impose (sometimes strict) limits on the length of an article, the chapters do not significantly exceed the 8000 words mark. Therefore, the lengthy discussion of various examples to illustrate one particular point often had to be abandoned in favor of a concise presentation and analysis – a feature which might strike a reader as somewhat uncommon for the text type dissertation.

The remainder of this introduction is organized as follows: A review of the relevant fields of research and methodologies as laid out by the main proponents in their significant contributions situates this research as emerging from and belonging to

particular research traditions. The setting of the genre vlog, that is, the website YouTube, is described by way of a review of the literature so far amassed, including a specific genre description. Finally, the data corpus is introduced, alongside a note on transcription conventions and the ethics of using publically available internet data for research.

1.2 Methodology, Literature

1.2.1 The pragmatics of CMC

A study written from the perspective of pragmatics needs a more precise delimitation of the methods employed, especially since the boundaries of the set of methods denoted by that term pragmatics are fuzzy at best. In fact, pragmatics has been called "a perspective on language, rather than a component of a linguistic theory" (Verschueren 1995: 12), which is "*an approach to language which takes into account the full complexity of its **cognitive, social, and cultural (i.e. 'meaningful') functioning** in the lives of human beings*" (13f) [italics and bold print in the original]. Clearly, a more detailed description of the approach chosen is in order, which will be the objective of this section. This must be read as a basic introduction to the relevant fields, however. A more detailed description of the models used in the analysis and how they are applied will be provided in the respective chapters.

In their introduction to *Context and Contexts* (Fetzer and Oishi 2011: 1), the authors give a more concrete description of the field of pragmatics: "Pragmatics is fundamentally concerned with communicative action and its felicity in context". They continue: "The heterogeneous nature of context and the context-dependence of the concept itself have made it almost impossible for the scientific community to agree upon a commonly shared definition or theoretical perspective" (Fetzer and Oishi 2011: 1). This further reflects the nature of pragmatics as a wide field, whose foci can be fairly varied. Consequently, the crucial notion of context has been approached in various ways (see, for example, van Dijk 2008 or Duranti and Goodwin 1992 for discussions of

conceptualizations of context in the humanities). In the following, I will briefly outline a pragmatic approach developed for the study of online communication.

Herring et al. (2013) evokes Ariel's (2010) distinction between "border-seekers" and "problem-solvers" among pragmatics scholars, where the former are predominantly concerned with a "particular set of topics that are defined as pragmatic, rather than syntactic or semantic" (Herring et al. 2013: 6). The latter, on the other hand, focus on a particular problem that cannot be tackled "in terms of grammar, or 'the grammar' in a Chomskyan sense" (Herring et al. 2013: 6). Herring continues by applying this distinction to the *pragmatics of CMC* in particular, explaining that contemporaneous interest in the field reflects a stronger tendency towards the problem-solving approach. This approach, she continues, "includes the socio-technical constellations of the uses of language in CMC – the communication pragmatic constraints as captured in concepts such as 'mode' and 'affordances'" (Herring et al. 2013: 7).

Herring et al. (2013: 4), developing a research agenda for the pragmatics of CMC, recommends focusing on "three kinds of phenomena: 1) classical core pragmatic phenomena (e.g. implicature, presupposition, relevance, speech acts, politeness) in CMC; 2) CMC-specific phenomena (e.g. emoticons, nicknames, "Netspeak"); and 3) CMC genres or modes (blogs, SMS, wikis, chat, etc.)." The present investigation falls mainly into the third category (which, of course, does not preclude a concern with topics from the other two categories where applicable). Herring's own earlier work includes the development of a classification scheme for computer-mediated discourse. It aims at synthesizing and articulating aspects of technical and social context that influence discourse use in CMC environments (Herring 2007: 1). This scheme provides a scaffold on which to develop a description of mainly written CMC genres, and it has subsequently been revised and amended (see e.g. Puschmann 2010, Herring forthcoming). A pragmatic approach to providing a taxonomy of internet communication based on relevance theory (Sperber and Wilson 1986) is what its inventor terms cyberpragmatics (Yus 2011).

The term *genre* and its use, then, requires more intensive discussion as within the scope of the present study. As is explained in more detail below, I try to avoid deciding on what counts as a vlog based on other than the vloggers' own judgments. I follow this approach because at this point in time there does not seem to be an established definition, and, consequently, a data selection based on my own judgment would run the risk of being biased regarding my research agenda. Furthermore, as this study will show, there do not seem to be any fully conventionalized, that is obligatory, communicative strategies that make a video a vlog – at this point we can identify tendencies. Therefore, if we can speak of a genre according to more traditional linguistic understandings of the term (see below), it is in its formative stage and consequently difficult to grasp. Still, I do use the term *genre* with regard to vlogs, and beacause one of the main goals of this study is a genre description, there should be a note on my use of this terminology.

Giltrow and Stein (2009: 1), in their introduction to *Genres in the Internet*, an edited volume predominantly dedicated to genre theoretical issues, state that "[a]t the very least, *genre* works as a common intuitive concept – a sense that features of language aggregate in recognizable patterns, and that these aggregations indicate something important in the uses of language in context" [emphasis in the original]. The above quote by Herring, "… genres or modes …", (2013: 4) is an instance of the use of the word *genre* alongside the term *mode*, followed by exemplifications of those categories, without, however, explicitly distinguishing them. She then reports more specifically: "[The] drift in the theoretical orientation of the definition of genre recapitulates the drift in the study of language on the internet from a surface, form-based approach to a more broadly pragmatic approach to an emphasis on users as social identity carriers in dynamic, ad hoc, cognitive states" (Herring et al. 2013: 9). This observation is derived from Giltrow (2013: 731), who states that

> results of pragmatic study challenge the notion that genre is what is the same, what is formally fixed or even rigid, and what is rule-governed, at best instrumental, at worst imposing and exacting of conformity. With its capacity to

expose the coordinates of language-users' social positions and inclinations, pragmatic study can establish genre as what is versatile and variable, functional because flexible, and subjective rather than instrumental.

These statements reflect two currents in genre theoretical studies, a form based and a function based approach (cf. discussion of Discourse Analytic approaches below). Baumann (2006) explains that while schools such as systemic functional linguistics and applied linguistics (specifically English for special purposes) "conceive of genre as a nexus of the linguistic and the social, but with primacy accorded to the former" (750), social constructivism views genres as instantiations of "the solution of recurrent communicative problems" (750).

The present investigation seeks to give an initial impression of the linguistic structure and strategies employed in vlogs, and, where possible, their purposes. It is not, however, designed to provide a systematic genre analysis in terms of purposive, goal-oriented stages in the tradition of systemic functional linguistics or an analysis in terms of moves akin to practices in applied linguistics. This would entail a text-based analysis of lexico-grammatical patterns of cohesion marking the boundaries of the generic stages (cf Paltridge 1994; Hasan 1978; Richards, Platt, and Platt 1992: 62). Nor does this study pursue an approach that foregrounds the purpose of vlogs or stages within them. As vlogs are a multi-modal form of interaction, encompassing a virtual and a physical setting, two stages of production, and the potential for asynchronous exchanges with the viewers, an approach that tries to apply the notion of cohesion to this kind of data would result in a reductionist study. However, the meaning-making potential inherent in, for example, the gestures, gaze shifts, and further visual material employed in vlogs investigated in this study would be difficult to analyze in terms of cohesion. Regarding a purpose-driven investigation of vlogs, Askehave and Nielsen (2005: 122) explain that if a researcher is not part of the relevant discourse community, he or she must turn to its expert members to find out about a text's purpose. Ethnographic research has revealed that vloggers make and upload their videos to maintain existing social networks, or even create new ones (Lange 2007b). This insight is clearly not specific enough to match it to

linguistic practices in vlogs. Vlogs appear to be too recent, or maybe too marginal, a phenomenon for a more specific investigation of its purpose to have been developed at this point.

Thus, the use of the term *genre* in the present study is not meant to invoke any genre analysis tradition along these aforementioned parameters. It is, however, based on the vlog community's mutual understanding of the term vlog, that is, it shifts the basis of genre distinctions from an outside observer's analytic position to the insider users' perspective. Baumann (2006: 746) describes a similar reasoning in the works of Boas (1940) and Malinowski (1948), who investigate narrative genres of North American peoples and Trobriand Islanders respectively. The two scholars take an ethnographic approach in that they use "indigenous distinctions, expressed in native terms" (Baumann 2006: 747), essentially adopting an emic, rather than etic perspective.

When taking stock of the terminology used by vloggers to label their videos, it becomes obvious that the community of vloggers distinguishes different kinds of vlogs. Thus, a vlog can for example be a *haul*, or an *unboxing*, and a haul, in turn, can be further classified into *plus-size hauls*, *discount store hauls*, and so forth (see section 1.3.4). This demonstrates that there is a genre ecology present or under development that the members of the vlog community have recourse to and shape according to their needs. These more specifically topic bound vlogs may, at least if one foregrounds the parameter of purpose, be more suitable to fit a more traditional understanding of the concept of genre. Here, vloggers are possibly more restricted in the semantic and functional range, and possibly also in the linguistic strategies available to them (see section 1.3.4 and Jeffries 2011).

In light of the various approaches to the concept and the nature of the object of this study, I will proceed to use the term *genre* with regard to vlogs to mean what vloggers (and, where possible, viewers) understand to be a vlog. In other words, I treat *vlog* as an emergent category. This includes a cognitive state that comprises mutual knowledge and recognition, a basis on which interactants instruct each other as to how to make sense of each other's actions:

Classification of these types shows not simply that people develop functional ways of speaking under newly enabled conditions, but that they construct mutual knowledge of these ways: They not only know and recognise a way of speaking but also know that others know this speechway" (Giltrow 2013: 720).

1.2.2 Multimodality in CMC

The discussion of CMC genres, how to theorize them, and ultimately how to arrive at a fruitful analysis of them, has a major focal point in its inclusion of multimodality as a basic assumption about all communication online. In their characterization of linguistic CMC genre description, Thurlow and Mroczek (2011: xxxif) observe a) that it is a "favorite occupation" of researchers in that field, and b) that it no longer concludes by making statements about a genre's hybridity, but takes it as a starting point to attend to matters such as the inherent multimodality. In their introduction to *Digital Discourse*, they state that "multimodality is – or at least should be – a 'taken-for-granted' in new media studies" (xxv). And further:

> In their efforts to redress the relative absence of the linguistic in computer-mediated communication, scholars often overlook the fact that it is only ever one of many communicative resources being used. All texts, all communicative events, are always achieved by *multiple semiotic resources*, even so-called text-based new media like instant- and text messaging [emphasis in the original] (Thurlow and Mroczek 2011: xxvi).

As an approach to the study of communication, multimodality presupposes that multiple semiotic resources are at play simultaneously in any communicative event (see Page 2010: 4). In other words, making meaning draws on multiple modes, be it in a written text or spoken communication (Jones and Hafner 2012: 50). Based on this premise, one can argue that all text (in the comprehensive sense of the word) is multimodal. Words, then, are just one resource among many. The phrase "modal democracy" (Page 2010: 4) denotes this integration of various modes, while it does not

imply that all modes are equally prominent at all times. On the contrary, multimodality recognizes the dominance of, for example, verbal resources in communicative acts, but tries to steer away from approaches that make this resource their sole focus. It does so by drawing connections between the different resources (Page 2010), which is reflected in Kress and van Leeuwen's characterization of multimodality as "common semiotic principles [that] operate in and across different modes" (2001: 2).

A semiotic *mode* can be understood as "a system of choices used to communicate meaning" (Page 2010: 6). These systems encompass "aural and visual, as well as verbal and textual" modes (Jones and Hafner 2012: 50), that is spoken and written language, prosody, gesture, gaze, image, color, font and so forth. From an interactional perspective, Goodwin's approach to multimodal analysis is central. His discussion of visual semiotic systems in talk-in-interaction makes use of the term *field*, rather than *mode*:

> Images in interaction are lodged within endogenous activity systems constituted through the ongoing, changing deployment of multiple semiotic fields which mutually elaborate each other. The term semiotic field is intended to focus on signs-in-their-media, i.e., the way in which what is typically being attended to are sign phenomena of various types (gestures, maps, displays of bodily orientation, etc.) which have variable structural properties that arise in part from the different kinds of materials used to make them visible (e.g. the body, talk, documents, etc.). Bringing signs lodged within different fields into a relationship of mutual elaboration produces locally relevant meaning and action that could not be accomplished by one sign system alone (Goodwin 2000b: 166).

The ensemble of semiotic fields which makes up the contextual configuration is a basis for my analysis of multimodal interaction in vlogs. With regard to video data, Jones and Hafner (2012: 63) state that they "mix the affordances of image with the affordances of writing". The written mode, according to them, follows a sequential order, whereas images – the visual mode – are organized in a spatial/simultaneous manner (65).

Beißwenger (2008) convincingly argues that, for written CMC modes, researchers must distinguish two types of arenas of investigation. In his discussion of transcription methods that adequately capture all relevant information, he points out that one quite reduced record of interaction of Internet Relay Chat (IRC) is the written text exchange itself, while a comprehensive record includes chatters' gaze behavior, their positioning towards the computer screen, interactions in their immediate surroundings etc. In the first case, some might claim to be dealing with a monomodal type of interaction, although such elements as emoticons, the use of different fonts or colors, or simply capitalization could be argued to add another visual mode. However, it should be clear that all chatting is done in some physical location with specific resources, which adds layers of interaction with varying levels of impact on the chat itself. Stripping them necessarily results in a reductionist approach. Regarding research practices, however, it must be pointed out that access to these additional layers is, of course, extremely limited, and only to be gained through considerable effort, and then possibly only in laboratory settings rather than the natural homes of these interactions. Furthermore, as Beißwenger (2008) documents, the complex texture of interaction, if one wants to analyze it in terms of its multimodal character, poses challenges regarding the representation of the interaction in a written transcript. As Page (2010: 8) explains:

> Transcription must be at once systematic and replicable, but flexible enough to embrace the rich diversity of all that multimodality encompasses. This is no mean feat, and a definitive, established system of multimodal transcription is far from complete in the way that systems for representing single modes have developed (say, for language, prosody, music).

My own approach to data transcription is covered in section 1.5.

1.2.3 Conversation Analysis and Interactional Sociolinguistics

Discourse Analysis, like pragmatics, is a term that can be taken to cover a rather wide range of approaches. A useful distinction is that of formal (or structural) approaches to

discourse versus functional ones (Schiffrin 1994; Hymes 1974; Leech 1983). Where the former ones "focus on the way different units function in relation *to each other*" (Schiffrin 1994: 23) [emphasis in the original], the latter ones approach discourse as "interdependent with social life, so that its analysis necessarily intersects with meanings, activities, and systems outside of itself" (Schiffrin 1994: 31). While this distinction of types of approaches, and the resulting distinction of definitions of discourse, holds theoretically, "actual analyses of discourse reveal an interdependence between structure and function" (Schiffrin 1994: 42). My own approach, too, is grounded in both paradigms, reflecting the necessity to account for both form and function equally. It does so by borrowing from Conversation Analysis (CA), a structural approach based on the belief that context is created through interaction, on the one hand, and from Interactional Sociolinguistics (IS), a functional approach seeking to explain language use in terms of its indexicality to "background cultural understandings" (Schiffrin 1994: 105). The two fields are introduced in this section.

Conversation Analysis, originally developed by Harvey Sacks, Emanuel Schegloff and Gail Jefferson, is rooted in ethnomethodology. Its aim is to describe human action towards the creation of social order as displayed in conversation (Garfinkel and Sacks 1970). Its main methodological tool is the sequential analysis of turn-taking, whereby a turn at talk contains information both about the previous turn, as well as about how the next turn could be appropriately constructed (Sacks et al. 1974). In this sense, contributions are both shaped by the local sequential context and at the same time renew this context. It is presupposed that speakers encode and decode this interactional information in their turns to help each other 'make sense' of their contributions, and thus enact the joint production of coherent actions through time.

The methodology of CA must be understood in terms of its reliance on what is a participant's understanding rather than that of an analyst. Hence, what is of analytical interest is the orientation of interlocutors towards each other's contributions. In this way, we can observe the methods people apply to make sense of each other. As a consequence, contextual factors such as age, gender, social status etc. can only be taken

into account if they are demonstrably oriented to by the participants of the conversations, or in other words, if they are manifest in the turn-taking practices displayed in the data (Schegloff 1992). Or, as Rawls (2004: 4-5) puts it: "The point is to find these constraints at work in interaction without performing a conceptual reduction or starting with assumptions about macrostructures".

This rather restrictive and strictly empirical approach to data analysis is, indeed, well suited to performing a micro-analysis of the 'members' methods' and can thereby help to uncover the ways people structure and organize their contributions in a local context to achieve orderly interaction. CA has, however, come under criticism for not providing a very good theoretical basis to link the micro-structures to the macro-context if it is not demonstrably present in the sequential unfolding of participants' contributions. In an article that discusses, among other things, precisely the problem of evidence that arises from treating macro-variables, such as interactional history, ethnic background, speakers' relationship etc., as constituents of a communicative situation, Gumperz (2003: 115) states:

> Since situated understanding ultimately always builds on indexical signs, we cannot avoid ambiguity. But then, […] even properly socialized natives can understand different things by the same utterances. In fact, we all know political discourse, advertising, committee negotiations and similar genres rely on ambiguities for rhetorical effect (Cook-Gumperz and Gumperz 1994). Interpretation, moreover, is not a unitary assessment. Typically, inferences are interactively negotiated over several turns at speaking and such negotiations rarely yield uncontestable understandings. As I see it, the IS analyst's task is to reconstruct the understanding process, relying on a varied array of formally distinct sources of knowledge both at the level of content and form.

Gumperz thus takes a firm stance against the sole focus on sequential micro-structures advocated by CA. He propagates the inclusion of whatever (certified) background knowledge participants bring to a communicative event available to the analyst, even if it does not present itself as an emergent category on a micro-level. Only in this way can

an analyst actually trace the reasoning behind how participants in a conversation make sense of each other's contributions, as macro-level elements, such as interactional history, do not necessarily have an impact on the sequential order of turns, while still being essential to the participants' understanding of the communication.

Ten Have (no year [1990]: no pages), discussing criticism voiced by Garfinkel in an interview (Jules-Rosette 1985), points out the limitations of the CA approach in terms of its failure to conceptualize interaction as situated practice:

> [T]he strategy taken by CA, while starting from the same problematic as other kinds of ethnomethodology, is, by its reliance on a reification of its object through the 'overhearing' of tapes and the construction of transcripts, restricted in its study of conversational streams as situated practices. Instead, a sequential/structural representation is studied from a position of 'anybody'.

Further criticism of the CA paradigm is based on the potential alternative organizations of orderly human interactional behavior that cannot be described in terms of turn-taking. The perceived ideal of conversation "no gap, no overlap" or "one party at a time" (see, e.g., Schegloff 1968) does not necessarily hold for all settings, conversational or otherwise. Ceremonies that involve chanting, joint citations or other ritual texts are a point in case; likewise conflict talk (see Goffman 1981, Gumperz 1982, Kotthoff 1994). While CA purists are largely aware of this and consequently limit the applicability of their findings to particular genres, there clearly is a lack of a model in CA that can handle a more general notion of communication.

An approach that can provide just that is the Ethnography of Communication, which is mainly informed by the works of Dell Hymes (Hymes 1974). One of the basic theoretical models he implements is the three-way distinction between *speech situation*, *speech event*, and *speech act*, which all exist or are enacted within a *speech community* (Hymes 1974). One of the characteristics of a speech community is the dimension of intelligibility, made up of both "knowledge of at least one form of speech, and knowledge also of its patterns of use" (Hymes 1974: 51). In this sense, participating in a speech community requires not just grammatical knowledge but also knowledge of

"cultural appropriateness" (49). The notion of speech situation denotes general settings such as ceremonies, fights, or hunts. While they are not themselves governed by rules of speaking, they can "enter as contexts into the statement of rules of speaking as aspects of setting (or of genre)" (51). Speech events, on the other hand, are "directly governed by the rules or norms for the use of speech" (52). Thus, this model incorporates macro-level information that speakers necessarily have to participate successfully in communication and therefore makes contextual knowledge used by speakers accessible to the analyst.

Interactional Sociolinguistics, as propagated by John Gumperz, tries to integrate both macro- and micro-structures:

> The key to Gumperz's sociolinguistics of interpersonal communication is a view of language as a socially and culturally constructed symbol system that is used in ways that reflect macro-level social meanings (e.g. group identity, status differences) and create micro-level social meanings (what one is saying and doing at a moment in time) (Schiffrin 1994: 102).

This is achieved with the help of a model that encompasses three interrelated concepts (Gumperz 1982): contextualization cue, contextual presupposition, and situated inference.

> A contextualization cue is any feature of linguistic form that contributes to the signalling of contextual presuppositions. Such cues may have a number of such linguistic realizations depending on the historically given linguistic repertoire of the participants. The code, dialect and style switching processes, some of the prosodic phenomena we have discussed as well as choice among lexical and syntactic options, formulaic expressions, conversational openings, closings, and sequencing strategies can all have similar contextualizing functions (Gumperz 1982: 131).

These contextualization cues relate speakers' verbal and non-verbal behavior to their contextual knowledge of the activity they are engaged in. In other words, they represent

the link between a cognitive dimension and its realizations in interaction. Based on contextualization cues, and with the help of people's cognitive capacities, they can make situated inferences, i.e. "interpret what is going on" (Gumperz 1982: 3), which, at the same time, requires what he terms *conversational involvement*.

Goffman's (1981) notions of participation framework and footing complement Gumperz's basic model. Footing refers to "the alignments we take up to ourselves and the others present as expressed in the way we manage the production or reception of an utterance" (Goffman 1981: 128). Shifts in footing are achieved, among other things, by what Gumperz terms contextualization cues. Furthermore, Goffman (1974 (1986)) developed the concept of interpretive frame, which can also be indicated with the help of contextualization cues and thus represents a theoretical basis for contextual presupposition and situated inference (Schiffrin 1994: 103).

1.2.4 Monologue

Linguistic studies of monologue per se are few and far between so that we can hardly speak of a field *monologue studies*. The general distinction already mentioned between a situational definition of monologue and a structural one, which result in either a binary or gradual classification of the status of discourse as monologue, is rarely discussed. The following review of the literature remarks on this where applicable, as there seems to be little awareness of this issue and its theoretical implications. Many studies use monologic data without making this an explicit point of interest (e.g. Schwarz 2010, Fortanet 2004, Alajoutsijärvi and Määttä 2011, Kazandjian et al. 2007, Strodt-Lopez 1991). This section gives an overview of the research that has so far been published on monologic speech.

An important, very basic assumption that I consider essential to the present study has been brought forward by several scholars in various contexts. The assumption is that monologues, and humans' ability to produce them, are based on conversational genres. Haviland (2007: 150) writes "interaction is a compelling model for talk, even apparently

monologic talk". Similarly, Schegloff (1987: 222) claims that speech exchange systems, and their turn-taking organizations, are the product of transformations or modifications of the one for conversation, which is the primordial organization for talk-in-interaction. Swales (1990) calls conversation a pre-genre, thus it is ubiquitous to such a degree that it informs many genres, but does not itself have this status. He does not, however, consider conversation as a precursor to monologue. Instead, he suggests narration as an obvious candidate to be a pre-genre of monologue (Swales 1990: 61). I cannot quite agree with this view in that conversation can be called the "natural home of narrative" (Norrick 2002: 258, 2007: 127). Thus, ultimately, even if it is via the detour of narrative, it would appear to go back to conversational settings.

Within the CA paradigm, one finds a section on applying methods of sequential analysis to monologic speech in Hutchby and Wooffitt (1998: 185-201). Without giving an actual definition of monologue, they implicitly characterize it as "long uninterrupted stretches of talk" (185). Two interrelated problems arise: the default unit of analysis in CA is the turn, which, by (situational) definition, does not occur in monologues. Consequently, a crucial means of providing evidence for the existence of analytic categories, namely interlocutor's reactions in their own next contributions, is missing. Hutchby and Wooffitt (1998) suggest, through their analysis of monologic passages, that self-repair and insertions are signs of audience design, or even anticipation of potential criticism by the interlocutor, and can thus be used analytically instead of another interlocutor's reaction.

Without an explicit definition of monologue, Coulthard and Montgomery (1981) analyze the structure of monologues in a discourse analytic study of academic lectures. Following a caveat about the preliminary and underdeveloped nature of their model, they suggest the notion *member* as a unit of lectures, which make up either *main discourse* or *subsidiary discourse*. It is that latter one, consisting mainly of asides and glosses, they speculate, which makes freely produced lectures easier to follow than read-aloud papers. Likewise, the presence of such subsidiary discourse is what makes a lecture interactive rather than "an undifferentiated chain of informing members" (Coulthard and Montgomery 1981: 39).

The genre lecture monologue has been investigated further, particularly in terms of its status as an institutionalized device to convey knowledge to learners. Therefore, many of these studies have an EFL/ ESP background. Crawford Camiciottoli (2008) investigates the use of questions in instructional settings, Fortanet (2004) the use of the pronoun 'we' in academic speech, and Jung (2005) the role of markers in L2 listening comprehension. Further topics are: genre-defined functions and multifunctionality of pragmatic force modifiers in lectures (Lin 2010), the use of *okay*, *right* and *yeah* in academic lectures (Othman 2010), the achievement of the coordinated transition between phases of talk and phases of silence (Rendle-Short 2005), asides (Strodt-Lopez 1991), and metadiscoursal and intonational signals of organization (Thompson 2003).

As opposed to the approach chosen in the present study, Mindt (2008), in her work on certain adjectives in different discourse settings, defines monologue as one end of a cline (with dialogue at the other end). The resulting problem of identifying features of monologue is solved by applying the model by Koch and Oesterreicher (1985), who distinguish speech of closeness and speech of distance ('Sprache der Nähe', 'Sprache der Distanz' [translation Mindt 2008]), where the former comprises "dialogue, turn-taking, the people talking to each other know each other well, face-to-face interaction, free choice of topic, private and spontaneous communication, strong involvement of discourse participants" (Mindt 2008: 1509), and the latter "monologue, no turn-taking, people talking to each other do not know each other, communication partners are not present, a set topic, public and planned communication, low involvement of discourse participants" (1510). Mindt thus adopts this *speech of distance* as a marker of monologicity.

Other monologic contexts have also received attention: The answering machine message, e.g., has been researched mostly with respect to the structure of these spontaneous monologues (cf. Crabb 1996, Gold 1991, Dubin 1987, Liddicoat 1994, Rosen 1987). Learner monologues have been investigated in the case of first language acquisition (Jisa 1987) as well as second language acquisition (Hincks 2004). Research in child language acquisition has also been interested in children's self-talk. Several terms

have been coined to describe monologic talk during stages of a child's cognitive development (adapted from Hasegawa 2010: 13): egocentric speech (Piaget 1959), private speech (Vygotsky 1986), inner speech (Vygotsky 1986), crib speech (Weir 1962) or self-talk (Meichenbaum 1977). Glick (2007) and Wells and Bull (2007) study stand-up comedy, whereas Bavelas et al. (1992), Bavelas et al. (1995), Kazandjian, Borod and Brickman (2007) and Smith et al. (2005) use elicited monologues to research gestures, facial expression or the introduction of referents.

Hasegawa (2010), in her monograph *Soliloquy in Japanese and English*, investigates soliloquy as a phenomenon that has a special status in Japanese as compared to its status in other languages (e.g. English). This special status derives from the fact that soliloquy, that is, unaddressed speech, is grammatically marked as such, due to the complex system of honorifics and terms of address that is central to Japanese language use. Hasegawa cites several scholars who consider all language use, whether in the form of thought, monologue or dialogue, as dialogic (cf. Watson 1925, Mead 1934, Peirce 1960, Bakhtin 1984). Their statements regarding, for example, "different phases of the ego" (Peirce 1960: 4), or the word as "the product of a reciprocal relationship between speaker and listener" (Bakhtin 1984: 86), are too cognitive in nature to be helpful in terms of a linguistic, data-driven investigation of the properties of monologic speech. Furthermore, they are clearly based on a structural definition of monologue (as opposed to a situational one, see my discussion of literary research on monologues in the follwing paragraphs). Hasegawa (2010: 2f) also differentiates further in that she argues that despite any degree of dialogicity, there should still be different properties inherent in soliloquy as compared to dialogue because there is no discrepancy between addresser's and addressee's knowledge and perspective, or because of the lack of potential face-threat which would lead to absence of linguistic politeness.

Literary research has also contributed to the study of monologue, in particular dramatic monologues. According to Pfister (1977: 180f), the basic distinction between two different approaches holds here, as well: two criteria can be applied, a situational and a structural one. The situational criterion for the classification as a monologue is that the

speaker a) is alone on stage, b) thinks he or she is alone on stage, or c) another character's presence is ignored. The structural criterion says that the label monologue applies depending on length and coherence of the speech in question. As mentioned previously, the structural approach allows for a further distinction into monologue and soliloquy, the former one being addressed, and the latter lacking addressivity. Thus, one approach allows for a binary classification, whereas the other evokes a scale depicting the degree of monologicity or dialogicity.

Regarding the factors that determine this degree, Pfister mentions 'semantic directionality' [*semantische Richtung*, translation MF] and a tendency for relative length (Pfister 1977: 182). Dialogue is characterized by short turns (and subsequent speaker change), and frequent shifts in semantic directionality. Monologue features long turns (or a long turn) and semantic unidirectionality. Thus, two characters whose contributions are semantically uniform (i.e. display agreement, for example) can produce dialogue that assumes a fairly monologic character. On the other hand, two characters whose turns do not cohere at all ["beziehungslose Repliken" (183)] likewise each produce monologue-like speech. A monologue can be made more dialogic through address of a non-present entity, as this creates a shift in semantic directionality. Address of self, such as when a character reveals a second personality and carries out an inner dialogue, and address of the audience in an *ad spectatores* also moves a monologue towards the dialogic end of the continuum.

This section has given a brief overview of research on monologic speech. It has demonstrated that a more comprehensive study of the notion of monologue is desirable, which the present study approaches by providing a first step towards the comparison of two monologic genres.

1.3 YouTube as a phenomenon

This section describes the website YouTube with regards to the communicative channels it offers to users and the ways users employ them for their communicative purposes. A

review of the literature about the (video) content on YouTube delineates the kind of content vlogs represent. Comparisons with other sites of communication and a discussion of the differences in reception situation result in a construction of YouTube as a site where social interaction takes place. The role written comments play for YouTube users is central to this perspective and therefore receives in-depth treatment in this section. Thus, this section situates vlogs as linguistic data occurring in a highly interactive CMC environment.

In its "about" section, YouTube describes the services it offers:

> Founded in February 2005, YouTube allows billions of people to discover, watch and share originally-created videos. YouTube provides a forum for people to connect, inform, and inspire others across the globe and acts as a distribution platform for original content creators and advertisers large and small (YouTube n.d.).

In addition to the viewers of YouTube videos, this statement identifies two main types of users: original content creators and advertisers; and it characterizes YouTube as a forum for interaction. These self-proclaimed goals allude to the main characteristics of *Web 2.0* sites. The website YouTube was launched around the same time the term Web 2.0 (O'Reilly 2005) was coined, in the middle of the first decade of this millennium. Web 2.0 is characterized by, among other features, decentralized content creation which is driven by user contributions (O'Reilly 2005). YouTube's architecture firmly situates it as a Web 2.0 network: anyone with access to high speed internet can upload and view video content free of charge; video content is selected, titled, described and tagged by the uploader; video content is searched, watched, rated, recommended ("shared") and commented on by the viewer; the embedding function allows video content to be incorporated into other online spaces, e.g. blogs (Marlow et al. 2006; Gehl 2009). In light of these features, YouTube can be characterized as occupying "an institutional function – operating as a coordinating mechanism between individual and collective creativity and meaning production" (Burgess and Green 2009: 37).

The December issue of *Time Magazine* chose "You." to be the person of the year 2006 (Gehl 2009, cf. Figure 1.1). This choice was supposed to reflect the decentralized organization of the Web 2.0 network. The computer screen forming the central part of the image is based on the then current design of the video frame of the YouTube website with its few basic navigating functions.
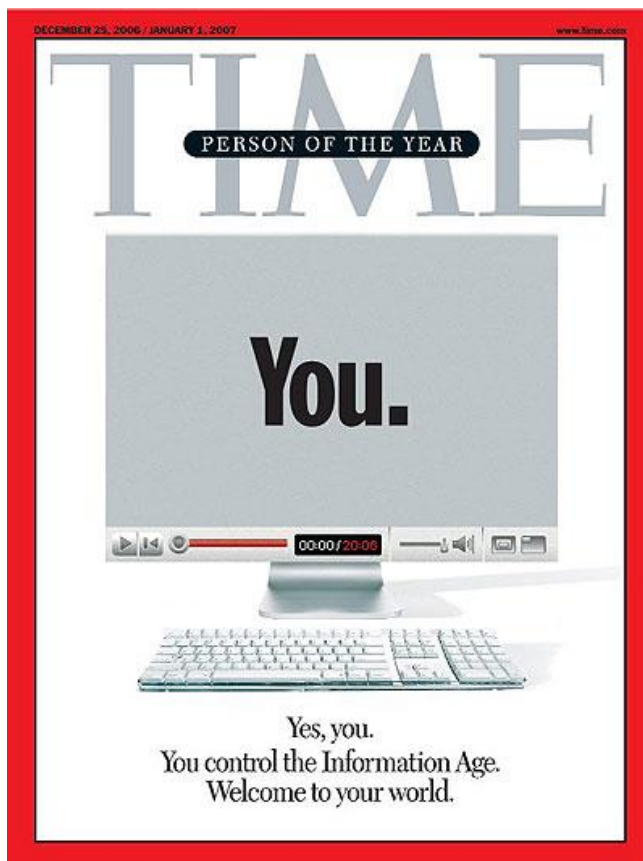
Figure 1.1, Cover of *Time Magazine*, December 25, 2006

Tolson (2010: 285) discusses YouTube in relation to television: YouTube is a network where clips of media professionals, non-professionals and celebrities are interlinked without an institutionally pre-imposed hierarchy. Unlike television channels that produce and compose a program, YouTube allows its users to make their own choices. Tolson follows Lister et al. (2009) in calling YouTube a form of 'post-television'. YouTube has also been likened to vaudeville performances. Jenkins (2006: no pages) draws parallels between the video platform and the performance culture: both forms are

characterized by their mixture of content, immediate emotional impact, mastery and virtuosity on display, memorability of successful contributions and use of technology.

1.3.1 Functions on YouTube

YouTube allows users to create accounts. Each account has a *channel page* that shows some personal information (name; age; country of residence; characteristics/character traits; etc.) if the user chooses to disclose it. It also lists the user's own video uploads and those that he or she *favorited*. To favorite a video can be interpreted as a display of identity (Burgess and Green 2009: 46), in that one publicly states a preference of certain types of content over other types. In order to upload a video, one needs an account. Until recently, the length of videos on YouTube was restricted to ten minutes, which now has been changed to fifteen minutes[1]. Videos can be uploaded in various formats; they are then converted to *.flv files*. The uploader can choose a title for the video which is displayed above the video frame. Likewise, he or she can add a video description. This appears in a box underneath the video frame (until March 2010, this box was located next to the video, on the right hand side). The uploaded video can be tagged: the uploader chooses terms that describe the content of the video (these and the title play a role for YouTube's search function). Videos can be set to one of three modes: One, the video is available for anyone to watch; two, the video is shared privately, which means up to 25 "YouTube friends" can be invited to watch it; three, it can be marked as "unlisted", which means that a direct link to the video is needed to watch it.

Viewers find videos to watch either because they are linked to on the YouTube homepage or in other environments (Facebook, emails, blogs, websites etc.). Alternatively, viewers can actively search for videos by using the website's search engine. Upon submitting one or several search terms, videos are presented as search results that contain one or several of the search terms in their title, description or tags. Search results can be filtered according to various parameters, such as upload date,

---

[1] Both the 15 minute time limit and the default format .flv file have been changed again since the writing of this text.

subject category, video length etc., and they can be sorted by relevance, upload date, view count, etc. YouTube presents the search results in form of a list that includes some basic information: next to a still image (usually a frame from the video), there is the title, length and (beginning of) the video description, as well as the time that has passed since the video was uploaded and how many times it has been viewed.

Viewers can then start a video by clicking on the still image. Videos are played in an embedded player on either the channel page of the uploader, or the uniform YouTube website environment (or, if embedded or linked to another website, they can be played in a completely different surrounding). In the YouTube environment, there are links (in form of images) to other videos next to the video being played (labeled as "related videos"). Viewers can rate a video by clicking on either a thumbs up or thumbs down icon. The results of this voluntary survey are displayed underneath the video as a bar graph, representing the total number of votes as "likes" and "dislikes". This rating system replaced the star system that allowed viewers to rate a video with up to five stars in March 2010. Viewers can also favorite a video, that is, add it to a list of videos that is displayed on the viewer's channel page. The most interactive forms of viewer response subsumed under the heading "social action" in YouTube's own terms (cf. YouTube n.d.) are video responses and written comments. They both require creativity that goes beyond clicking on icons on the website. A video response is a video that is posted in response to another video. It appears below the first video as a clickable link in form of a still image. Written comments can be typed into a box below the video, with a maximum length of 500 characters. They can contain so-called *deep links*, time stamps that link to a certain moment in the video. Thus, a commenter could post "the dog is so smart, especially at 1:40!!" (example from YouTube's help article, YouTube n.d.). Another YouTube user clicking on the time stamp will be referred to the frame one minute and forty seconds into the video. Written comments can also be rated: each comment has thumbs up and thumbs down icons. The uploader's comments are displayed at the top of the comments section, followed by those that receive the most thumbs up ratings ("Top Comments"); finally all other comments are shown ("All Comments"). Users can post direct answers to comments. These answers are directed at

another commenter by the automatic indentation of the post. Finally, a YouTube user can subscribe to other users. This entails being listed as a subscriber to another channel both on one's own as well as on the other's channel page. Additionally, whenever the subscriber logs on to YouTube, the (subscriptions') new uploads, favorited or liked videos of the users he or she subscribed to are displayed.

1.3.2 YouTube as an archive and a self-proclaimed site for social interaction

YouTube represents an archive (Gehl 2009; McKee 2011), or even a "democratic digital archive" (Gracy 2007: 193). As content is not created by YouTube itself but by the users and it is the users who describe, tag, rate and comment on the content, they are "the curators of storage and classification" (Gehl 2009: 46). YouTube's search function uses an algorithm that is based on the titles, tags and descriptions that video uploaders attach to the content (Gehl 2009). Thus, entering the term *car*, for example, into YouTube's search function will yield a list of videos that have *car* in the title, description or as a tag. Neither the uploader's, nor the video's popularity (in terms of view counts), nor the actual content of the video influence the order in which the search results are presented (Gehl 2009). With increasingly large amounts of content uploaded, accurate tagging and titling of videos is indispensable if uploaders want viewers to retrieve videos successfully. YouTube itself offers little guidance or prefabricated terminology, except for a broad categorization, in the way of systematic storage (cf. Gracy 2007: 194). Social tagging systems, such as YouTube's tag function, "rely on shared and emergent social structures and behaviors, as well as related conceptual and linguistic structures of the user community" (Marlow et al. 2006). The obvious shortcomings of this non-professional curatorial practice, that is, possibly inconsistent and unreliable labels, must be viewed in context with the virtues of such an approach: the decentralization of this work allows the handling of large amounts of data, which can be accessed fast, and by many people (cf. McKee 2011: 157, applying Fallis' 2008 argumentation about Wikipedia to YouTube).

The amount of video content uploaded and viewed on YouTube has increased steadily since its launch. In its section on statistics, YouTube states that "48 hours of video are uploaded every minute, resulting in nearly 8 years of content uploaded every day; Over 3 billion videos are viewed a day" (YouTube n.d.). Compared to this, in the fall of 2008, users uploaded 15 hours of video every minute (YouTube n.d.). These statistics alone indicate that the website's central services, that is storing video material and making it available for viewing, enjoy immense popularity. Apart from these fairly impressive statistics on its core functions, YouTube also provides data on those functions that render it a site where interaction takes place. Under the category "social", YouTube gives information about connections between YouTube and other websites as well as YouTube internal interaction. According to these statistics, a large number of users ("nearly 17 million") have connected their accounts to one or more social networking site, such as *Twitter* or *Facebook*. Over 500 *tweets* per minute contain links to YouTube videos, and 150 years of YouTube videos are watched via Facebook every day. As for interaction taking place within YouTube, every week "100 million people take a social action ... (likes, shares, comments, etc)", "millions of subscriptions happen each day", "millions of videos are favorited every day", and over half the videos on YouTube have been rated or commented on by the community (YouTube n.d.).

This information is interesting in two distinct ways. Firstly, the great amount of diverse traffic documented between YouTube and other sites as well as within YouTube illustrates users' acceptance of this online space as a place to interact with other internet users. Secondly, the fact that these statistics are published shows the website creators' efforts to position their product as a social network rather than a site that allows the storage of video material. The choice of terminology ("social", "connect", "social action", "community") and the above-cited description of YouTube's services reflect this positioning. Both of these observations represent reasons for undertaking a study such as the present one, exploring the interactive possibilities of a particular virtual space.

Not only YouTube itself uses the term *community* to describe its users and their actions on (and off) the website. The users, who commonly refer to themselves as YouTubers, also conceive of themselves as an online community based on their social contacts that result from the use of YouTube. Evidence of the use of this terminology can be found in many vlogs. The term *community* is likewise used in the literature (linguistic, anthropological, educational etc.) on YouTube and its users (see e.g. Boyd 2010; Burgess and Green 2009; Cayari 2011; Cheng, Dale and Liu 2007; Moor, Heuvelman and Verleur 2010; Gehl 2009; Gracy 2007; Harley and Fitzpatrick 2008; Hess 2009; Jeffries 2011; Jones and Schieffelin 2009; Lange 2007a; Lange 2007b, Lange 2007c; Lange 2007d; Snelson and Perkins 2009; Zink et al. 2009). The functions of flagging, rating and commenting on a video, which are considered typical Web 2.0 affordances, as well as the requirement to create an account (and thereby a username), enhance the sense of community (Boyd 2010: 5f). Hess (2009: 413f) considers the sense of community within users essential to YouTube's success, with some users spending hours posting responses. Jones and Schieffelin (2009: 1063) identify features of a speech community in written comments on YouTube, as they share linguistic features for producing and interpreting messages. The present study adopts this view of the users of YouTube as representing a community. The major goal of the study is to reveal some of the communicative structures that help form this sense of community.

1.3.3 Video content

The video content uploaded to YouTube is very diverse. For content survey purposes, it can be categorized into two groups: user created (non-professionally created) content and content created by and for traditional media (professionally created content). YouTube is seen as a site where these types of content merge, therefore this distinction is a simplified model (Burgess and Green 2009). User created content encompasses a range of amateur videos, such as home videos, fanvids (fan videos), vlogs, user created music videos, anime music videos, cooking videos etc. Traditional media content uploaded to YouTube contains excerpts from TV shows, movies, series or music videos

without substantial editing (Burgess and Green 2009). Copyright is often violated when users who do not hold the rights upload content that originated in the traditional media (Gracy 2007; Cha et al. 2007).

In a study designed to determine the composition of uploaded and consumed video material, Burgess and Green (2009) collected a sample of 4320 'popular' videos from YouTube. Videos were considered popular if they appeared in one of the following categories: 'most viewed', 'most discussed', 'most favorited' or 'most responded'. Overall, traditional media content made up roughly 42 percent of the sample, a little over 50 percent was user created material (the status of the remaining videos could not be determined). The distribution over the categories revealed a clear pattern: in 'most favorited' and 'most viewed', traditional media content outweighed user generated content. In 'most discussed' and 'most responded', user generated content dominated, indicating that viewers were motivated to post some sort of response (either as a written comment or as a video response) rather than passively consuming. User generated content thus has more potential to initiate exchange between visitors of the website than traditional media content.

The study also reports on who uploads what kind of content: roughly 61 percent of popular content is uploaded by non-professional users, only about 20 percent by traditional media; the rest comes from organizations (2 percent) and 'small-to-medium enterprises or Independent producers' (9 percent). The upload source of the remaining 8 percent is uncertain. Thus, the 'most viewed' and 'most favorited' categories are dominated by traditional media generated content. As the sample contains about 42 percent of traditional media generated content, but only 20 percent of the uploaders of that same sample are traditional media representatives, clearly, non-professional users upload content created by the media. Burgess and Green (2009) discuss this with regard to copyright infringement. The poor video quality and restrictions in video length make YouTube an inadequate tool for sharing full length media productions. Rather, the media 'quotes' that users upload should be seen as part of a meaning making process in a "redactional system" (49).

1.3.4 Vlogs

In terms of the 'most responded' and 'most discussed' categories, content that has been created and uploaded by non-professional users is most popular. The kinds of videos found in this category include mundane videos, short films, fanvids, hypercreative mashups, anime music videos, instructional video game walkthroughs and machinima[2] (Burgess and Green 2009: 52). Video blogs make up the most significant part of the sample. According to Herring et al. (2013: 14), conversational video exchanges or multimodal conversations more generally are among the phenomena "that can tentatively be identified as emergent and unprecedented, at least as common practices".

Moor et al. (2010: 1536) describe vlogging as "the video version of text-based weblogs". Lange (2007c: no pages) gives an inclusive description of the kinds of videos subsumed under the term *vlog*: "Video blogging is an umbrella term that covers a wide number of genres, including everything from short video footage of spontaneous, real-life, personal moments, to scripted and preplanned "shows" with characters, narratives, and professional acting". Vlogs are usually made available to a general public, but some are restricted to a select audience (Lange 2007c). They can have journalistic orientation, entertainment character, diary form etc. (Lange 2007c: no pages). The diary-based type of vlog features vloggers talking directly into a camera, relating their life experience (Lange 2007a: 39). Lange (2007c: no pages) further characterizes the members of the vlogging community as sharing "a commitment to video as a crucial means of expressing and understanding issues that the video blogger wishes to share". Thus, vlogging is a form of communication that uses video as its medium and is used as a means of interpersonal exchange.

Burgess and Green (2009:53) see its antecedents in "webcam culture, personal blogging and 'confessional culture' (Matthews 2007) that characterizes television talkshows (sic!) and reality television". They cite one project, "the show with ze frank", as the driving

---

[2] "Machinima is a term for animations made using video games: by using game characters as puppets, recording the visual output and recording a new audio track for them" (Gauntlett 2011: 259).

force in establishing and defining the genre, imprinting formal characteristics on it, such as rapid editing and snappy performance to camera. Matthews (2007: 438), studying the television project *Video Nation*, which was created for television and featured British citizens presenting aspects of their everyday lives, notes about their filming practices:

> Of my sample of 81 shorts from across the five-year run of the programmes, all but three include direct address to the camera. The contributors to *Video Nation* do not simply turn the camera outward toward the environment in which they live, but towards themselves, capturing themselves.

This clearly identifies these short films as predecessors of vlogs, except that they are situated in a television environment that lacks the possibility for interaction.

A crucial feature of vlogging is its conversational character (Burgess and Green 2009; Tolson 2010). This becomes evident first and foremost in the exchange that takes place through the written comments on the website underneath the video. Written comments do not necessarily comment on the vlogger, but they recycle the vlogger's statements, they answer questions, they ask questions, they show reactions to humor etc. (cf. Tolson 2010). That is, they contain various kinds of responses to the issues raised by the vlogger. They can be addressed to the vlogger, or they can be addressed to other commenters, which means that in some comments sections exchanges of written messages between vlogger and viewer or between several viewers take place. According to Tolson (2010: 282), both "spelling mistakes and non-standard punctuation contribute to their general impression of immediacy and spontaneity".

YouTube comments have been the subject of scholarly attention (cf. Boyd 2010; Harley and Fitzpatrick 2009; Hess 2009; Moor et al. 2010; Lange 2007d). In a case study of one (male) vlogger's first eight videos and the comments to them, Harley and Fitzpatrick (2009) argue that it is the social context of YouTube which enables the co-creation of his personal narrative and inspires the user to tell his story. This co-creation is driven by the support, interest and affection that are expressed through the comments to his videos. In comments to political speeches, users recontextualize the text of the original speech and create new, often discriminatory discourses (Boyd 2010: 3). Investigating comments

on commercials, Jones and Schieffelin (2009: 1063) note that viewers recycle catchphrases from the videos, "establishing them as quotable forms". The analysis of the discursive framework of responses (both written and in video format) to the "Above the Influence" antidrug campaign videos led Hess (2009: 412) to the conclusion that "digital activism is limited in the sea of responses, replies, and often dismissive and overly playful atmosphere found on the video blogging site". While this indicates that comments can lack seriousness and may not be conducive to democratic debate, other research concerns comments that disregard established norms of the YouTube community. Burgess and Green (2009: 95) identify *haters* as users "who haven't uploaded their own videos but leave abusive comments in the discussion threats of other users' videos". Lange (2007d), based on interviews with YouTube users, reports on YouTubers' attitudes towards abusive comments on their videos and their suggestions on how to prevent viewers from posting them. While some accept them as part of their activity on YouTube, others suggest moderating systems or user requirements. These offensive comments are a central topic for vloggers, as many of them address the issue in their posts. Moor et al. (2010) use the term *flaming* to describe this phenomenon (although they state that the term has been used too inconsistently to be assigned a clear definition). They explain that "while some YouTube users intentionally offend others for mere entertainment, most flaming seems to be meant to express disagreement or an opinion. Feeling disappointed by a video or feeling offended by either a video or another commenter were popular reasons for flaming" (1544). Thus, the communicative character of YouTube comments is highlighted in that even offensive posts can be seen as reactions to video, albeit unwelcome ones due to their hostile character.

Tolson (2010: 279), discussing whether vlogging can be assigned the label "authentic" vis-à-vis the "pseudo-conversations of traditional broadcasting", identifies three key dimensions of vlogging: presentation, interactivity and expertise (280). Besides user responses in the form of comments, he cites the "excessive direct address and the transparent amateurishness" (286) inherent in vlogs as features that might add to the impression of interactivity and consequently authenticity. Further, he characterizes the

manner of presentation in the data he discusses as foregrounding the production process, with clear evidence of the home-setting and lack of editing that is apparently due to the vlogger's naiveté rather than a desired effect (281). Finally, vlogging enables a form of communication, blending together traditional speech genres associated with expertise and colloquial talk, resulting in hybrid forms of talk featuring the "ordinary expert" (283). However, the term *authenticity* is limited in its analytical usefulness in that it is relative and thus subject to critique (286). Jean Christian (2009: no pages) discusses the rules and codes of vlogging in terms of authenticity from the perspective of the audience. This includes aspects of topic choice, filming and editing as well as the discussions about these matters within the vlogging community. He suggests that the reception of vlogs might steer away from considering authenticity as a crucial characteristic as long as they relate emotionally to the audience (2009: no pages).

The kinds of video content covered by the term *vlog* have yet to be delineated precisely (cf. various quotes above). One type of vlog that has been identified and discussed is the *haul*.

> The ostensible purpose of a haul is to display recent clothing, accessories, and makeup purchases, sometimes simply showing the items and sometimes offering a quick demonstration of how they will be worn; a typical video is under ten minutes long and is filmed with a webcam in the video blogger's, or *vlogger's*, bedroom (Jeffries 2011: 59, [emphasis in the original]).

Jeffries (2011: 61) demonstrates the established structures of normative behavior among the "young female consumers" that are identified as the main group in the demographic of hauls and their viewers. These norms include the exchange of compliments, especially regarding natural beauty, the possession of the products presented, and personal, subjective qualities, for example sweetness and cuteness (64). On the other hand, vloggers are expected to pay detailed attention to makeup, hair and generally the presentation of their bodies and to regularly purchase and present products (64). There is also a government regulated norm that is widely adhered to. It states that if the product was a free sample, the vlogger must disclose this fact (61). The

most prevalent norm in hauls is the projection of a friendly and pleasant persona with a positive attitude to the products presented (67). Subgenres of the haul are, for example, bargain or discount store hauls, hauls by Asians and Asian-Americans and plus-size hauls. Clearly, there is much potential for subgenres of the vlog to further diversify according to user demographics. A general survey of the genre vlog and its subgenres, resulting in a taxonomy of the genre and its users, would be a useful tool for analyzing vlogs from various approaches.

The interactional aspect of YouTube highlighted and illustrated so far can be complemented by another motive: Haridakis and Hanson (2009: 330) discuss YouTube viewers' co-viewing habits, postulating that it is a means of sharing content with others. As with television, watching YouTube can serve as a joint activity. Due to its continuous availability, it can enhance both "during viewing interaction and post-viewing social activities" (330).

This section has reviewed research on the website YouTube. The video sharing website represents a Web 2.0 network, which means that its users create, manage and consume the content. The interactive character of the website is a result of the various functions that YouTubers can make use of besides uploading and viewing videos. Written comments, in particular, prove to be the channel where the bulk of the interaction surrounding the videos takes place. The genre vlog, which is part of the content not recycled from the classic mass media, encompasses a range of subgenres, making it difficult to clearly delineate what exactly counts as a vlog. In summary, YouTube represents a web space that hosts various opportunities for people to interact. Despite its monologic character, vlogs also reflect this interactivity. This is the main research interest of this thesis.

1.4 The data

This study is based on a continually growing corpus of vlogs. To date, it comprises over 100 videos uploaded to YouTube, with the majority first published between 2006 and

2010. These videos have been assembled to form a two part corpus: Part 1 is a systematically selected sub-corpus of 46 vlogs that were included according to very specific criteria; Part 2 is a collection of vlogs that partially extends Part 1, and partially contains vlogs that feature points of special interest to the present study. Before I detail the respective criteria, I explain the reasoning behind this corpus design.

The motivation for this particular two-part structure of my corpus grew over the course of the project rather than informing it from the outset. When presenting my work at conferences, especially those with a sociolinguistic focus, or when submitting to journals, a recurrent question concerned the basis for my choice of examples that I analyzed and presented. It became increasingly clear that a systematic exploration of language use in vlogs requires a well-defined corpus with a set of stable variables, on the basis of which frequencies of selected phenomena can be established. This allows for a choice of examples reflecting practices that are central to vlogging, rather than speaker idiosyncrasies. Thereby, so-called *cherry picking* can be avoided: the choice of a particular data item that happens to reflect just what one wants to demonstrate, contrary to what the majority of the data indicates.

This should not be confused with the analysis of a deviant case, as discussed (and developed into a major theoretical contribution) by Schegloff (1968): The occurrence of a single phone call opening that did not follow the previously established distribution rule among 500 that did, led to a reconsideration of the ongoing action in terms of summons-answer sequences, which enabled a more general formulation of the rule which, in turn, also covered the deviant case.

The present study is concerned with a young (and societally marginal) genre, as opposed to the phone calls investigated in Schegloff (1968). I assume that any conventions regarding appropriate practices in vlogs are still being negotiated, if there should ever be a type of consensus as there is for phone conversations. This has direct implications for the use of my vlog corpus: I cannot (and, in fact, I do not) expect to find such clear deviant cases among a large set of uniformly structured comparable data sets, for there simply is no such established standard at this time. My corpus is much more a snapshot

of a genre at a particular stage of its development, from which certain tendencies can be derived. Thus, it serves as the balancing pole on a tight rope: it provides a picture of what are relatively typical practices in vlogs as opposed to rare practices, so that they can be recognized as such and subsequently analyzed as such. In this way, cherry picking is prevented, giving way to the analysis of practices precisely in terms of their status as rare or common rather than oblivious to this status.

The two parts of my corpus allow just that. **Part 1** was assembled following a vlog search on March 17, 2010. The search involved typing the term *vlog* into YouTube's search function. The algorithm of this function (at least in 2010) was designed to return videos that had the particular search term in the title. From the list of videos that this search returned, I included the first 46 that a) (predominantly) used English as spoken and/or written medium, b) contained spoken language, c) (predominantly) featured a single speaker talking into/for the camera (rather than a single speaker addressing a physically co-present, silent person).[3]

The choice of search term, and indeed the whole method of using the search function rather than navigating videos in other available ways (featured videos, suggestions etc.), is based on the lack of an established definition of the genre vlog. Thus, rather than choose certain videos from among the multitude of videos uploaded daily to YouTube, and label them as vlogs according to vague criteria, it seemed more sensible to choose videos that had been labeled as vlogs by their creators who are also the ones providing the title of a video that is uploaded. The advantage of this method is that it circumvents the genre problem, in that it relies on the knowledge of members of the vlogger community. The downside is, of course, that vlogs that were created and uploaded by a vlogger as a contribution in the form of a vlog, but did not have the word *vlog* in the title, could not be included.

---

[3] I arrived at the notably 'unround' number of 46 for practical reasons: initially I had planned for 40 vlogs. When I made the decision to collect comments to vlogs, some of the comments sections of the original 40 were not available, thus I included more vlogs from the original search to get up to the round number of 30 comments sections – with the effect of adding another 6 vlogs.

The sample thus collected contains vlogs by 19 women and 27 men, predominantly from the US, but also a few from Canada, Europe and Australia (according to the information they give on their channel page). All except two of them appear to be native speakers of English. The vloggers range in age from 10-60. The shortest vlog in the sample is 1 min 7 sec long, the longest is just under YouTube's limit of a little over 10 minutes (in the meantime the limit has been raised to fifteen minutes).

Part 2 has been assembled following the same formal criteria as Part 1 (spoken language, English, monologue), however, the choice of videos was not based on a search term, and it was not reduced to one point in time. In fact, some of the vlogs date back to 2006, while the latest ones are from 2012. What I tried to avoid in Part 1, an analyst's labeling of a video as vlog, was a natural consequence of my approach in Part 2. The 62 vlogs come from just a few, sometimes interconnected YouTubers. Thirty-nine vlogs alone were made and uploaded by a single vlogger. Some of these vloggers are particularly experienced or successful in terms of number of vlogs and view counts. In Part 2, the proper use of the genre label *vlog* hinges on the identity of the uploaders as vloggers. I established this status through the larger context of each video, that is I paid attention to whether the YouTuber in question referred to him- or herself as a vlogger in general, or used the label *vlog* in other contexts than the video title.

Written comments to the vlogs form an adjunct part of my corpus. The necessity to collect written comments as data pertaining to the genre also became obvious during the manifold re-consideration and re-shaping of the object of this study. As a result, 30 comments sections from Part 1 vlogs were collected, as well as selected comments or whole sections from Part 2 vlogs.

The particular data sets used will be laid out again in each of the subsequent analysis chapters respectively, for the specific research aims and methodologies demand a certain degree of flexibility. Moreover, these chapter-specific orientations demanded minor variations in the transcription conventions.[4]

---

[4] A description of the comparative data corpus of TED Talks will be provided in chapter 8.

1.5 Transcription conventions and ethical considerations

The spoken data used in this study has been transcribed carefully according to the transcription conventions which can be found in the appendix. The most striking feature of these transcripts is their structuring of the spoken language data into intonation units (cf. Chafe 1982, 1994). Furthermore, all dysfluencies (e.g. false starts, stutters, pronunciation difficulties) are faithfully represented, rather than edited out. This reflects my interest in language use as it naturally occurs, as opposed to an abstract competence that allows speakers to pass judgment on the grammaticality of a syntactic unit.

Various researchers have been paying considerable attention to transcribing as a method, one of the most prominent being Gail Jefferson, whose transcription system is the foundation for work in the CA paradigm and beyond. Her system of notation is presented and explained in the appendix to *A simplest systematics for the organization of turn-taking for conversation* (Sacks et al. 1974). She stresses the importance of crafting a careful transcript of speech that represents actual realizations of pronunciations, overlap, etc. Ochs (1979) argues for an organization of a transcript that is locally adapted to the research interests at hand. She cites the case of child language acquisition research, where a transcript developed according to adult conversation norms (e.g. top to bottom turn representation on the page) might suggest turn coherence that simply is not yet present at certain stages in a child's development. For a brief overview of transcription methods for various fields of interest see Crystal (1985), and for a synthesis of transcription conventions used in the past in discourse analysis, see Dressler and Kreuz (2000).

Regarding the transcribing of spoken language, it needs to be pointed out that this is always a selective process on several levels (Gerhardt 2008). Not only does a transcriber have to make conscious decisions as to what should be represented in the written account they are producing, but the recording from which those transcripts are furnished will influence this decision via sound quality, microphone range etc.

For the present data, this latter point concerning recording is of particular interest: unlike many other speech situations, talk in vlogs is created to be recorded. This

characteristic of vlogs elegantly eliminates the problem of the observer's paradox (Labov 1966): the 'camera observation' is a defining feature of the genre. Moreover, a vlogger (or uploader of the material) can edit the footage. Therefore, the decision as to what is audible, visible, or somehow noticeable in general in the video lies with the producer. Consequently, only what has been approved of by the video maker is eligible to go into the transcript in the first place. This represents a crucial difference from recordings of speech situations that are not produced 'for the camera' on a regular basis, such as dinner table conversation, or office hour talk. For the transcripts (as well as the analysis) this means that the editing a video was subjected to before it was uploaded as a vlog is part of the finished product. As such it is transcribed and, where applicable, commented on in the analysis.

Also innate to this arena of language use is the multimodal setting that is assigned via the technology used: vlogs are audio-visual information, as well as being embedded in a Web 2.0 space that itself is a complex, dynamic site. Where multimodality is an analytic dimension, researchers often try to represent their precise object of study not only by using a transcript of words, but with the help of additional visual material, e.g. pitch contours to represent intonation issues or images of gestures or gaze behavior. This study also makes use of this type of material: as vlogs are received on a computer screen (and recently also other mobile devices that have such screens), one reasonable device to create a faithful representation is the *screenshot*. I have adopted this technique to provide images that complement the transcripts of spoken language. Just like the transcripts, screenshots are selective; for even if a screenshot comprises the whole of the screen, the video itself often is only a smaller part of the whole website (which could be accessed by scrolling). Consequently, screenshots, too, demand a decision as to what should be included, and what can be left out. Therefore, the present study contains images with a varying number of elements. Some of these images only show the video screen, some of them show the video screen and its immediate surroundings, some only show comments, and (very) few show the whole screen. This choice, of course, depends on the goal of the analysis, and it was always influenced by the attempt to make the data representation as accessible as possible.

Where research on humans and/or their behavior is conducted, there should be strict standards regarding the protection of said subjects and their privacy. This applies to behavior that is recorded online just as much as anywhere else. The Association of Internet Researchers (AoIR) has published guidelines for the conduct of ethically acceptable research (Ess 2002, Markham and Buchanan 2012). Issues revolving around the concept of *human subject* and the public/private distinction are discussed in these guidelines, resulting in a process approach that involves ethical reasoning in every step of a research project, and which is case-based, that is, individually adapted to the case at hand.

The present research project uses audio and visual data from a number of internet users. The type of data and the limited availability for contact to its creators poses a challenge regarding ethical issues. Considerations of various points need to be balanced to explain the decisions I have made in this respect:

- The data used in this project were uploaded to a public platform, accessible to anyone without any limitations (limiting access to certain people is an option on YouTube).

- YouTube's terms of use state: "You also hereby grant each user of the Service a non-exclusive license to access your Content through the Service, and to use, reproduce, distribute, display and perform such Content as permitted through the functionality of the Service and under these Terms of Service" (YouTube 2013: no pages).

- The two above-mentioned points are likely to be understood quite differently by vloggers themselves, if, in fact, they are known to them: a vlogger whose vlogs regularly attract only a handful of views (perhaps by the same people) might not think of YouTube as a public forum; and it is quite feasible that an average YouTube user does not even read the terms of use.

- Linguistic-pragmatic analysis per se is in no way directly harmful to a vlogger.

- The publication of this research might cause harm to a vlogger if their identity can be linked to cited material (both in the form of images and text) and if the vlogger no longer wishes to be associated with this material.

- Minors in particular, who might not have developed the faculties to apply elaborate reasoning regarding the long term effects of their actions in online space, need to be protected from such possible harm.

Wherever I could not avoid using elements of the data that might help reveal a vlogger's identity, that is, when using screenshots, I contacted the vlogger to ask for permission to do so. All of them were well over 18 years of age. In those cases, I also retained the original (user) names of the vloggers in the transcripts. Elsewhere, the data were anonymized. This practice is in line with Kozinets (2010: 142), who says that "analyzing online community or culture communications or their archives is not human subjects' research if the researcher does not record the identity of the communicators". In conclusion, the research project reported here adheres to ethical standards that protects those people whose behavior it studies to the best of my knowledge and abilities, without, however, compromising its goals or methods.

1.6 Overview

The remainder of this investigation is made up of three major analytical parts, each subdivided into two chapters, a comparative analysis, and a conclusion. The first set of chapters covers core CA topics, namely the structure of openings and closings of vlogs. The second set builds on topics originating in Interactional Sociolinguistics, as it deals with participation frameworks and with involvement through repetition. The third set tackles issues of multimodality, discussing pointing gestures and written comments. Following the three sets of vlog analysis chapters, a comparative analysis of another monologic genre, the TED Talk, is provided to illuminate what is unique to the vlog setting. Finally, the conclusion provides an agenda for the study of monologue based on this comparison.

2 Beginning a monologue: the opening sequence of video blogs[5]

2.1 Introduction

This chapter describes the opening sequences of video blogs (vlogs). It presents the various ways people make use of conventions from other genres, how they make up and try to establish their own conventions, and how these conventions facilitate the production of a vlog, especially the opening sequence. Research on openings and research on monologues is applied to vlog data, integrating the two strands.

Schegloff (1968), in his article "Sequencing in Conversational Openings", focused on the allocation of speaker roles (a and b, or first speaker and second speaker) between two parties. Assuming that conversation between two or more participants is the default, most frequent pattern of spoken language use, then monologues, which lack an opening phase constructed around fairly quick exchanges of turns, must pose a challenge. Garrod and Pickering (2004) claim that dialogue is easier to produce than monologue because of an interactive processing mechanism that leads to the alignment of linguistic representations between partners. However, in the production of a monologue, there is no second speaker to negotiate these roles with. Therefore, single speakers must develop strategies that compensate for the missing co-construction which entails a lack of turn-taking and negotiation of speaker roles. These strategies are a rich area of study whose exploration will help illuminate generally how speakers adapt to their speech situation, not only in the context of monologues, but also in the context of newly developing media more generally. The current study gives an overview of the different strategies vloggers employ, proposing a taxonomy of different practices in opening sequences. As vlogs represent a fairly young genre which is still developing, this categorization cannot encompass the complete range of openings.

The present investigation applies previous findings from research on monologues and research on openings to the novel area of vlogs. The study lays out the features of talk that is bereft of salient dialogic elements, that is, phatic communion and sequentiality, which cannot occur due to the missing interlocutor. I demonstrate how the transition from silence to talk is managed instead, using strategies borrowed from other monologic genres, while taking into consideration the technical parameters of the vlog genre. These parameters render the data unique in the study of computer-mediated communication or mediated discourse in general, since two phases of production of unscripted, spoken material need to be distinguished, as I show in the next section.

Everything that is usually negotiated in conversational openings (e.g. availability for interaction, identification, social status, alignment of participants) is irrelevant for vlog openings, since (some of) these things happen via different channels. So if opening sequences in vlogs are not actually needed, one must ask why some speakers have an introductory phase. As early as 1956, Horton and Wohl coined the term *para-social interaction* to describe practices used in TV presenters' talk, such as direct address of the audience. Tolson (2005: 10) states about greetings directed at the audience in monologic media talk that "the talk constructs a place for *potential* interaction, whether or not it is taken up in practice". He continues: "It might simply be a way of reaching out to the active listener, provoking a basic form of active listenership". Transferred to vlog openings, these statements assign interactive elements of the vlog text the function of persuading the viewers to make use of the various features of the website that allow them to reply: writing comments, rating the video, sending a personal message or posting a video response.

On a theoretical level, this chapter (as well as this whole study) contributes to the description of monologues as interaction by revealing what resources speakers use to appeal to their audience. Speakers on vlogs clearly try to make their videos interesting to their audience, to include them in the interaction by discussing relevant topics and addressing them directly, or to persuade them to engage in two-way interaction via channels outside the actual video. The basic underlying assumption is that audience

design is a driving factor in this genre, even though the audience is largely unknown to the speaker and does not take an active part in the spoken interaction. The first moments of a video are highly relevant for the viewer's decision whether to continue watching it or not, and this represents an incentive for vloggers to make their openings particularly relevant to the audience.

This chapter gives an overview of previous research on openings and on monologues. Data analysis forms the main body of this chapter, and it concludes with a discussion of the findings regarding openings, monologues, and CMC.

2.2 Research on openings

Schegloff (1967: 51) observes about the beginnings of phone calls that "openings [...] provide [...] continuity into the body of conversation" and that they are not separable or dispensable. He explains how the non-terminality of summons-answer (SA) sequences helps establish the coordinated entry into *abab* sequencing in phone conversations and how SA sequences "align the roles of speaker and hearer" (122). In presupposing that conversation is a "minimally two party activity" (1972: 1093) he implicitly states that monologues have a different status. Hence, there does not seem to be any ground for summonses in monologue openings at all.

Furthermore, Schegloff (1986: 113) states that "a multiplicity of jobs ... get done" in openings, such as checking the availability of some co-present person to engage in interaction, (re)constitution of the relationship of the parties, and negotiation of topic. Since in vlogs none of these functions are necessary for the talk to proceed, a recognizable opening sequence seems dispensable. The functions listed by Schegloff are non-negotiable in the vlog situation. The viewer can choose not to enter the situation at all by not playing the video. However, this decision is based on factors relevant before the talk starts (e.g. the username of the vlogger, the title of the vlog, its description, and the thumbnail image one sees on the website). It is not subject to some joint activity in the talk. Therefore, we can assume that openings in vlogs must fulfill other functions.

Since viewers can also decide to stop watching the vlog at any moment, one principal function might be to raise the viewers' interest and induce them to watch the whole vlog. Similarly, Labov and Fanshel (1977) use the concept of reportability to describe storytellers' justifications for their narratives to prevent rejection from the interlocutors on the basis of being ordinary.

Laver (1975, 1981) observes that the use of routine formulae is most dense in the marginal phases of interaction, that is openings and closings, and that negotiations about social status and role are conducted with the help of formulaic phrases, address terms and phatic communion. This serves to "defuse the potential hostility of silence" (226). In this sense, phatic communion cannot take place in a vlog since there is no immediate negotiation. The opening sequence can nonetheless be used by the vlogger to display status or social role.

Gold (1991) identifies four sections in answering machine messages and describes the devices callers use to compensate for their absent interlocutor. She observes that the "greeting section" (246) contains elements borrowed from other genres: greetings are reminiscent of the ritualized salutations in letters; information such as date and/or location and time also occur in the letter frame. Furthermore, she likens answering machine talk greetings to written language as both contain features that compensate for the time/space gap between interlocutors. Gold (1991) describes self-identification as a feature of answering machine messages, while she deems it pragmatically inappropriate in both face-to-face and regular telephone conversation. Vlog openings contain features borrowed from this genre as well: some contain self-identification and the date the video is posted. Nevertheless, vlogs and answering machine messages differ in important respects: vlogs are audio-visual rather than audio data, and they are recorded with the knowledge that one is about to produce a monologue which can therefore be planned in advance. Answering machine messages are unplanned recordings by callers who expected to have a dialogue with another person. Vloggers can edit their material after recording it; callers speaking on answering machine tapes cannot do so in most cases.

Liddicoat (1994: 308), expanding the research of Gold (1991) to cover both the pre-recorded message by the called party and the spontaneously recorded message by the caller, points out "that technological means of communication place constraints on talk which lead to systematic modification of the available stock of conversational routines". Interactive routines that regularly take place in face-to-face or telephone interaction may be truncated or avoided (307). Since vlogs, too, place constraints on talk, it is to be expected that modifications or a different usage pattern of routine phrases will apply.

Aijmer (2007) analyzes two sets of answering machine message data, studying the use of routine phrases. She says of answering machine message openings that "a greeting phrase suggests that you are aware of addressing a person rather than a machine". And "a particular event may, but need not be accompanied by a routinized phrase" (329). Thus, she distinguishes between messages where callers imagine or construct a recipient and those where callers view the answering machine only as a machine. The concept of an imagined or constructed audience helps explain why vloggers greet and address their viewers. But it is not one-to-one applicable to the vlog situation since it is difficult to argue that vlogs without greeting and address were not filmed with an imagined audience in mind. Moreover, the study illustrates that routine phrases used in certain speech events or certain media develop over time. This suggests that for vlogs this process might be in its early stages, and quite possibly it will take another few years until comparable conventions can be observed in vlog data.

Two studies that are concerned with openings in written CMC contexts are Rintel et al. (2001) and Waldvogel (2007). Rintel et al. investigate openings in internet relay chat (IRC). They discuss the various channel entry phase (CEP) progressions, made up of automatically generated messages by the server and the users' own posts, which can take place when a new user joins a chat room. Waldvogel studies email openings in work place settings. As in this study of vlogs, there are email openings that contain both greeting and term of address, neither of the two, or just one of the two. The choice of combination (and the choice of words) reflect back on social variables, e.g. gender,

power structures within the work place, seniority, and on the degree of harmony and support among staff.

Duman and Locher (2008) explore monologic YouTube videos presented by the presidential candidates Obama and Clinton. The study focuses on the *video exchange is conversation metaphor* that the candidates make use of while coping with the fact that "technical limitations of YouTube do not permit the immediacy/synchronicity of conversation" (205). This metaphor is created through the "actual visual representation of the candidates in the imitation of face-to face interaction" (202). In some of these videos, the candidates answer selected questions that were posted as comments online – which is a function of vlogs, too. These presidential candidates' videos are very similar to regular vlogs, though the words spoken are presumably very carefully composed by a whole team. Vlogs, too, then might contain features that "draw upon the source domain of conversation" (219).

2.3 Data collection and methodology

The research reported in this chapter is based on the opening sections of Part 1 of my corpus as well as selected openings from Part 2. Transcribing the first few moments of each vlog and listing their features provided me with statistics on which to base my choice of examples. I noted down whether:

- there was an opening sequence resembling opening credits of a movie or TV show
- it was shot indoors or outdoors
- there was a greeting
- there was a term of address
- there was self-identification
- the vlogger stated the date of filming
- there were discourse markers
- the vlogger used cuts as an editorial feature

Apart from that, I noted down the vloggers' age, the number of subscribers, and the country of residence, which can all be found on the users' profile page (except sometimes the age, which some users do not give – in these cases I made a rough estimate). To check whether there was an effect of time difference throughout the English-speaking world on country of residence, I checked another 40 vlogs at a later time in the day (first 41[6]: 12 noon; second 40: 17 CET) for this feature. In my data collection I did not include those vlogs that featured more than one speaker, since in this study I am only concerned with monologues without a present audience. The 41 vlogs yielded the following statistical data:

18 were made by female, 23 by male vloggers;

11 vlogs were headed by a sequence containing written opening credits (and sometimes music),

30 started directly with a person talking or images of other things (e.g. trees, a computer screen, a street scene);

8 vlogs were filmed outdoors, 32 indoors,

1 had audio only accompanied by a still image;

24 contained greeting terms, 17 vloggers did not;

22 vloggers used terms of address for their audience, 19 did not (as Figure 1 illustrates, 3 without greeting had a term of address, five with greeting had no term of address, 14 without greeting did not have term of address either, 19 with greeting also had a term of address);

---

[6] As one of the early parts of this study, this chapter is based on a pre-final version of the corpus, hence the number 41.
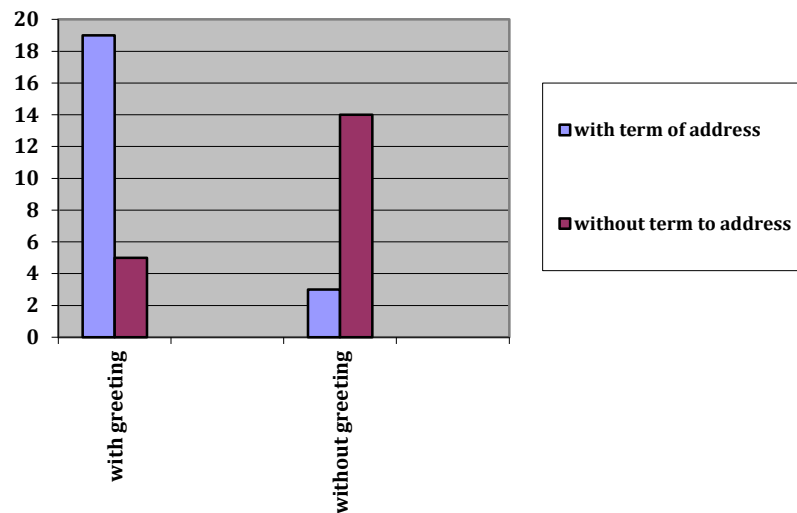
Figure 2.1

26 vloggers had no self-identification,

5 had their (user)name in the opening credits or the title, 10 verbally stated their (user)name in the opening sequence;

6 vloggers used cuts to edit their videos, 35 did not;

24 vloggers had discourse or boundary markers in their opening sequences, 6 of which used them as their first word, 17 used no such markers;

29 did not mention the date at all, 8 had it in the title or credits, and only four mentioned it verbally.

The youngest vlogger was 10-15, the oldest 50-60, the average age was 25; the lowest number of subscribers for a user was 1, the highest 263,426, the average was 10,862. Of the total of 81 users, 62 were US Americans, 9 were British, 6 were Canadians, 2 Australians, 1 each were French, German, and Dutch, with the distribution of nationalities about equal in both sets.

The next analytical step was to examine more vlogs by the same users to establish how many of them repeatedly made use of the same established patterns in their vlog openings. Twenty vloggers' videos were further examined, resulting in the following statistics:

Two had only one video uploaded; 4 had no recognizable pattern in their opening sequences; 14 had identical opening sequences in at least 50% of their vlogs. Of these 14, 8 always started their vlogs with a greeting and term of address. The most common greetings are "hi", "hey", "hello". The most common terms of address are "everyone", "everybody", "guys", "YouTube" and "YouTubers". Three vloggers regularly employed only one of the two features – either term of address or greeting. The remaining 3 used linguistic markers "okay", the date, or had several patterns that they used equally often. This small sample suggests that a majority of vloggers recurrently use identical or very similar opening sequences in their vlogs. Generally, there seems to be a tendency for vloggers to employ the structure *greeting plus term of address*.

The examples analyzed in the present chapter were chosen based on these statistics. The selection represents the most common elements of vlog openings. In section 2.4, there will be close analyses of vlog openings that contain greetings, terms of address, self-identification, discourse or boundary markers, cuts, the date, objects or animals that are the subject of talk, opening credits, and there will also be an example where none of these elements occur. Finally, the analysis will illustrate how vloggers who have already established a pattern in their opening sequence deviate from this pattern.

2.4 Types of openings in the data

The first part of this section presents examples that contain the features of opening sequences listed in the previous section in various combinations. It also describes examples that are marked by the absence of these features. The second part deals with examples by users who have established their own recognizable opening sequence.

2.4.1 The features of opening sequences

2.4.1.1 Opening containing greeting, term of address, self-identification, linguistic markers

The first example is by a male US-American in his late fifties.

Example 2.1

1	((laughs)) {waving}
2	hey YouTubers it's me Zipster.
3	okay,
4	so,
5	today I was thinking about um,
6	a couple of weeks ago me an' um,
7	me an' Sign543,
8	had to go to the DMV 'cause,

This vlogger's opening sequences are started off by laughter and a stylized kind of waving. He greets his audience with the informal "hey" and addresses them as "YouTubers". This term of address reflects the affiliation with the video hosting website (YouTube) that he assumes for his audience. At the same time, it subsumes a yet unknown number of people under their only relevant common property, that is, they are all users of YouTube. Therefore, this reflects a certain degree of anonymity inherent in some communication on the internet: in this case it is single-blind, in that at the time of production the speaker does not know who will be watching the video, whereas at the time of reception, the viewer knows exactly who is talking. Later, the vlogger can see at least how many views his or her vlog has, and if there are comments or video responses, is given a sense of who the viewers are.

"It's me Zipster." introduces the vlogger by his user name, which is also the name he gives to his character. "Zipster" is in apposition to "me", which implies that this is not the first video he has posted, and that his viewers should recognize him. The whole

clause is reminiscent of the beginnings of answering machine talk, where a caller identifies him- or herself. Since in the case of vlogs, identification is facilitated by the accompanying images of the speaker, and since the navigation through the internet/YouTube that is involved in playing a vlog raises strong expectations about whose vlog one is about to watch, I assume that self-identification might only be a secondary purpose of this clause. Rather, Zipster uses this recurrent bit in most of his vlogs to coin his own catchphrase and thus establish a unique character.

The fact that Zipster borrows this sort of introduction from answering machine talk, or possibly radio or TV talk, reflects the lack of conventions for the genre vlog. This is an example of a vlogger who apparently feels the need to have an opening sequence that is recognizable as such, and since there is no standard wording available, he provides a make-shift adaptation from another genre.

In lines 3 and 4, there are two discourse markers "okay" and "so" that signal the transition from the introductory unit to the main body of the vlog, which in this case begins with a narrative. Not only is there a shift topic-wise, but there is also a change from an informal mode (laughter and waving), to a more serious tone (change in facial expression) that allows the speaker to set up his story, leaving room for humorous build-up later on.

2.4.1.2 Opening containing greeting and term of address

Example 2.2 was posted by a male US-American in his mid-thirties.

Example 2.2 (Title: Re: Who are you....Who, Who...Who, Who)

1   hey Renetto.

2   uhm, {rubs his right eye with fingers}

3   let's see. {looks at piece of paper}

4   first up.

5   yes I am a person.

6   no I'm not a kid.

7   yes I have a job?

8   yes I have a life?

The significant difference from the previous example is that here, a single, specific, real life person is being addressed (who is most likely known to most viewers of this vlog, since this video was posted as a video response to a vlog by the user he is addressing: Renetto). This speaker displays some nervousness through a gesture he makes (cf. l. 2) while he bridges the gap between the introduction and the topic with a filled pause/hesitation marker. He then continues in line 3 with the main topic, which is headed by "let's see.". That introduces the answers to a list of yes/no questions which he apparently reads off a piece of paper. Both the addressing of a concrete person and the prepared topic outline facilitate the speaker's performance of the beginning of the vlog.

The title also points to the fact that the whole vlog is a unit of interaction, that is, the answer to a question. "Re:" is automatically added to the title of a video if it is uploaded as a video response. Thus, this instantiates a type of interaction where the basic unit is not a turn, but a whole video, which could of course be interpreted as a multi-unit turn.

2.4.1.3 Opening containing term of address, self-identification and date

Example 2.3 was posted by a male American in his late fifties.

Example 2.3

1   everyone,

2   it's Peter Schiff,

3   this is Tuesday.. uh March sixteenth two thousand and ten.

4   well,

5   the Dollar was broadly weaker today,

6   uh,

This example starts with a term of address "everyone" (l. 1), followed by a self-identification (l. 2). Only 3 vlogs out of 41 contain a term of address but no greeting. In line 3, the vlogger states the date of the video production, mentioning the day of the week and, after a short hesitation, the date. The transition from these three elements to the first topic is instantiated by the discourse marker "well" (l. 4). The first three lines are the regular beginning this vlogger uses to introduce most of his vlogs, which usually cover political topics. The date is not only found in the video description that is automatically generated on the website, but also in the (written) title of this vlog. The additional mention of the date in the spoken part would seem redundant if the only function was to inform viewers of the date. However, since giving the date is part of the vlogger's fixed routine, it also stresses the topicality with which he discusses events. Moreover, it lends the vlog the character of professional journalism.

2.4.1.4 Opening containing linguistic marker, greeting, term of address and foregrounded object

Example 2.4 was posted by a female US-American in her fifties. It illustrates two particular features: the use of a linguistic marker as the first word, and the use of an object to distract the attention.

Example 2.4

1	well-
2	((laughs))
3	hi vlog fans,
4	uhm,
5	{puts on glasses}
6	°like the glasses,°
7	[{takes glasses off again}]
8	[°this time,°]
9	um I was uh,

10   yesterday.

11   uhm,

Before greeting the audience, the speaker uses "well" as the very first utterance, followed by laughter. This is a boundary marker indicating the beginning of the vlog. Combined with laughter, it may evince the speaker's nervousness in front of the camera or her insecurity of how to start the monologue without instant feedback. The marker "well" and the laughter do not only mark the transition from the situation before the vlog begins to the greeting plus term of address; they also mark the fact that this particular vlogger has decided to make this the moment when she starts recording, though she has no obligation whatsoever to do so.

The greeting and the term of address, both conventional in openings in multi-party conversations, show that she transfers conventions from another genre into a new genre. Her choice of term of address, "vlog fans" (l. 3), not only reflects that she does not know her audience so that she has to identify them by their function as vlog viewers; it also assumes that these people are fans. This semi-humorous device seems to have the function of acknowledging the monologue situation: by saying something that might conceivably be contested, albeit rather harmless, she reminds her that, obviously, they cannot react immediately.

In lines 5-8 follows an interactive feature in form of a question (the direct gaze into the camera and the intonation suggest that this is a question), "like the glasses, this time,". It is accompanied by the speaker's putting on the glasses and taking them off again. Since the glasses are not used in their normal function to correct her vision, they serve some other purpose in this sequence. Formally, they constitute the subject of a mini sequence that is inserted between the greeting (l. 3) and the first topic of the vlog (l. 9ff). The distraction caused by the glasses is presumably a device that the vlogger uses to postpone the main body of talk in order to familiarize herself with the situation.

The added "this time," (l. 8) suggests that there has been a previous occasion involving glasses. It can be speculated that the vlogger has received comments about her glasses

(written comments via YouTube or through other channels of communication) that she acknowledges in turn. As in Example 2.2, this reflects the interactivity and recipient design of vlogs. The same is true of the following example, where the vlogger explicitly mentions that she has received questions that she now wants to answer.

2.4.1.5 Opening containing linguistic marker and cuts

In the next passage, editing, especially cutting, plays an important role. This example is by a US-American woman who, after posting several other videos, is shooting her first vlog:

Example 2.5

1	so instead of the usual HappySlip video,
2	I wanted to just connect with you guys,
3	and show you the girl behind the slip.
4	CUT
5	I'm thinking that one of these vlogs is overdue.
6	just to answer some of the questions that you have,
7	CUT
8	number one asked question is,
9	are you Filipino.
10	YES I AM. {gestures with paper roll in her hand}

The whole vlog shows no signs of hesitation, nervousness or other struggle with producing a monologue on camera. However, a regular and salient feature is the frequent cuts, which probably helped eliminate all features that might have rendered the vlog less fluent. The first three lines, up to the first cut, and lines 5 and 6, up to the second cut, introduce the vlog by way of mentioning its purpose ("to connect") and an outline of its contents ("answer questions").

These lines are headed by the discourse marker "so". Based on the intonation (level) and the inclusion in the intonation unit that constitutes line 1, it can be assumed that this "so" originally bracketed two units of talk. It is a marker of result rather than a boundary marker indicating the beginning of the vlog. Thus, the vlogger made an editorial decision that violates the conventions of regular talk, creating a situation where viewers of the vlog are given the impression that they have missed the beginning.

In lines 6 and 8, the vlogger mentions viewers' questions, and in line 9 she quotes the question she receives most frequently: "are you Filipino.". She subsequently answers that question in line 10. Consequently, this example instantiates interaction between a vlogger and her audience. Her taking up questions, quoting them as direct speech, and answering them clearly indicates that vlogs are geared towards an audience. Discussing topics that are relevant to the audience is an incentive for viewers to continue watching the vlog.

2.4.1.6 Example containing cut and linguistic marker

Example 2.6 is by the same vlogger that posted example 2.4.

Example 2.6

1	-nd I uh::
2	so I tried to do a little shopping,
3	didn't really get the shopping done,
4	uuuh didn't have any shopping to GET done.
5	I just didn't feel in the mood to shop,
6	I went to my favorite sto:re,
7	nothing.. nothing.. uh called out to me=
8	=as needing to be added to my... wardrobe.
9	((laughs))

The first line is headed by the discourse marker "-nd", and it is aborted after the subject. Like in Example 2.5, this marker indicates that there was previous talk which fell victim to an editorial decision. In this case, the editing of the video is foregrounded in that the beginning of the word "and" is not even audible. Hence, the vlog starts in the middle of a word. Line 2 is then a completed utterance, which is also headed by a discourse marker, namely "so". This utterance is a repair of the incomplete "-nd I uh::" in line 1. Whether that repair took place because the speaker realized "so" was a more appropriate organizational marker than "and" in that context or whether she started out to say something completely different which she then discarded, cannot be determined. But still, the editor's decision to keep the first line in the video reflects that she considers it a good place to start the vlog, perhaps because it is a good entry into that particular unit or topic, even though viewers end up feeling that they have missed a part of the vlog. From an editor's point of view, she retrospectively assigned that whole unit "and I uh::" a new function, that is, that of a boundary marker that marks the beginning of the vlog.

2.4.1.7 Opening containing none of the elements

None of the elements that regularly occur in vlogs are present at the beginning of the following example (i.e. greeting, term of address, linguistic marker, self-identification). Furthermore, no signs can be found that any previous part of the video footage is missing. In other words, there are no ties to foregoing talk or footage. This example is by a woman who presents a comedic vlog about the possibility of clouds as presidential candidate:

Example 2.7
    1   what really pisses me off is that no one understands that clouds are the only
         possible good president there could ever be.
    2   no president.. in history has been able to float through the sky?
    3   okay?

4   neither Obama nor Clinton NOR MCCAIN.. are cumulus?

5   CUT

6   they're not stratus?

7   they're not cirrus?

8   okay.. motherfuckers?

This passage starts with a complete grammatical sentence, including a wh- cleft and four clauses. There are no breath groups or other signs of spontaneous spoken language. This may change in the lines afterwards. However, the opening sentence sounds like the speaker either previously composed, or even practiced it. This absence of features of spoken language makes the speaker appear determined and strongly opinionated, which is in accordance with the angry expression on her face as well as the exaggerated use of offensive language. Given the unrealistic subject of her vlog, it is obvious that the speaker is not in a serious mode, but that she is displaying fake anger. As a means of stressing the (pretense) immediacy of her concern, she chooses to skip the opening sequence.

2.4.2 Variations of established, personal openings

Established, idiosyncratic phrases or behavioral patterns are an important part of some vloggers' routines. Zipster (cf. Example 2.1), for example, uses the introductory phrase "Hey YouTubers it's me Zipster." in almost all of the videos he posts online. He deviates from his routine when acting as a different character, using another phrase. As of date, Zipster has uploaded over 840 videos to his profile page. Out of 20 randomly picked vlogs in which he appears as the character "Zipster", 15 contain the established opening. 4 contain a variation, and 1 does not adhere to the pattern at all.

The following example illustrates a seemingly unplanned alternate of his usual pattern caused by a distracting noise. Often these atypical openings contain an allusion to the vlogger's typical opening or draw on them as a resource of humor. At times, the established introductory phrases are simply postponed to a later point in the video.

Example 2.8

1	Zipster:	[((laughs))]	
2	voice:	[yahoo mail,]	
3	Zipster:	yahoo mail,	{bends over towards camera,}
4		Jesus Christ.	{turns down speaker volume}
5		firstest thi^{hi}ng in the da^{ha}mn video.	
6		"hey YouTubers it's me Zipster."	
7		I gotta tell you.	
8		I wasn't even gonna do a video tonight,	

This opening sequence starts out as expected with laughter, but the usual sequence is disturbed. The interruptive voice in line 2 is an automated sound alert that notifies the user of this email provider that a new email message just arrived: "yahoo mail,". Zipster takes that up immediately by repeating the exact line "yahoo mail,", which is followed by "Jesus Christ.", expressing annoyance at the disturbance. At the same time Zipster turns down the volume of his speakers to prevent further incoming messages from interrupting him. The immediacy of his verbal and physical reactions suggests that this situation has occurred before, and that Zipster is following a strategy developed in response to it. This sound alert occurs in other videos by Zipster and is therefore known to regular viewers of his vlogs. Zipster even sometimes thematizes that alert and the disruption it causes by addressing the voice (e.g. by answering "thank you, Christine" or "shut up, Christine" – where he personifies the pre-recorded voice of an unknown woman by giving her a name). Thus we can assume that there is a shared history between Zipster and his viewers that he presupposes when he says "firstest thing in the damn video." in line 5. He can rely on the viewers to understand who is talking and why, and that he finds that interruption undesirable. The form "firstest" possibly goes back to the first vlogs he posted, where his character was supposed to portray a mentally challenged person whose use of ungrammatical forms and mispronunciations is extremely prominent. This and the laugh pulses in "thing" and "damn" signal that he is in a less serious mode than his wording might suggest.

It is important to remember that vloggers have two phases of production: first the actual uttering of their intonation units, and secondly the editing process. That is, vloggers can go back in time, so to speak, and take back or alter what they said by cutting and deleting scenes. Here, Zipster chooses to leave three intonation units (l. 3-5) that are clearly deviant from his usual pattern, in the vlog. They show him in a situation where he briefly loses his temper – or pretends to do so. Only afterwards does he proceed with his regular introductory pattern, namely in line 6 "hey YouTubers it's me Zipster." which he utters with a clearly altered voice. The vlogger's decision to leave this passage in the vlog instead of cutting or reshooting the scene indicates that Zipster assumes that something can be gained from it at the cost of being inconsistent with his well-established characteristic opening.

In the analysis of this vlog opening, but especially in the three intonation units 3-5, identity construction is a salient issue. The viewers see an irascible Zipster who is prompted into swearing for a short time. This introduces humor (cf. laughter pulses in l. 5) as a strategy to mend the pattern and return to the usual opening sequence. At the same time, Zipster establishes another recurrent (though unplanned) element of his vlogs, namely the prerecorded message spoken by a voice he calls Christine. Instead of turning off the speaker before taping the vlog, he retains the element of surprise, which allows him to digress and possibly introduce humor. Also, this makes the whole vlog look less structured and less professional. This feature allows for interaction, and it introduces a second character with a certain will of its own, since it is not directly manipulated by Zipster himself. A situation where one has to react spontaneously can help the production of an entertaining monologue insofar as it possibly makes it less monotonous.

Considering both levels of production allows for a further distinction. In this case, the interruption was not planned (though not prevented either) at the level of speech production since email messages, just as their sound alerts, come in unexpectedly. However, at the level of editing there was a clear decision to retain the interruption in the video. The next example also illustrates a deviation from an established pattern.

The vlogger is a female US-American in her late twenties. She is preparing to watch a live video broadcast while filming herself for the vlog. The phrase she regularly starts her vlogs with is "hello everyone it's Katie again.".

Example 2.9

(music sequence with images of the vlogger, her dog, a painting, noodles; the user name "k80blog!" appears at the end in orange letters)

1 okay.

2 so this YouTube live is gonna start in one minute.

3 and.. hello everyone it's Katie again.

4 and this this uh YouTube live event,

5 of the uhm of our lives really.

6 uh it- it seems to be- the way YouTube has advertized it=

7 =it's going to be... hisTORic.

8 uhm.

9 the wa:y they're going to.. IMitate television=

10 =it's really going to be historic.

This vlog starts with opening credits and music. The background images show entities that regularly appear in this user's vlogs. It is reminiscent of opening credits of TV series, where images of familiar characters and places are shown to a signature melody with an episode title shown on screen.

K80blog opens with a variation of her usual opening sequence in that the first two intonation units make it seem like she is going to start the vlog without introducing herself or greeting the audience at all. All she puts in front of the main topic is two markers "okay" and "so". The first, "okay" is a boundary marker. It marks the boundary between a pre-posed sequence of music to the sequence that shows the vlogger during speech production. Since there clearly is no foregoing talk, it is not a discourse marker in the sense of Schiffrin (1987: 31) who states that discourse markers "bracket units of talk.".

Sinclair and Coulthard (1975: 45), referring to classroom talk, note that the boundaries of what they call transactions are "typically marked by frames", a category which comprises five items: *okay*, *well*, *right*, *now*, and *good*. They say about teachers' talk that "a frame invariably occurs at the beginning of a lesson, marking off the settling-down time" (22). In the teacher's monologic passages, these frames often co-occur with so-called foci, as Sinclair and Coulthard's (1975: 49) example demonstrates:

frame   "well

focus   today I thought we'd do three quizzes"

They argue that this sort of topic introduction is typical of classroom discourse as opposed to conversation outside such settings because of the distribution of control that allows a teacher to decide on a topic or an activity. Vloggers are in a similar situation: as they are the only speakers, the choice of topic is completely up to them. Unlike the teacher in the classroom situation though, the vlogger's viewers can turn off the video at any time during the vlog. Vloggers might probably do well not to stress their control of topic too much. This way, the viewer's sense of voluntary participation is preserved, a feature lacking in classroom discourse. The downside of this control for both vloggers and teachers is the obligation to fill a turn that often by far exceeds the usual turn length in multi-party conversation. Given these similarities, it is not surprising that individuals in such situations develop similar strategies.

"So", which is at the head of the second intonation unit (l. 2), serves to reintroduce a topic that has apparently been discussed before. This is indicated by the use of the demonstrative determiner "this", which presupposes common knowledge. It is a matter of speculation whether the knowledge derives from previous multi-party interaction or just from the title of the vlog ("On 'YouTube Live'"). The topic is an event that is central to many vloggers because they themselves, namely the active users of the website YouTube, are its subject. K80blog uses a context which is the sum of experience of all kinds of interaction that has taken place among the vlogger community about "YouTube Live". Thus, on the local, sequential plane, "so" links two units of talk. On a global level, or in this case a super-global level, it links back to discourse that might have taken place

in a different setting, through different media, where K80blog might have been just a passive listener etc. The vlogger makes an implicit assumption about her audience, namely that there is a collective experience that enables the viewers to understand what she is referring to.

The next line finally provides the element that usually comes right at the beginning of K80blog's videos: "and.. hello everyone it's Katie again". It is headed by another discourse marker: "and". The switch from the introduction of the topic in lines 1 and 2 to the prefabricated opening in line 3 is accompanied by a shift in her facial expression. It changes from a somewhat serious mode to an open friendly smile exactly when she utters "and" while slightly shifting her head from right to left (viewer's perspective). Thus, the shift from main topic talk to introductory intermission that is verbally enacted by a discourse marker is underscored multimodally by shifts in expression and posture. The lexical choice of "and" as shift facilitator (rather than "so") is interesting in that it seems to be completely devoid of its grammatical role (Schiffrin 2001) as a conjunction, as it occurs between two elements that are structurally not just different, but incommensurable.

Finally, in line 4 there is a shift back to the topic introduced in the first two lines. Likewise, the facial expression gradually goes back to a more serious mode. Again, she uses the discourse marker "and" to realize that shift, which, this time, fits better into the syntactic context.

2.5 Conclusion

In this chapter I have presented and analyzed various opening sequences in vlogs. I have shown how speakers approach the self-imposed task of producing a monologue on camera, and especially how they handle the transition from what happens before the vlog begins to the initial verbal expressions in the video. Speakers on vlogs use different strategies in their openings, which can be verbal or non-verbal. These strategies include linguistic means such as recognizable patterns, or the use of other media (computerized

voice), non-verbal distractive devices (showing glasses), which can trigger side-sequences, and video editing. Vlogs include interactive elements such as questions, greetings, and terms of address, which are borrowed from dialogic genres or conventionalized monologues. These phenomena are evidence of audience design, which underlies the production of most monologic media talk.

The use of these strategies results in a categorization of vlog openings. The main distinction is that between openings that follow a user's idiosyncratic pattern, and thus appear in the majority of that vlogger's videos, and those that do not rely on such a routine. This distinction is represented in Table 2.1 as 1) and 2). As this is not a structural distinction (openings from either category could have the exact same structure, e.g. greeting, term of address, self-identification), the study of several vlogs by the same user is required (represented as a – c).

Opening sequences	
1 a) sequence containing a combination of any of the following features: opening credits, greeting, term of address, self-identification, linguistic markers, cuts, distractive devices	→1 b) comparing several vlogs by the same vlogger reveals the class of prefabricated, established personal openings ↓
	1 c) once a user has established their personal opening phrase, they can use variations on it
2 vlog opening containing none of the above listed elements	

Table 2.1

The following are my conclusions regarding monologues: Assuming that dialogue is the default setting for spoken interaction – the mode in which language is acquired and most frequently used – monologue, in being reduced to the productive part of dialogic conversation and stripped of the receptive part, should be more difficult, or at least

different than face-to-face conversation. The suspension of turn-taking disallows all standard formulaic exchange patterns that involve sequencing, unless the speaker uses incomplete forms, or impersonates both speakers. Therefore, the observation that speakers employ alternative, adapted strategies in monologues is not surprising. The fixed introductory phrases presented are reminiscent of conversational openings from other genres (TV, radio, answering machine messages, etc.), which can also be very ritualistic. Editing, which constitutes the second phase of language production in vlogs, allows for strategies that would be very difficult to realize or unconventional in other monologic genres: cutting off parts of words or phrases, or moving elements to initial position that, when they were first uttered, served as brackets between two units. Monologues in vlogs also display elements that are employed in other monologic genres: greetings and terms of address occur (e.g. news reports), and so do boundary markers/frames (e.g. classroom talk, lectures). Furthermore, multimodal features such as shifts in gaze, facial expression and posture support the vloggers' verbal output.

Regarding openings, one can conclude that standard conversational openings such as question/answer- and summons/answer-pairs do not work in vlogs. On the other hand, alternate conversational elements such as terms of address and self-identification work and occur. What is interesting with regard to vlog openings in general is that some vloggers try to conventionalize them and turn them into routines. Others completely deny a need for an introductory phase, so they do not establish personalized routine formulae. That there are different strategies to begin a vlog, and which one a vlogger chooses, ultimately reflects pragmatic and sociolinguistic variation: speakers use different degrees of formality, choose different registers, address different imagined audiences, discuss different topics, have different aims, have different socio-economic backgrounds etc. Some vloggers possibly have introductory phases to prepare their viewers for topics discussed and create a sense of completeness of the vlog as a unit or to justify their discussing it in the first place (this also occurs at the ends of vlogs with phrases like "I just thought I'd share this with you"). Viewer feedback in forms of comments and ratings probably play a role when a vlogger designs their vlog content so

that it passes as tellable. Others, by leaving out a formal introduction, perhaps seek to create a sense of immediacy.

The fact that some users do have an introductory phase, even though negotiation of factors such as availability for interaction, identification, social status, and alignment of participants does not take place in vlogs, suggests some functions of vlog openings. It reflects that some vloggers have an imagined audience in mind that they are addressing, which might be helpful in the process of producing a monologue to a camera. It also acts as incentive for the audience to understand the vlog as part of an asymmetric, asynchronous interaction, inviting the viewers to respond via the communicative channels that YouTube offers. An interesting avenue for future research would be to compare the impact of vlogs with and without opening phase, in terms of the interaction they initiate. Such research could be conducted, for example, by comparing the number of views, comments, video responses and the ratings. However, since these features reflect back on the complete video, not just the opening sequence, it would be very difficult to isolate that variable.

Regarding the genre vlog as part of CMC, we can say the following: Why people engage in an activity that seems more demanding than other pastimes is a matter of speculation. An initial explanation is that these monologues reaching a potentially large audience are an ideal ground for representation of the self, creation of a persona, or generally the construction of identity. Many vloggers state, in metareferential comments about their hobby, that they like the interaction and the community. Clearly, the written comments users can make below a video and other features of the website are used extensively by vloggers and viewers to communicate and create identities. The fact that this happens online allows users to establish relationships with people that are physically so distant that they would generally not meet offline. As part of these relationships, vloggers and their viewers develop recurrent topics and thus establish common knowledge. Hence, in vlogs there is exophoric reference to both common world knowledge and very specific community based knowledge.

Further research on vlog openings is required to shed additional light on what functions these sequences fulfill. Such research should reveal more about the use of terms of address, which were only touched upon in passing here. Also the occurrence of silences and pauses at the beginning of vlogs should result in interesting insights into monologues (cf. silence/pauses in lectures). Observing vlog openings over time will also reveal whether there is a tendency to establish general conventions for the genre, such as fixed phrases that are used by all vloggers. The following chapter addresses a very similar desideratum of research in CMC and vlogs specifically. It focuses on endings of vlogs to reveal how this part of spontaneous monologues is structured, if there is a pattern, and if so, how it is realized.

3 The closing section of video blogs

3.1 Introduction

In this chapter, I present a descriptive analysis of the closing section of a type of computer-mediated, asynchronous monologue as found in video blogs. Closing sections of conversational data have received scholarly attention, most prominently by Schegloff and Sacks (1973). Their research demonstrates how the turn-taking mechanism is suspended to allow for a stopping of the interaction that is not perceived as a gap or silence in the conversation. The present chapter takes these insights on conversational closings as a starting point to compare them to the structure of vlog closings, identifying similar elements and their functions. Further, the vlog data is likened to answering machine messages (AMM), another genre of mediated, asynchronous monologic talk. The closings of AMMs often display a conversational character (Gold 1992), however, "the contribution may be terminated without the need for negotiating a closing" (Liddicoat 1994: 303). Based on these findings, this chapter hypothesizes that some practices in vlogs can be traced back to speakers' experience of conversational interaction, while the medium video influences the talk to such a degree that conversational conventions are being reconfigured in vlogs.

Consider an initial example to obtain a better understanding of the object of study:

Example 3.1

286	it is the mass of popular delusions,
287	and the madness of crowds.
288	anyway,
289	that's it for today,
290	take care everybody,
291	bye bye.

The speaker (male, in his sixties) finishes a topic (l. 286/287), and follows up with a boundary marker (l. 288) which indicates the transition between the vlog body and the

closing section. Thus initiated, the closing section progresses with a statement announcing that there are no further issues the speaker would like to discuss (l. 289) and two standard formulaic leave-taking expressions (l. 290/291). This vlog closing contains elements that could occur in conversational closings (e.g. transition, making sure there are no further mentionable topics, leave-taking formulae), however, in terms of intonation/prosody the monologic nature of the data becomes obvious in the lack of transition relevant places (pauses). Thus, we find both similarities as well as deviations from the conversational pattern in this excerpt.

The remainder of this chapter is structured as follows: the next section gives a summary of previous research and an outline of the methodology applied in the present chapter. A detailed analysis of vlog data, both quantitative and qualitative, represents the main section, leading to a discussion of the findings and the problems encountered.

3.2 Monologue, conversation and closing sections

Different types of monologues have been treated in the literature (Hasegawa 2010, soliloquy; Pascual 2006, legal monologues; Gold 1992, Dubin 1987, Liddicoat 1994 answering machine talk; Rendle-Short 2005, lectures; Mindt 2008, monologue/dialogue; Thompson 1994, cohesion in monologue; Jung 2006, academic monologues; Frobenius 2011, vlogs). Only few studies, however, are concerned with the structure of closing sections of these monologues; the most relevant of them will be introduced below.

Schegloff and Sacks (1973) provide an analysis of the structure of closing sections of conversations with a recognizable beginning and a recognizable end, in their case phone conversations. Conversations are characterized by turn-taking, which results in the sequentiality of speaker turns. The turn-taking mechanism of conversation, which enables the formulation of the phrase "no gap, no overlap" (Sacks et al. 1974: 700) as an ideal in conversation, explains why a failure to fill the next turn by either of the participants would be heard as silence, or a gap, in the conversation. Consequently, to properly close a conversation, both speakers must be relieved of the obligation to talk,

which can be achieved by suspending the turn-taking mechanism. This is done with a pre-closing: an adjacency pair that gives both speakers the chance to contribute so far "unmentioned mentionables" (303) to the conversation. These can be exchanges such as "okay"/"okay"; but a closing can also be initiated with various kinds of comments, e.g. "I gotta go" or "I don't want to tie up your phone". Once the pre-closing has been completed, conversationalists can close the conversation with a terminal exchange, most commonly "bye"/"bye" or similar standard wordings.

Closings of messages callers leave on answering machines can have several elements that are realized according to whether they are left on a business or a domestic answering machine (Liddicoat 1994). The terminating phase can include terminating components "goodbye", "bye"; closing components "okay", "alright"; acknowledgements "thank you"; and arrangements "see you on Saturday". The message can end after any of these elements as the lack of turn-taking precludes a negotiation of a closing that would offer both interlocutors the opportunity to contribute anything else worth mentioning. Liddicoat (1994: 304) further states that "it is necessary only to produce a move which is identifiably topic bounding and that suspension of topic is adequate to achieve a closing". His research (1994) shows that in institutional settings, AMMs predominantly end with an arrangement, or the next in frequency, an acknowledgement as last move. In the domestic setting, a terminating component is the most common option, followed by the arrangement category to close the message (302).

Gold (1991: 249) reports research on a smaller sample of messages left on a domestic answering machine. She identifies a further section in AMMs that can optionally follow the closing, that is, the post-script. In her data, the lexical items and prosody which are used in the closings of the messages sound typical of those used in closure to 'live' conversations, either on the phone or face-to-face. Speakers employ "preclosure gestures" (similar to what Liddicoat (1994) calls closing components), such as *okay*, which in some cases are followed by a pause as if to give an imaginary interlocutor the

opportunity to make another contribution. Thus, she describes behavior that clearly reflects conversational structures according to Schegloff and Sacks (1973).

Vlogs are similar to AMMs in that they represent mediated, asynchronous spoken language, with a recognizable beginning and a recognizable end. However, vlogs differ from AMMs in that they are audio-visual, and they can be viewed by a multitude of people, known and unknown to the vlogger. Vlog production involves two phases, the recording and the editing, and the vlogger knows that there will not be an interlocutor, unlike a caller who expects to reach a person to interact with. The similarities of the two genres warrant a comparison, though the various differences give reason to expect significant deviation from the regular pattern for AMMs.

3.3 The distribution of closing elements in vlogs

As indicated in the previous section, vlog closings deviate from the pattern of AMM closings. There is a larger number of elements that vlog closings regularly contain. The elements that are common to both genres are (following Liddicoat's terminology as outlined above):

- closing components ("yeah"),
- terminating components ("bye"),
- arrangements ("uhm we're just gonna keep in touch with that stuff") and
- acknowledgements ("just thank you for watching").

These functional slots are realized by a range of moves: closing components include statements that indicate that there are "no unmentioned mentionables" (Schegloff and Sacks 1973), e.g. "that's all I have to say". They also include a restatement of the reason for making the vlog (cf. Schegloff and Sacks' 1973 observation on similar moves in phone calls), such as "I really just want to say ha- happy St. Patrick's Day.". Arrangements sometimes have the form of a request for interaction "I wanna hear your opinions in the comments section," or they comment on a possible future encounter: "uhm we're just gonna keep in touch with that stuff.". Note that requests for interaction are strongly influenced by the affordances of the website, mostly in the sense that vloggers and

viewers allude to the various channels YouTube (and the internet in general) offers to contact other users. Example 3.2 contains a reference to the written comments underneath the video frame.

Beyond these elements that are created in the recording phase of the vlog making process, there are features that can be added when editing the footage. Vloggers frequently cut their material, insert captions, add a sequence similar to movie or TV show credits etc. This often results in a two part structure of vlog closings: the spoken part and the credits. Credits can even completely replace a spoken closing section.

Some elements prove difficult to assign to one of the above categories. Some vloggers include self-identification when ending their vlogs, e.g. "I'm DeoGenZ", which perhaps is inspired by radio show hosts' practices, or other mediated monologues, such as news reports or oral history stories published on the [murmur] project website[7]. The sample also contains closings in which the vlogger refers to the medium video: "am I even recording?". Self-identification might be part of a vlogger's routine when closing a vlog, and thus helps the speaker accomplish the task of ending a stretch of talk unilaterally, when, arguably, most other closings are negotiated bilaterally. Medium relevant talk seems to fulfill a similar function, that is, helps compensate for the lack of a familiar closing pattern.

The following table shows elements in AMM closings and the corresponding realizations in vlog closings.

---

[7] I would like to thank Dr. Ruth Page for pointing this out to me.

AMM	Vlog
closing component	closing component
	'no unmentioned mentionables'
	(re)stating reason for vlog
	reason for leaving
terminating component	terminating component
acknowledgement	acknowledgement
arrangement	arrangement
	request for interaction
	filler: self-ID, well-wishing
	editing: cuts, credits, captions

Table 3.1, Closing elements

The frequencies of these elements are as follows:

Out of 46 vlog closings in the sample:

32   contain a terminating component (22 of them in final position);

25   contain arrangements;

18   contain a no unmentioned mentionables marker;

18   contain an acknowledgement;

17   contain a closing component (e.g. "okay");

17   contain a request for interaction;

13   contain credits or captions;

12   contain self-identification (7 of which in the credits);

7   contain a well-wishing;

4   contain a reason for making the vlog;

3   contain a reason for ending the vlog.

These frequencies show clear tendencies regarding vloggers' concepts of vlog closings and what elements they contain. Three-quarters of the vlogs contain a terminating

component. Half the samples even have the terminating component in final position (except for two vlogs where the spoken part is followed by credits). Thus, it can be observed that a dominant number of vloggers adopt this element from conversation and assign it a similar function in a monologue, that is, act as the final element in a stretch of spoken language with a recognizable ending. Twenty-three vlog closings of those 32 that contain a terminating component also contain at least one of the four realizations of the closing component function. This would imply that there are vestiges of the conversational closing sequence present in vlogs, similar to that of AMMs.

Of the vlog sample, 30 closings contain an arrangement, that is a reference to some future communication, be it through another vlog, face-to-face interaction, or some other form of mediated communication. This is a reflection of the predominantly social character of vlogs and its function as a tool to connect with other people. While in AMMs, arrangements can be the final move, especially in institutional settings, vlogs hardly ever close with them: there is only one instance of this in the sample. Liddicoat (1994) explains that arrangements, as topic bounding moves, lend themselves to end talk without need for a negotiated closing. The difference between AMMs and vlogs is likely due to the expectations of the speaker. A caller who leaves a message on an answering machine usually expects to engage in conversation rather than monologue when making the call; so one goal of the message the caller then leaves will often be to indicate interest in interaction at a later time. A vlogger who knows they will be talking into a camera does not expect to engage in conversation, and they are not talking to a specific person or company they could have specific arrangements with. The arrangements remain vague and might therefore not qualify as final moves.

Another striking feature of vlogs, which sets them apart from AMMs and most other monologues, is the presence of editorial features. In the sample, one fourth of the closings are complemented by written credits, which is an element possibly borrowed from movies and TV genres. Of course, such editing requires certain skills, which might not be available to all vloggers. The written content of these credits predominantly

comes from three categories: self-ID, acknowledgement, request for interaction. In all 13 cases, the credits are in final position.

In sum, vlog closings can be classified as containing fairly typical elements or patterns on the one hand, or idiosyncratic practices on the other. While the present chapter attempts to identify tendencies for patterns in the data, there is no claim that it is exhaustive. Certainly, one could find vlogs containing elements that this discussion does not cover or that do not fit into the categorization developed here.

3.4 The elements of vlog closings and their functions – close analysis

Some of the closings in the sample appear to be fairly similar in structure, containing elements familiar from closings of other genres, most notably AMMs. These elements fulfill functions similar to those that they have in these other genres. In contrast to that, other closings simply lack these features or redress the balance with other elements (often genre related). The following analyses delimit the spectrum covered in the sample, and they illustrate how the functions of the contained elements are adjusted to the monologue situation.

The first example features a man in his twenties. The first five lines represent the end of a story including constructed dialogue before, starting in line 6, he wraps up the vlog, and finally, in line 20 he initiates the core closing:

Example 3.2

1	"I had an issue yesterday with my card",
2	and if it happens again today,=
3	=I'm gonna be very upset.
4	"they were like,"
5	"of course Mister Wolf." (2.5)
6	and that's it...,
7	it's nice to be back?

8   please be expecting videos from me,=

9   =that, .h

10   aren't so much my personal history of the last year,

11   but you know,

12   a little more entertaining.=

13   =but I just needed to fill you guys in,

14   on where I've been,

15   and,

16   what I'm doing…

17   and,

18   uhm,

19   looking forward to seeing you again,

20   alright,

21   I'll talk to you later.

22   bye bye.

The vlogger progresses from the vlog body to a mediating phase, which leads to the closing. A pause of over two seconds (l. 5) indicates the end of the preceding narrative and positions the following talk as a separate unit. "and that's it..," (l. 6) signals that the speaker has related everything he wants to say, and thus represents an instantiation of the 'no unmentioned mentionables' category. While in conversation this type of move has the function of offering the interlocutor an opportunity to make a further contribution, this is certainly not the case here, for there is no interlocutor. Although there is a pause of over one second, as if to create a transition relevant place, this line has an organizational function within the monologue, that is, it prepares the viewer to expect the end of the vlog, rather than demanding a speaker shift (cf. Coulthard and Montgomery 1981 for organizational features and the structure of monologues). Interestingly, the video shows the speaker turning his head about 90° to his left during this pause, apparently searching for something (cf. screenshot). This gaze shift is owed to the adaptation of the pre-closing move to the monologue situation, in that it

compensates for a speaker shift and gives the vlogger time to follow up with more talk himself.

Figure 3.1, Pause in line 6

Lines 7-19 can be subsumed under the term relationship management. The speaker expresses his positive feelings regarding his vlogging activity (l. 7), he promises more communication via vlogs (l. 8), and he qualifies the content of his vlogs (l. 9 through 16). Contained in that is the reason for making the vlog in lines 13 through 16: "=but I just needed to fill you guys in, on where I've been, and, what I'm doing…," which is reminiscent of AMMs. Line 19, "looking forward to seeing you again," is of course semantically odd, since there has not been and will not be a physical encounter between him and his viewers. However, since it constitutes a fixed phrase, it nevertheless works as reference to a possible next virtual encounter, thus falls in the arrangement category. Compared to the initial example, where there does not seem to be any relationship management in the closing section, this stretch from line 7 to line 19 is fairly long. Whether that difference is due to speaker personality, topic, or both, is a matter of speculation. But the function that it fulfills seems to involve providing for a smooth transition to the closing, which in conversation would be realized through negotiation.

After the mediating phase is finished, the final moves are accomplished: "alright," (l. 20) acts as a closing component, which is then immediately followed by two terminating components in lines 21 and 22: "I'll talk to you later. bye bye.". All three elements are standard devices which might occur in telephone conversations.

Generally speaking, this example does not offer anything in the way of surprise regarding the structure, the elements employed, and the functions they fulfill. Clearly, the speaker borrows conventions from conversation and AMMs, some of which are adapted to fit the genre. He apparently favors a smooth transition over an abrupt ending, which results in a vlog section that emphasizes the implicit goal to inspire interaction with the viewers.

The following examples illustrate other vloggers' choices when accomplishing closings, which are not all as straight forward as the previous one. The next example (3.3) was recorded by a woman in her late thirties. It is structurally interesting in that the terminal component is followed by a coda to a previous topic in the vlog. Thus, the leave-taking formula does not constitute the final move.

Example 3.3

185	yeah,
186	she ate her cheese and took off. (1.5)
187	so I might do a a vlog before I leave tomorrow so,…
188	that's about all,=
189	=I really just want to say ha- happy St. Patrick's Day.
190	uhm,
191	I almost wore one of my pink tie dye shirts today,
192	then I realized what day it was,
193	and I.. this one was clean so,
194	I had to get on the green though….
195	even Dad wore green today.
196	and I don't even think he did it.. because of St. Patrick's Day,
197	I think it just happened….
198	.h a:nd it was 65 degrees today and I opened up my sun roof in the car?
199	"it was awesome,"
200	"awesome I say,"

201   all right.

202   so.

203   bye:.

204   I gotta take this thing off my head,=

205   =my head is starting to sweat.

Lines 185 and 186 represent the closing of a topic, namely the speaker's cat, which also appears in the vlog several times. The transition between topic and closing is accomplished with the help of an extended pause during which the vlogger chews food (compare the gaze shift during the pause in the previous example). The closing is first initiated in lines 187 through 189, when the speaker refers to the next possible vlog encounter, states that she has made all relevant contributions, and restates the reason for making the vlog, here in the form of well-wishing. Following these three elements that are all common elements in closings (both monologic and conversational), she reopens topic talk. The first topic is related to the well-wishing in line 189, thus it can be assumed that her narrative-like comment on her choice of clothing is triggered by her "happy St. Patrick's Day."-wish. There is no detectable connection to the next topic (weather/her car, in lines 198 through 200), except that she is referring to events that happened the same day. The actual closing of the vlog is subsequently taken up again: lines 201-203 contain two standard pre-closings and a terminal component, each representing complete intonation units. Thus, the topic talk inserted into the closing (l. 190-200) is comparable to a conversational closing sequence which is aborted when one participant takes up the opportunity to make a further contribution. While the quick succession of turns in conversational closings would make such a lengthy digression subject to negotiation, the speaker of this vlog can proceed comfortably with pauses and hesitation markers without any risk of losing the turn.

The two lines following the terminal component refer to a large green hat she is wearing for St. Patrick's Day (l. 204-205). The hat actually represents the very first topic of the vlog (the vlogger starts the vlog with a sing-songy "happy St. Patrick's Day"). Along with the whole theme it symbolizes, the hat topic provides a frame for the monologue. In the

closing section, it adopts a particular function: While the "bye:" in line 203 provides a conventional closing, signaling the end of a stretch of spoken language, it is clearly relieved of part of that function here. The position of final element is occupied by the hat-coda. Thus, structurally, the "bye" loses its function; however, semantically it is still necessary since the hat topic is not a conventional way of ending talk. The vlogger can choose to alter standard functions of elements for two reasons: one, there is no negotiation possible, so it is up to her to assign the relevant meanings and the viewers have to accept them; two, the actual closing is effected by the final cut, an editorial feature – therefore, whatever happens last has the function of closing the vlog. The vlogger might have cut the video off right after the "bye:.", thereby preserving the conventional form-function unit. However, her decision illustrates that in vlogs, conventions can be challenged without a significant loss of communicative force.

Apparently, vlogs constitute a genre that fosters experimentation, or at least vloggers do not necessarily adhere to conversational constraints. While in this example, editing is only minimal (there really is only the final cut, but none throughout the vlog), and thus the coda is produced as such on the spot while recording, in other vlog closings, a coda is later added using another cut. It seems that when adding a coda in the editing process, it involves a different quality of effort to violate established conventions, as it is not done under the time pressure of online speech production. The following short excerpt shows an edited coda:

Example 3.4

82	alright guys,
83	until next time.
84	it's Christoph Von,
85	and <I SEE YOU.. WHEN I SEE YOU>.
86	CUT
87	oh master,

After a standard closing including closing component (l. 82), self-identification (l. 84) and terminal component (85), there is a cut (l. 86), followed by a coda (l. 87). It consists of a direct quote from the footage, a short clip taken from the vlog body. The self-identification ("it's Christoph Von") is placed in the functional slot before the terminal component, thus it functions as a closing component, along with "alright guys" (l. 82). In contrast to the previous example, this vlog closing represents a discontinuous unit. This impression is enhanced by the difference in body posture, pitch, volume and content between the closing up to and including the terminal component on the one hand, and the coda on the other hand. Editing thus plays a significant role in the production of vlogs, offering vloggers ways to structure their videos that can be used to challenge familiar conversational patterns.

The following example also illustrates the use of editing in vlog closings. A man in his twenties performs what appears to be his standard closing twice in the same vlog:

Example 3.5

263   uhm look forward to.. some of that,

264   and other than that,

265   I got nothing more to say to you guys,

266   except watch out for my new let's play in the near future,

267   and the finales of Mass Effect and Realms of the Haunting,

268   a:nd I guess this is gonna be Joseph Vissarionovich Stalin,

269   signing out for now.

270   BLEND

271   one quick thing I forgot to mention,

272   note about my.. side project?

- - - - lines 273-293 not reproduced here - - - -

294   anyone can really do it,

295   so.

296   if you're intereste- interested,

297   please message me,

298	and we'll talk about the details THERE.
299	so.
300	other than that,
301	uh you guys have a fantastic evening,
302	day,
303	morning,
304	wherever it is you happen to live.
305	and this is gonna be Joseph Vissarionovich Stalin,
306	signing off,
307	and I'll see you guys later.

In lines 264-268 of this example, we find a fluently delivered closing containing a statement that all relevant contributions have been made (l. 264-265), immediately followed by a reference to future videos (l. 266-267), self-identification (in form of an obviously adopted name; l. 268) and a terminal component, apparently borrowed from radio broadcast, where hosts 'signed off the air'. A very similar passage appears in lines 300-307, where many of the elements are repeated using almost the same wording. One notable difference is the substitution of the 'next vlog encounter' statement (l. 266-267) for a well-wishing (l. 301-304); another is the additional terminal component "and I'll see you guys later" (l. 307) supplementing the phrase from radio broadcasts. The striking similarity of these two passages suggests that the vlogger has developed this as his routine closing for vlogs, thereby avoiding the need to find an original strategy each time he makes a vlog.

The double occurrence of the vlogger's signature closing requires some discussion, especially in the light of what has so far been established about the role of editing. Of course, if a closing section occurs with a considerable amount of talk following it, including another closing section, this means it has retrospectively been assigned another function. Initially, the structure of the example seems simple: the speaker ends the vlog, realizes he has another contribution to make and continues, and finally he closes the vlog again. On second thought, it becomes obvious that the vlogger made an

interesting decision in the editorial process. Instead of cutting out the first closing section and blending the vlog body over into the "note" (l. 272) immediately, he left it in. Note also his choice not to use an abrupt cut, but a blend of both the audio and video track of the vlog for a smooth transition (l. 270). If lines 264-269 do not serve to close the vlog, what needs to be determined is the function(s) they are assigned instead.

Clearly, their closing character is not lost entirely, but rather than representing a final vlog closing, the first routine closing in the example seems to accomplish a section or topic closing. This view is supported by the following line "one quick thing I forgot to mention," (l. 271), which is a standard reopener or topic shifter. But rather than only acting as a text-structuring device, the passage introduces a shift in footing (Goffman 1981; also cf. Clark and Carlson 1992 and Morley 2006 for audience design). While the main vlog is over, there is a further topic that is potentially relevant for only part of the audience. Lines 296-298 refer to another communicative channel, (email/YouTube message) that is to be used by those viewers interested in the project outlined after the first closing. Thus, there is a shift in audience design, for the audience is split into those interested in the "side project" (l. 272) who become the addressees and those not interested who become overhearers.

This example further illustrates the significant impact the editing stage has on mediated asynchronous communication. Editing options are readily adopted as tools to structure talk and to assign participatory status. The next example shows less subtle use of editing. Here, a man in his early thirties places cuts after every intonation unit, and the vlog ends with a credits section.

Example 3.6

	102	CUT
	103	they're good people down there,
	104	CUT
	105	shout out to Ricky (O)?

106	CUT (2.0)	
	{looks at smart phone, wipes it clean, then holds it into camera during pause}	
107	yeah that's my phone,	
108	CUT	
109	I think it's pretty cool.	
110	CUT	
111	I can do anything with this phone.	
112	CUT	
113	(bet) I can call your grandma,	
114	CUT	
115	am I even recording? (1.0)	
	{closely approaches camera before and during pause}	
116	CUT	
117	peace out,	
118	CUT	
119	{written} Ridiculous Productions	{music}
120	{written} Edited by Rick	{music}
121	{written} Help me keep vlogin by visiting www.ridicGraphics.com	{music}
122	{written} Thank you for watching	{music}

Lines 102-118 represent the spoken part of the vlog closing in this transcript, lines 119-122 the credits. The spoken part is delivered in a playful tone, and the multiple cuts create the impression of fast-paced talk. There are no hesitation markers or filled pauses, suggesting that the vlogger systematically deleted them in the editing process. Two pauses (l. 106; l. 115) are both filled with multimodal elements of communication: the first is a gaze shift towards and placing of an object in the viewers' sight to direct attention to it; the second is play acting for humorous effect through creating a feeling of awkwardness with an exaggerated close-up shot of his face (cf. Figure 3.2).

Figure 3.2, close up, line 115

As far as standard closing elements of the types so far discussed are concerned, only one can be detected: "peace out" (l. 117) is a colloquial, but recognizable standard terminal component. There is no lead-up to it, that is, there is no element that could be interpreted as a marker indicating the initiation of a closing. There is no reminiscence of a negotiation of the closing section, not even an element announcing it, making this the closing with the least resemblance to conversation and AMM of the closings inspected so far.

The section following the spoken part, the credits, contains traces of several media/genres. The credit section is a sequence of written captions, with music playing simultaneously. It is a convention borrowed from TV shows/films, which are in most cases followed by a listing of people, places, works of art etc. involved in the production process. "Ridiculous Productions" (l. 119) appears to be the name of the company that produced the vlog (it is unclear whether this company exists or whether it is imaginary), which alludes to a media-professional status. "Edited by Rick" (l. 120) is a self-reference in third person (his profile information states his name as "RidicRick"). Lines 121 and 122 deviate from the TV broadcast theme and resemble vlog talk: the vlogger asks viewers to visit his website, which is a request for (online) interaction; finally he makes an acknowledgement: "Thank you for watching". Both elements recur in vlog closings, as discussed above, and both could potentially fill a functional slot before the terminal component, announcing the closing. In this example, the vlogger chose to move them not only after the terminal component, but also into a different, i.e. written, mode.

Perhaps it was the vlogger's goal to produce a fairly entertaining spoken part, and therefore to leave out elements that would have slowed it down. These were then included in different form in the separate credits. This kind of mode-shift is only available in mediated audio-visual communication, such as TV genres, but not in face-to-face interaction. (Of course, one could imagine scenarios where interlocutors close the spoken part of their communication and switch to writing notes; or, more likely: people sitting in the same room interacting via online chat. But this would always involve a switch in communicative channel, which is not the case in the example under discussion.)

One vlog in my sample is interesting for its complete lack of a recognizable closing sequence. The whole vlog is shot outdoors, while riding on a motorbike through city streets. The footage shows the streets, not the vlogger. The vlogger is male, in his early twenties and his native language is French.

Example 3.7

279	the only place where you can have fun is between the lights,
280	and there must be like three or four turns in the whole city and that's it.
281	which which uh makes me a bad driver actually,
282	a bad rider,
283	I'm not very good.
284	I could be better,
285	at riding,
286	but I'm not.
287	{accelerates, 15 seconds of footage of the streets}

The passage does not contain any of the elements thus far identified as pertaining to closings, neither regular spoken nor edited ones. It is possible that the vlogger produced a closing that was then cut off in the editing process, which cannot be determined. Either way, the vlogger uploaded the vlog to YouTube without a closing, indicating that he considers it complete. On the one hand, one could argue that the special context of

the recording allows for an interpretation of the last 15 seconds that show streets from the perspective of a motorbike as some sort of closing (perhaps also reminiscent of movies that end in a character's moving along on his or her way). On the other hand, this is by no means a standard closing of any spoken interaction, nor to vlogs. While this closing might be an exception, it still illustrates that there are simply no conventions of spoken interaction that are strong enough to govern all vlog closings. Most vloggers show awareness of conventions of spoken language, but if someone does not, there do not seem to be viewer comments remarking on this behavior. In the video description of the vlog Example 3.7 is taken from, the vlogger mentions time constraints: "It's been edited very roughly to fit under the 10-minute youtube law. Could have been smoother but I couldn't be bothered to care about it". It is hard to believe that the vlogger had to give up a proper closing section because the video was too long; he could have easily shortened the last 15 seconds of simply riding in favor of two more intonation units closing the talk. The "couldn't be bothered" comment suggests a better explanation: no convention requires a vlogger to include a closing, so if it is too much effort for whatever reason, it might as well be dropped.

A similar example features a woman in her thirties. While her editing is fairly elaborate, involving a number of cuts, there is also no recognizable closing element.

Example 3.8

95	CUT
96	{plays song, sings to it, dances to it}
97	CUT
98	but I like .. sentimental music too?
99	CUT
100	but I like tons of music.
101	CUT (3.0)
102	this was pretty pointless huh.

The final topic, music, here demonstrated and commented on in lines 95-100, is followed by a cut, a long pause (l. 101) and finally by an evaluating comment. "this was pretty pointless huh." (l. 102) could refer either to the previously treated topic of music, or to the vlog as a whole. At best, this assessment could be classified as showing awareness of the expectation for vlogs to 'have a point', and mentioning the lack of it might allude to the 'stating reason for vlog' category. It is worth noting that the "huh" tag turns the line into a question. Both its placement as the final line and the fact that this is a monologue preclude any response, and this further illustrates the inherently interactive character of spoken language, even in monologues. Possibly, the vlogger uses slightly self-deprecating humor to make up for the lack of a negotiated closing section. Still, this closing lacks any recognizable conventions and is therefore to be situated at one end of the spectrum (along with Example 3.7), as opposed to the ones previously discussed, which borrow conventions from various genres.

The differences in vlog closings may be due to the kinds of sub-genres that exist in the genre ecology of vlogs. A systematic exploration of this issue would require research into the different types of sub-genres that exist, with a view to determine whether there are consistent patterns.

3.5 Conclusion

The present chapter reports research on the closing sections of monologues, in particular vlogs. Like openings, closings are structural elements that have to be 'achieved'. While this achievement entails co-construction and negotiation via turn-taking in (face-to-face) conversation, vloggers are alone in managing the task they set for themselves.

This study has shown that vloggers borrow conventions and phrases from other genres of talk, both dialogic and monologic, synchronous and asynchronous: the analysis presented elements that occur in conversation, AMMs, as well as radio and TV broadcasts. Most notably, many vloggers adapt a conversational closing to monologues:

a pre-closing and a terminal exchange are reduced to one or several contributions each instead of one adjacency pair each. In most cases, this is also done with standard lexical items and phrases, which makes closing sections easy to recognize. The lack of negotiation and the resulting brevity of the section are optionally compensated by the insertion of fillers, or further elements that could occupy the closing component slot before the final move. For those vlogs that demonstrate a familiar structure, AMMs represent comparable data.

Analysis has further demonstrated that while many vloggers adhere to a familiar pattern, there is a whole spectrum of vlog closings that display alterations or even a complete lack of a recognizable pattern. Often (but not always), editing plays a role in the design of closings. In this second stage of the production of a vlog, crucial decisions as to what remains in the vlog, and where the elements are placed, are made. While the production of talk happens under time constraints and is a highly developed and frequently used skill, the editing process can take as long as necessary and probably requires a certain amount of attention. Therefore, one would assume that editorial decisions are made with a greater level of awareness. In the sample, editing is used for various effects: speakers can shift footing, change the function of standard elements by assigning them certain slots, shift from spoken to written mode and change the pace of talk. This greater level of awareness suggests that vlog closings can be the site of a stage-like performance. Vlogs at the high end of the spectrum display originality, while those at the low end retain familiar patterns.

If, as the examples cited have shown, a whole range of vlog closings are acceptable, from those resembling conversational interaction to abrupt cut-offs, then what needs explaining is not the absence, but the presence of elaborate closing sections in vlogs. If the viewer has to accept the vlogger's decisions in any case, without any chance of negotiation or influence, then why do vloggers go to the trouble of producing a closing? An obvious answer, which probably applies to many vloggers, is that people's conversational experience is an established standard that is sufficiently habitual to spill over into other genres. Further, a mutually achieved closing is a highly interactive part of

a conversation. If the purpose of a vlog is to connect with the viewers, then the inclusion of this part may be considered a helpful feature.

Vloggers in the data set most commonly adopted terminal components and references to the next encounter (arrangements) as part of their closing section. These are elements typical of conversation as well as all other formats that figure as sources of vlog closing strategies. Thus, they seem to be the choices representing the most familiar elements. This could be due to reasons of efficacy in production as well as viewer accommodation. Retrieving those elements that occur frequently in interaction in certain slots facilitates online speech production. Likewise, it facilitates comprehension, as these common elements represent expected language use. Thus, the use of familiar elements can be characterized as a form of involvement (in Tannen's 1989 sense).

The data surveyed have begun to yield a picture of how the new media naturally incorporate strategies from everyday conversation and how they expropriate strategies from other genres. Both sorts of strategies in part retain their old forms and functions, but also adapt to their new environment. The borrowings from old media into newer media forms such as online video, whose format closely resembles that of movies/TV shows but which are produced by non-professionals for a specific online audience, deserve more attention. This appropriation of old media text is not only interesting from a linguistic perspective, but for any discipline concerned with new and traditional media.

The vlogs discussed in this chapter were all recorded at approximately the same time, thus they represent examples of the general conception of vlogs at a certain point in time. From this data set it is impossible to say if there is any standardization of vlog closing conventions taking place at all, or if closings will continue to appear in various forms. Conducting a study that takes the development of genre conventions over time into account certainly represents a fruitful direction for future research. Likewise, sociolinguistic studies might reveal how factors like age and sex influence vloggers' use of conventions in vlogs. Monologue closings in various settings require further research to reveal more about people's ability to adapt their predominantly dialogic experience to the demanding task of producing talk independent of interlocutors. The comparison

of vlogs to TED Talk closings presented later in this thesis shows the major impact a present audience can have on closing structures (see chapter 8).

4 Audience design in monologues: how vloggers involve their viewers[8]

4.1 Introduction

The aim of this chapter is to investigate speakers' strategies to create involvement with their audiences through monologues. Researching these strategies seems particularly relevant in the context of the re-evaluation of the concept of *Participation Frameworks* regarding its application to traditional and new media, the context in which the research presented in this chapter originated. The relevance to this topic is tied to the functions vlogging fulfills for its users: vloggers implicitly and explicitly state that the main purpose of their vlogging activity is to communicate with other people through the channels offered on YouTube. While many of them ask for feedback in the form of written comments and videos, some recount how they have been able to make new friends through vlogging, people they meet online as well as offline. The use of the term *community*, both by scholars (e.g. Lange 2007) and YouTubers themselves, to describe active participants on the website, indicates that exchange among the members is essential. Likewise, high view counts and ratings of their videos are frequently thematized by vloggers, leading to the fixed phrase "please comment, rate and subscribe", which frequently occurs in my dataset. This, too, reflects a desire for active participation by the audience.

As Goffman (1979: 12, 1981: 138) states regarding similar kinds of monologue, that is, TV and radio talk, they are "not addressed to a massed but visible audience off the stage, but to imagined recipients". In this chapter, I illustrate how talk in vlogs reflects this lack of an immediately present audience, or more precisely, how speakers adapt to a speech situation where there is not even minimal immediate feedback and the speaker has to address imaginary viewers. Goffman's (1979, 1981) *participation framework*, Clark and Carlson's (1982) *audience design* and Bell's (1984) *audience design*

are central constructs in the present study. The phenomena examined in this chapter include, for example, terms of address, directives/directed language, questions, voicing the audience, and categorization. These are discussed in terms of how they contribute to establishing or altering established participant statuses and thus reflect audience design.

This part of my study is based on Part 1 and 2 of my vlog corpus. Part 2, which is a random collection of vlogs, includes examples of sub genres (e.g. 'unboxings', 'hauls' – where vloggers present their newly purchased video games or clothes respectively), or vlogs by well-established vloggers (whose view counts are relatively high) versus vlogs by first time vloggers. I assume that popularity and topic, among other things, have an influence on how the vlogger imagines his or her audience to be, and consequently how he or she addresses and involves them. An experienced vlogger with a high view count and many viewer comments might possibly use different involvement strategies than a vlogger with less experience.

4.2 Audience design and audience involvement

Several studies have produced varied notions of audience, speaker and hearer roles, role assignment and audience design. Goffman (1979) distinguishes several roles that participants in conversations assume. Speaker roles are *animator*, *author* and *principal*; hearer roles are *addressee*, *bystander*, *overhearer*, and *eavesdropper*. Clark and Carlson (1982) present a similar model to develop the notion of *audience design*. The hearer roles they propose roughly correspond to Goffman's *participants* (encompassing *addressee* and *side-participants*), and *overhearer*. Addressee and bystander/side-participant are ratified participant roles, whereas overhearer and eavesdropper are unratified participant roles.

Clark and Carlson's (1982) notion of audience design is crucial to understanding that receiving a monologue, even if it is as remote and asynchronous as in a vlog, is not merely a passive role. As vlog viewers process the input, they must actively understand

what hearer role they are being assigned: "The speaker must designate which hearers are to take which roles. It is essential that the speaker and participants, the speaker and addressees, mutually recognize which hearers are being designated as participants and which as addressees" (1982: 222). If active recognition of the hearer's participant role is a prerequisite for a successful 'speech act' (their terminology), then a shift in participant status requires that recognition. In consequence, when vloggers (re)assign participant roles, the audience is, if they choose to follow and understand the vlogger, involved via their cognitive efforts. Such a cognition-based understanding of involvement can be derived from the works of, e.g. Tannen (2007 (1989)), Chafe (1985), or Gumperz (1981).

Role assignment can be achieved by a number of devices. Clark and Carlson (1982: 346) list five categories: 'physical arrangement', 'conversational history', 'gestures', 'manner of speaking', and 'linguistic content'. The next section (4.3) explores these categories in more detail and discusses their applicability to vlogs. The structure of the analysis section is oriented towards these categories, too, as reflected in the sub-section headings.

Bell's (1984) notion of audience design also differentiates between hearer roles: addressee and auditor (corresponding to addressee and side-participant in Clark and Carlson's model) are ratified, *overhearer* is not. Moreover, Bell's (1984: 186) model introduces the concept of *referee design*: "Referees are third persons not physically present at an interaction, but possessing such salience for a speaker that they influence speech even in their absence". As an example of referee design, he cites the use of French by a bilingual in Montreal to a monolingual English speaker. He discusses the applicability of audience design and referee design to mass media, arguing that there are traces of both to be found in mass mediated communication. Online video certainly does not represent a classical mass medium. However, as an audio-visual medium it clearly resembles TV talk, in that speakers may talk into a camera. Most vloggers, however, never reach the large numbers of viewers that make up a TV audience, and their express desire for personal communication with the audience is not typical of mass media. Vlogs therefore are likely to bear traces of referee design, in that "the media

audience is, for the communicator, unspecific: an image, an ideal, a perceived class of persons" (Bell 1984: 191f.). At the same time, if a vlogger has a number of subscribers he or she has been in contact with, the vlogger has a sense of who makes up at least part of the audience. Thus, as most vlogs bear traces of previous communication, for example when vloggers cite from viewers' comments, vlogs can also be designed for a concrete audience. Vlogs could therefore be classified as a genre situated in a meso-medium (Feldmann and Zerdick 2004). Meso-media "address small groups between a hundred and a hundred thousand participants" (2004: 23).

Goodwin (1986) analyzes the active co-participation of the audience in face-to-face storytelling, showing that expertise and competence, the interpretation of the story and the participants' alignment is constantly being negotiated. These findings are difficult to apply to the vlog situation: the visual and the audio channel are not only asynchronous, they are also unidirectional – the vlogger cannot see or hear the audience, nor can members of the audience see or hear each other (unless they watch a vlog together in the same room). In face-to-face communication, much of the negotiation is done via multimodal communication: gaze, gestures, nodding, shrugs etc. In vlogs, however, subsequent communication can take place via a different channel, namely written comments that vlogger and audience can leave underneath the video frame. As an analysis of these written comments in terms of how they can replace multimodal features used in talk would go beyond the scope of the present chapter, I only discuss audience structure and participant status based on the video footage. The participation of the vlog audience through the written channel is given detailed treatment in chapter 7.

4.3 Role assignment in vlogs

Interlocutors use a range of devices to designate hearer roles. This section adapts a model outlining these strategies (Clark and Carlson 1982) to the vlog environment with its particular properties.

4.3.1 Physical arrangement

The physical arrangement of the conversation even determines who is eligible to even have a participation status: "The people must be near each other relative to the space available – not separated by obvious physical or psychological barriers, and accessible to each other auditorily and visually" (Clark and Carlson 1982: 223). The physical space that the interlocutors are situated in, however, is not the space where communication takes place. Instead, the communication is moved to a digital or virtual space, namely the website, which has different properties. The following example illustrates vloggers' creative use of this space:

Example 4.1

      174   enjoy it,

      175   (I acted) stupid for you guys,

      176   so,

      177   leave me a comment BELO:W. {points down}

      178   say how you liked my video,

      179   say how you liked this game,

      180   if you picked it up or not,

      181   whatever,

Figure 4.1, "leave me a comment BELO:W.", line 177

The vlogger makes requests for interaction in the form of imperative constructions (l. 174, 177-179). Line 177 is accompanied by a medium-specific gesture: the vlogger points downwards with a large, sweeping motion, verbally identifying the referent of his deictic gesture (see Fig 4.1). The comments section he is pointing at is not located in his physical surroundings, but on the website (cf. Frobenius 2013 for a discussion of pointing gestures in vlogs). Simultaneously, "BELO:W" is pronounced with increased volume and a drawn out second syllable. Both of these multimodal features, prosody and gesture, underscore the vlogger's sincerity of his interest in the viewers. But more importantly, he demonstrates his ability to imagine the viewer's perspective as recipient of a vlog embedded in a website with a certain design.

The communication is mediated with the help of a camera, which, in many cases, is static and therefore has a static focal range. This restricts vloggers' movements as they risk being out of the frame, which in face-to-face conversation could easily be amended by the addressees' gaze shift. Pointing gestures in face-to-face communication are executed in a three-dimensional, physical space. Their referents are located in that same space, be they concrete or abstract. Vlog pointing gestures originate in a three-dimensional space, but they can be directed at referents outside that space. In other words, a vlogger filming him- or herself in any location can employ pointing gestures that are meaningless in that same location: here, the vlogger points at the floor of his apartment, which, however is not the referent. The referent can be found in the context of the website, a two-dimensional space that is created by uploading the video. Pointing gestures can be restricted in size, depending on the vlogger's proximity to the camera and the resulting picture frame. This is a crucial difference between talk in vlogs and face-to-face interaction. As shown in the example above, the creative use of gestures can be used to emphasize a desire for interaction. In an ongoing conversation, this would not be necessary, since a mere question, for example, expresses the expectation of an answer. In conversation, a question can be used as the first pair part of an adjacency pair to make relevant a contribution by an interlocutor, and thereby assigning participant status. This resource is not available in vlogs, thus gestures are employed together with the design of the website to achieve a similar function.

Gaze shift in vlogs cannot be used to assign different roles to different vlog viewers, as vloggers look into a camera (or at some other object in their surroundings), but have no access to their viewers' gaze. Likewise, viewers cannot signal their understanding of the role assignment, and consequently, there is no acceptance or challenge of it either. However, as is illustrated in the next example, gaze does play a role if there are people present in the vlogger's surroundings.

4.3.2 Gaze/gesture

The following lines are the opening of a vlog in which the vlogger addresses both her non-present asynchronous vlog audience and a person present with her in the room.

Example 4.2

    1   well hi this is the first vlog [I'm doing on my:: computer.]

            [{gaze shift to speaker's right; then back}]

    2   John thinks I'm talking to him?

    3   but in fact I'm no:t,

    4   I'm talking to you,

    5   to everyone in the world,

    6   how are you. (1.25)

    7   this is my first vlog on the computer.

This vlog opening contains a sequence of four intonation units (l. 2-5) in which the vlogger clearly and directly states who she is addressing; apparently she feels the need to clarify, which indicates her awareness of a potentially ambiguous participation framework. The first sign of this is the gaze shift that accompanies the second half of line 1 (cf. Fig 4.2, 4.3, 4.4). Supposedly, this is a reaction to some act of non-verbal communication by John, the other person in the room, e.g. his signaling attention as if he thought he was being addressed. After this, her gaze shifts back into the camera, confirming the vlog viewers' status as the direct addressees. While her meta-comment

excludes John from the group of addressees, which is underscored by her reference to him in third person (l. 2), this information is directed at him as much as at the vlog viewers. It explains to John why she is talking, and why his attention is not required. To the audience, it explains the vlogger's gaze shift. In line 7, the speaker repeats what she was saying before clarifying participant statuses. "this is my first vlog on the computer." marks the completion of her adopting and assigning footings.

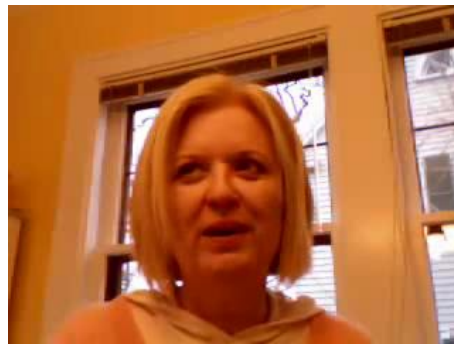Figure 4.2, "well hi this is the first vlog"

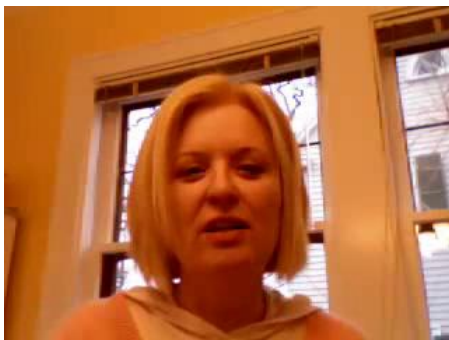Figure 4.3, "I'm doing on my:: computer."

Figure 4.4, "John thinks I'm talking to him?"

This example illustrates that in this passage, the speaker treats the interaction with another person in the room as separate from her vlog recording. (See Norris 2004 for a model that places higher level action, such as the two interactions the vlogger is participating in, on a cline between foreground and background. Here the vlogger explicitly deals with placing the different sites of interaction on that cline.) When this separation is at stake, the speaker a) addresses the issue and clearly assigns participant roles through propositional content; and b) returns her gaze to the camera quickly to

confirm it multi-modally. The subsequent treatment of the issue is evidence that vlog viewers, who only see what is on camera, depend on an explanation of what is causing the instability of the current participation framework. This would not necessarily be the case in face-to-face interaction, as in most cases interlocutors can identify the cause for loss of attention on their own. The same vlogger later uses the co-presence of two separate sites of interaction to deliver a playful reproach (cf. Example 4.4).

4.3.3 Conversational history

The history of a conversation usually includes those as addressees of the utterance that were addressees of the previous utterance, unless the speaker reassigns roles (Clark and Carlson 1982: 346). This holds true for vlogs as well. The communicative history has a slightly different dimension though. As vloggers do not talk to people face-to-face, but always into a camera which never actively participates, there are certain constants in the speech situation. Likewise, once a video is uploaded to the website, it stays there until removed by the vlogger. Some vloggers have accumulated several hundred vlogs on their accounts, which are available for viewing anytime. This contributes to the assumption that the themes of previous vlogs can be relevant to the participant status of a vlog viewer, which is reflected in some statements about new viewers as opposed to old viewers. In this context, the subscription function on YouTube is relevant. As subscribing to a vlogger's channel is an active choice by individual users, one can assume that users base this decision on some sort of knowledge of the vlogger's output. Furthermore, the subscribers display an interest in watching future videos by the same vlogger, thus help constitute an established audience ("old viewers").

The following example illustrates such a differentiation of the audience. It features a female vlogger in her twenties. She is instructing her viewers on how she expects them to behave when communicating with her online.

Example 4.3

1   hh. let's take a little second to talk to all the new people?

2   CUT

3   I don't know where you came from new people,

4   okay,

5   um we're cool,

6   okay,

7   it's fine that you're here, (2.0)

8   but.. just try a little bit, (2.0)

9   just try a little bit.

10   okay?

11   I got people lolling all over the place already?

12   I got people who don't.. seem to, (2.0)

13   just try.

14   okay just try for us.

15   together we can get through this.

This passage is headed by the vlogger's announcement of a participant status shift. Line 1 invites her regular viewers ("let's") to jointly address the "new people". Thus, on the production side she figures as the animator and her regular viewers are the principal alongside herself. The noun phrase "new people" is used to distinguish between two groups of viewers and undergoes a change from referential NP (l. 1) to vocative NP (l. 3) (cf. Zwicky 1974, Biber et al. 1999). Not only does that single out a subgroup of viewers as the direct addressees, but the repetition of the phrase in a different grammatical function introduces a humorous note, especially since it reflects a high degree of anonymity between vlogger and viewers (cf. Norrick and Bubel 2009 for the use of inappropriate forms of address in humorous interaction).

Lines 3-10 and 13, 14 are a highly repetitive admonition for the vlogger's new viewers, alerting them to her personal rules of conduct. It consists of multiple variations on the major theme of "just try" (l. 8, 9, 13, 14) and several instances of the pragmatic marker "okay" (l. 4, 6, 14), sometimes functioning as a boundary marker, sometimes as if to ask for a backchannel to signal understanding (l. 10). The rules which are alluded to are

never explicitly stated. They are treated as pre-existent knowledge, and in fact, they have been thematized in several previous videos. The only concrete mention of a behavior she does not accept occurs in lines 11 and 12: She is referring to the acronym "lol" (laughing out loud), here inflected as a verb form, which she condemns but regularly finds in written comments to her vlogs. Exactly those two lines represent a shift in the participation framework: The viewers whose comments she finds offensive are no longer addressed directly, but they are referred to in third person ("people lolling", "people who"), which leaves all other viewers as addressees. Her regular viewers are all the more aware of the vlogger's disapproval of the term "lol" since she actually repeatedly banned 'offenders' from writing comments to her videos. So here, the regular viewers are at an advantage over the new ones in that they have concrete knowledge as to what constitutes a breach of the vlogger's rules of conduct.

Finally, lines 14 and 15 introduce personal pronouns "us" and "we", inducing another shift on the production side of the participation framework. The first person plural pronouns indicate that her talk represents not only herself, but all viewers, inviting the new viewers to share their views.

This strict distinction of 'good' viewers (who know and abide by her rules) and 'bad' viewers (who break her rules), also expressed on both the production side and the reception side of the participation framework, creates solidarity among her regular viewers and constructs an in-group. Her presentation appears non-negotiable and based on strong opinions, partly because of the great amount of repetition throughout the passage. In a conversational context with turn-taking as organizational feature, this firm stance could be subject to negotiation. The regular viewers could support or reject the vlogger's stance, and the new viewers could defend their behavior, agree with the vlogger, or ask for more information. So here, by applying the participation framework the way she does, the vlogger acts as if the regular viewers agree with her, and the new ones do not say anything in their defense. This treatment works towards making the new viewers part of the group of old viewers. In conclusion, a vlogger can present and

use the results of a pretend conversational negotiation of participant statuses, and treat them as valid basis for further talk.

4.3.4 Manner of speaking

Another differentiation of participant status can be achieved by a change in manner of speaking. Whispering or using a markedly loud voice can exclude or include people as participants or addressees in face-to-face conversation, but these devices cannot be used to distinguish between different viewers of a vlog. Only if there is another person present in the physical surrounding of the vlogger (or in the physical surrounding of the viewer) do changes in volume have an impact on participant roles. The following example illustrates such a situation; the passage is the continuation of the opening passage cited above (Example 4.2).

Example 4.4

7	this is my first vlog on the computer.
8	°it's my birthday°. (2.4)
9	I will have some presents to open but, (2.6)
10	I can't get the uh action going,=
11	=to actually get the presents given to me.
12	so that I can [open them. (1.0)]
	[{gaze shift to speaker's right; then back}]
13	but uh John uh has them over there,
14	but he's doing something,
15	he's writing something on-
16	OH NO,
17	h- actually he's trying to help me.
18	he's look-
19	h- he- he's doing me a favor so,
20	so I'm not really criticizing.. him,

21   because in fact,

Figure 4.5, "°it's my birthday°."

Figure 4.6, "open them."

This passage sets in after the vlogger's efforts to establish a participation framework with the vlog viewers as addressees and the other person present in the room with her as an overhearer (cf. Example 4.2 above). She gives up this stability of the participation framework immediately by whispering into the camera (l. 8), which is accompanied by a posture shift towards the camera (cf. Fig 4.5). The subsequent lines (9-12), which refer to her birthday presents, gradually increase in volume. Initially, by means of whispering, John is staged as an unratified participant who is not permitted to hear her talk. It seems, though, as if this is only a play frame, since the vlogger's talk about her birthday presents is presumably soon (if not throughout) audible to John as well. A prolonged gaze shift at the end of line 12 that probably has John as a target (cf. Fig 4.6) is the only indicator (verbal or multimodal) that, after all, the main addressee is not the vlog audience. The vlogger is predominantly communicating her discontent to John over not receiving her presents. This framework is continued when the vlogger remembers that the delay is due to John doing her a favor: although she relates to the audience that John is not to blame, this serves mainly as a signal to John that she recognizes her criticism as unwarranted (l. 16-20). This is an instance of what Goffman (1979: 10; citing Fisher 1976) refers to as

> innuendo, whereby a speaker, ostensibly directing words to an addressed
> recipient, overlays his remarks with a patent but deniable meaning, a meaning
> that has a target more so than a recipient, is typically disparaging of it, and is

meant to be caught by the target, whether this be the addressed recipient or an unaddressed recipient, or even a bystander.

It is worth pointing out that the vlogger's use of pretend whisper is actually quite similar to this practice in face-to-face interaction, when it is not used to exclude overhearers, but actually to address them indirectly. The difference between this passage and a similar use in face-to-face conversation is the amount of contextual information the vlogger is giving her viewers: lines 13-18 describe to the vlog audience what John is doing, which would not be necessary if the audience could turn around to see for themselves.

A change in manner of speaking can also be used to effect a switch on the production side, as the following example shows. Here, a female vlogger in her thirties changes footing by voicing the audience.

Example 4.5

1   I don't know.

2   all my new subscribers are gonna be like,

3   "okay.

4   never mind this shit makes no sense,

5   she is insane?

6   goodbye."

7   but.. I hope you stay tuned,

8   cause,...

The vlogger talks about her new subscribers in third person, distinguishing them from the old subscribers and other viewers in general that are thus implicitly addressed (l. 2). The quotative "be like" (l. 2) introduces a shift on the production side: by speaking for the new subscribers, she makes the new subscribers the principal and reduces her own role to that of the animator (cf. Tannen 2004 for instances where people frame their dog as principal). This new alignment allows her to deliver fairly harsh criticism of her

own persona as presented previously in the vlog. By labeling her vlog as "shit" (l. 4) and herself as "insane" (l. 5), she achieves several things: she characterizes her new subscribers as potentially quite critical, and she precludes potential criticism from her non-verbal audience by stating it herself. At the same time she displays both self-awareness and a sense of humor, while bonding with her old subscribers. Couper-Kuhlen (1998: 15, 1999) shows that prosody and construction of voiced utterances "hint at stances which they [the voiced interactants] are constructed as assuming - and/or which their constructor, the current speaker, wishes to take towards them". The high functional load of this fairly short passage (l. 3-6) demonstrates that voicing someone else is, in monologues as much as in talk-in-interaction, well suited to create one's identity, as it displays stance (cf. Ochs 1993).

In line 7, two shifts in footing occur: on the production side, the vlogger returns to the role as principal, author and animator in one; on the side of the projected audience, the new subscribers, who were first bystanders and then principal, are now the direct addressees, in that the vlogger addresses her request directly to them. These two shifts are marked with the help of the discourse marker "but" and a shift back to the speaker's normal voice. Couper-Kuhlen (1998: 15) discusses how changes in prosody support the change in footing when a speaker voices another person, helping the recipient to recognize that shift: "The speaker, in other words, is not only 'doing' voices but also doing something with those voices which it is incumbent upon recipients to infer". Though this research is based on talk-in-interaction rather than monologue, it also stresses the active role, and thus involvement, of the recipient.

As in Example 4.5, this strategy can be employed similarly in face-to-face interaction: a speaker voices another party in the conversation for a certain effect (such as to mock them, to emphasize a certain stance or characteristic etc.). In a vlog, however, the speaker has to do this based on speculation about the audience. Likewise, the part of the audience which is thus characterized cannot respond immediately, which means that the negotiation of its actual stance is out of the question.

4.3.5 Linguistic content

In face-to-face conversation, interlocutors can designate the addressee(s) and other participant roles through the content of their utterances, e.g. by using terms of address, personal pronouns, questions, and other types of directed language. In vlogs, speakers do not have their interlocutors physically in front of them, in fact, they normally do not know who they are. Therefore, their directed language is based on their assumptions about who is going to watch the vlog. This uncertainty is reflected in the use of if-clauses to describe parts of the audience (e.g. "if you're interested..."). Terms of address remain generalized, e.g. "you", "you guys", or general but website specific "YouTubers", "vlog fans".

The following example was recorded by a woman in her twenties. She is informing her viewers about the progress she has been making in her effort to lose weight before asking the viewers to relate their experiences. The passage does not illustrate a switch in participation framework so much as it demonstrates the difference between soliloquy and monologue, which, according to the definitions given below, differ in that one has traces of audience design, and the other does not: According to Hasegawa (2010), soliloquies are characterized by the lack of an addressee in the talk. In a monologue, the "audience isn't expected to interrupt" (Clark 1996: 5), but that does not exclude addressing them. Line 122, at the beginning of the passage, and lines 133 and 134, in the end, represent soliloquy in the sense of Hasegawa (2010), as there is no marker of address in them. According to Clark (1996), the complete data excerpt is a monologue, simply because of the constraints of the medium video.

Example 4.6

122 it's going really well,

123 h so my question is,

124 how are you guys doing.

125 I wanna know how you guys are doing.

126 uh I wanna know how uhm your first day was,

127	and uhm.. do you feel a change already.
128	a:nd.. if you haven't started yet,
129	are you trying to get started,
130	or have you.. have you made plans to get started.
131	let me know,
132	leave it in the comment section below.
133	uhm,
134	so yeah,

Lines 122 and 123 represent the transition from the vlogger's report to her request for information from the viewers. She announces her first question (l. 124) and then continues with two further requests for information, which are not in the syntactic form of interrogatives (l. 125-126). Three more questions follow, one prefaced by an if-clause (l. 127-130). The detailed questions suggest significant common knowledge for the speaker to draw on. Clearly, there must have been communication between vlogger and viewers about the topic prior to the recording of the vlog. If the vlogger can talk about the viewers' "first day" (l. 126), she must have been in contact with them to find out that they are embarking on a project stretching over several days. Lines 131 and 132, containing an imperative structure each, hint at the preferred channel of future communication: the comments section is commonly referred to by vloggers when they are trying to motivate their audience to reply (cf. Example 4.1 above).

Although there has obviously been communication previous to this vlog, the vlogger uses only unspecific terms "you guys" (l. 124-125) and "you" (l. 127-130) rather than real names or user names. In conversation, these terms simply represent a casual style or neutral (plural) pronoun, used to address several people at the same time. In vlogs, these and similar terms are the default. Addressing a single person specifically in a vlog can only be done by using their name or alluding to a very specific characteristic. In conversation, multimodal features such as gesture, gaze shift and volume can play a major role in indicating addressee status. Thus, when a vlogger uses *you* without having verbally indicated a different addressee, the whole audience is addressed. If an

interlocutor uses *you* in a conversation, he or she might address only a part of all present, indicating this with the help of multimodal features.

The talk in Example 4.6 is designed as the basis for future communication, and it reflects the lack of an immediate exchange system such as turn-taking, but also the website specific options, namely the transfer of an exchange to another communicative channel. Therefore, the request for information through the comments section (or other channels, e.g. email) does not only involve the audience through their recognition of the relevant participant status; here, their active participation is requested.

Another device used to induce an addressee shift in vlogs is the switch from third person to imperative, as illustrated by Examples 4.7 and 4.8. The first is by a woman in her twenties (the same vlogger as in Example 4.3 above), the second by a man in his fifties.

Example 4.7

1	CUT
2	this is a great welcome video for all the new subscribers.
3	<so watch every fuckin' second of it.>
4	CUT

Figure 4.7   Figure 4.8

This short excerpt illustrates the shift of participation status of the vlogger's new subscribers. Their role changes from that of participants to that of direct addressees. The speaker in this example immediately switches to the imperative mood. That causes

a grammatical ambiguity: strictly speaking, in line 3 the vlogger could still be addressing the complete audience as in line 2. That interpretation seems less likely, simply because in the larger context of the utterance (i.e. the whole vlog) nothing indicates that there is a reason for all viewers to follow the video closely as a welcome video. Aside from the content of the utterance, the discourse marker "so" and the drastic shift of the vlogger's facial expression from friendly to threatening (cf. Figures 4.7 and 4.8) and the accompanying shift in voice quality (l. 3) support the hypothesis that there is a status shift in the participation framework. Line 3 also displays the unexpected use of a swearword, and it is performed in a distinct rhythmic pattern: "every", "fuckin'" and "second" each receive pronounced stress on the first syllable, adding to the overall impression of fake threat imposition. Couper-Kuhlen and Selting (1996: 28) show that contrastive prosody can make "a substantial contribution ... to the situated interactional meaning of talk", here illustrated by the low pitched, rhythmic and slow delivery of line 3. This stance shift is in line with the vlogger's efforts in Example 4.3. Both passages contain shifts in the participation framework to address and thereby educate a subgroup of the audience, quickly helping them to adapt to the audience as a whole. In both examples, the vlogger makes an assumption about her new viewers, that is they lack knowledge about how to be an appropriate viewer of her vlogs. Example 4.8 is similar in that also an imperative triggers a switch in participation statuses.

Example 4.8

57	uh so,
58	if somehow he actually sees this video blog,
59	and and wants to defend himself,
60	you know,
61	have at it.
62	I'd love uh uh to debate him in any form on this subject.
63	but anyway,

The "he" (l. 58) in question is Paul Krugman (a Nobel Prize laureate in economics), whose opinions on political and economic matters are discussed in the vlog. In line 58, the vlogger employs an if-clause which does not serve to address Krugman directly. The content of the utterance actually even points out the low probability of Krugman's status as a participant (ratified or not) of the vlog. In line 59, the vlogger continues the if-clause, still clearly talking about Krugman in third person. Line 61 introduces an abrupt and short switch (it lasts only for that one intonation unit): the imperative "have at it." addresses Krugman directly. Grammatically, it might be an alternative continuation of line 59 ("wants to defend himself"), but the falling intonation, which is in contrast to the previously established pattern of constant or slightly rising intonation (l. 57-60), leaves "have at it." as an imperative as the most likely interpretation. Curiously, the vlogger does not start addressing Krugman directly as early as line 58, for example by saying something to the effect of "if you actually see this video blog,". The vlogger is clearly struggling with the fact that he does not know who is watching the vlog, and that Paul Krugman might not be part of the audience after all. In a face-to-face conversation, this problem would not occur, as interlocutors can perceive who they are interacting with.

4.6 Combining different types of involvement strategies

The previous examples have served to illustrate the different types of devices that vloggers use to induce a shift in participation status, and how they differ from their use in conversational settings. The final example contains a combination of these devices. The vlogger (the same as in examples 4.2 and 4.4) is commenting on videos of people playing a Django Reinhardt song.

Example 4.9

     51   what I found was=

     52   =about a hundred people have recorded themSELves.

     53   playing.

     54   Nuages.

55   uh,

56   you know.

57   of course they admire Django Reinhardt,=

58   =would like to,

59   try to play guitar like that.

60   but uh::,

61   you know,

62   it's a little pain-

63   uh ((laughs)),

64   °it's a little painful to watch°

65   °you know,°

66   °people trying to play guitar,°

67   you know,

68   if you're gonna try to play guitar,

69   uh TUNE the damn guitar.

70   if you're gonna record yourself and put it up on uh YouTube,

71   TUNE the guitar at least,

72   there's no excuse for that,

73   not-

74   I mean there's an excuse for your bad playing.

75   you're not a good enough guitar player.

76   there's no excuse for not tuning your guitar.

77   that's my advice to the guitarists of the world.

78   so anyway,

79   uh.

80   so I'm waiting to open these presents,

Lines 51 through 62 are characterized by the vlogger's use of the third person to describe amateur guitarists and their interpretation of a Django Reinhardt song she viewed on YouTube. They are treated as a possibly non-present, but definitely not

addressed, party. The first sign of involvement of these musicians as potential audience is the interruption of the intonation unit in line 62, which consists of a clearly negative evaluation of the musicians' efforts. It marks a shift in the vlogger's projected participation framework. There are four more markers of a shift in footing: hesitation marker "uh", laughter (both in l. 63), the sudden and noticeable decrease in volume, and a subtle shift of her posture towards the camera, see Fig 4.9 and 4.10 (l. 64 - 66).
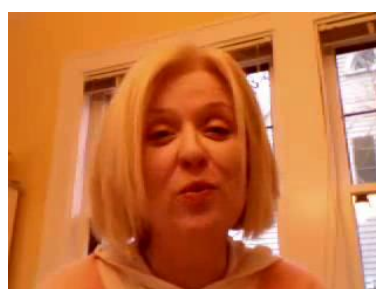
Figure 4.9, "it's a little pain-"   Figure 4.10, "°it's a little painful to watch°"

These features result in a pretend separation of the audience into non-ratified participants, namely the guitarists, and ratified participants, namely everybody else. As in Example 4.4, her whisper in lines 64-66 cannot actually differentiate in that way. Thus, in this example, the speaker is not effectively excluding the musicians she is referring to from participating; however, if both in the production and the reception of this vlog the participants apply their knowledge of face-to-face communication, then the viewers should be able to infer this virtual classification into participants and non-participants. Van Leeuwen (2009: 71), discussing loudness, describes this knowledge as follows: "at 'close range' our voice will be a soft whisper, whether for reasons of intimacy or conspiracy". The functions of her whispering then are a) to introduce a humorous note, and b) thereby to achieve a less serious keying of her criticism.

A further shift in participation framework occurs in line 67: the sudden increase in volume accompanies the transition to a footing that allows the vlogger to address the musicians directly. Several occurrences of second person possessive and personal pronouns (e.g. "your bad playing", l. 74; "if you're gonna try", l. 68) as well as directives (e.g. "uh TUNE the damn guitar." l. 69) distinguish the guitarists as the ratified

participants and addressees, while the rest adopt a bystander position. The aggressive tone of voice, the use of an expletive ("damn", l. 69) and the harsh criticism frame this passage (l. 67-76) as a reproach. The coda-like shift back to a uniform group of ratified participants occurs in line 77, where not just guitarists on YouTube, but all "guitarists of the world" are mentioned in third person instead of addressing them directly. With that shift accomplished, the vlogger finishes the topic altogether and transitions to another topic with the help of discourse markers and a filled pause (l. 78-80).

This combination of several strategies in a short episode of talk demonstrates a general quality that these camera monologues potentially have: While a real-time shift in participation frameworks as negotiated in a face-to-face conversational setting is impossible due to the lack of immediate feedback, vloggers can employ a range of strategies that make their speech resemble multi-party conversation. Viewers are not only required to recognize their own participant status, but oftentimes they are requested to establish contact with the vlogger through written comments etc. This renders the monologue lively and, to a degree, entertaining.

4.5 Conclusions

Monologues are characterized by a lack of active audience participation in the sense that no immediate contributions to the interaction are to be expected from them. This is true for vlogs especially, since the asynchronous, one-directional channel chosen in this genre precludes any immediate participation of the viewers. The culture of YouTube, however, values communication between its members (see, e.g., Burgess and Green 2009), with YouTubers frequently asking for feedback of any kind on their videos. If there is any verbal contribution from the viewers, it is delayed, usually in written form, and mediated through a different channel (written comments, email, etc.). In the present chapter, I have argued that, despite the asynchronicity of the exchange, there is a form of audience involvement present in vlogs resembling that of face-to-face conversation, distinguishing this kind of monologue from the notion of soliloquy.

Audience involvement is closely tied to audience design and footing or participation status, two theoretical frameworks that further distinguish the coarse notions of speaker and hearer into more precisely defined roles both on the production side and the reception side of interaction. It is essential that a speaker assigns these roles, and that the audience recognizes what role they are being assigned. Even though the audience's recognition of their roles does not necessarily require a signal of their understanding, it requires at least the correct processing of that information for successful communication. In receiving a monologue such as a vlog, the audience is involved because they are expected to decode their and other participants' statuses correctly. In short, vloggers involve their audience by assigning them participant statuses.

Vloggers use a variety of devices to assign participant statuses. Some of them resemble those used in face-to-face interaction; some are very specific to the medium online video as they are adapted to the technical restrictions. Adopting a model from Clark and Carlson (1982), they can be classified into five categories: physical setting, conversational history, gestures, manner of speaking/prosody and content of the talk. The data analysis in this chapter reveals how these devices apply to talk in vlogs.

The vlogger and the audience do not share a physical surrounding. The place they share and where communication takes place is a website, thus a digital or virtual space. Hence, the choice of medium has an impact on participation status: once a person accesses the shared virtual space, he or she is a participant. The choice of medium also has consequences for the physical spaces of vlogger and audience: The vlogger has to stay within his or her camera's focal range and the microphone's audio range, which in most cases is static, and he or she will not see the audience or know what their surroundings look like. The viewer has to stay within reasonable distance from his or her computer, and will only see a small, usually static, part of the vlogger's physical surroundings. These circumstances rule out many devices present in face-to-face interaction: As the contextual configuration does not permit visible co-orientation, gestures (e.g. pointing, nodding, posture shifts) and gaze behavior cannot be used by

the viewers to negotiate participant status. Physical proximity or distance is not available as a resource to distinguish different vlog viewers.

With these restrictions to battle, vloggers are fairly creative in their use of multimodal features. A change in volume, for example whispering, was shown to be an important component in separating the audience into two groups when issuing criticism. It is argued that, although lower volume does not exclude any specific viewers as participants, viewers recognize that the vlogger is differentiating between viewers because of their experience of the usual effects of whispering in off-line encounters. Likewise, a sudden increase in volume can reverse the addressee/participant status. Other aspects of prosody play similar roles when applied contrastively: distinct rhythm, markedly changed pitch and fast or slow delivery can set off utterances from their surrounding talk and thus clarify grammatically ambiguous utterances, or they can signal voicing of another interactant.

Gaze behavior plays a role in assigning participant status when another person is present in the vlogger's surroundings; it is used as a means to distinguish vlog viewers from that other person and to assign them a different participant status. The same is true for whispering, which can exclude a person that is physically present from participation (even if, as in the cited example 4.4, it is just a pretend exclusion), while vlog viewers remain participants. Gesturing, especially pointing, is used to refer to entities in the shared, virtual space of the website, which is a sign of accommodation towards the viewers.

Shifts in the participation framework are often accomplished by switching from second person or imperative to third person ('talking to' as opposed to 'talking about') or vice versa. Terms of address can distinguish groups of viewers from others, but this is usually based on one of their very general characteristics (e.g. the time a viewer has been subscribed to a vlogger's account, as in "new people"). Rarely does a vlogger address a person by their real name. As vloggers know little about their audience, this relative anonymity is not surprising. It also explains the use of if-clauses to invite participants to self-select as addressees. Quotatives are used to introduce shifts on the production side,

as they result in the voicing of a figure, which is an important means of identity creation. The analysis has revealed that those shifts can be accompanied by discourse markers, by swearwords and by shifts in prosody, as discussed in the analysis.

Returning to the discussion whether vlogs can be characterized as primarily referee-designed or audience-designed (in the sense of Bell 1984), I have presented a range of devices that seem to be designed to move vlogs towards audience-design. Vloggers specifically address viewers, or subsets of viewers, using the knowledge available to them: viewers are identified by their ability to leave written comments; they are identified by their status as new or old subscribers and their subsequent knowledge of the vlogger's ways; they are identified as dieters or guitar-players or experts in economics; they are identified by their inability to see the vlogger's surroundings beyond the camera range. These forms of identification are a way of replacing one of the fundamentals of face-to-face interaction, that is a sense of who one's interlocutors are. Thus, any of these devices make vlog-monologues resemble conversation to some degree, which works towards the general goal of YouTube members to establish contact amongst each other. Therefore, these strategies are attempts to create familiarity between the members of the YouTube community, which again fosters communication between vlogger and viewers. On the other hand, the anonymity reflected in the use of neutral personal pronouns and other terminology ("you", "you guys"), is a reminder that vloggers talk into a camera and therefore to an unknown audience. Thus, there is a clear trace of referee design.

Much research on monologue, especially in the vlog setting, remains to be done to uncover how vloggers involve their audience. Future research might try to quantify the devices discussed here, correlating them with a vlogger's success in terms of written feedback. A more thorough analysis of vlog subgenres and their specific audiences, as well as the written feedback that these viewers leave, could also show how this ongoing, asymmetric communication taking place in different media represents a self-sustaining system. This is approached in chapter 7.

5 Repetition as involvement strategy in monologues

5.1 Introduction

This chapter reports research on the creation of involvement in monologic talk through the use of repetition. It investigates instances of lexical repetition and patterned rhythm in video blogs (vlogs), exploring both their form and function in context. Some researchers have pointed out that the production of a monologue is based on speakers' experience of dialogic talk-in-interaction: Haviland (2007: 150) writes "interaction is a compelling model for talk, even apparently monologic talk". Similarly, Schegloff (1987: 222) claims that "speech exchange systems, and their turn-taking organizations, are the product of transformations or modifications of the one for conversation, which is the primordial organization for talk-in-interaction". This forms the basis of my assumption that the monologues featured in vlogs display characteristics typical of conversational data.

One such characteristic is the creation of involvement. Involvement is understood as describing both listening and speaking in conversation as active participation, where both include traces and elements of the other (Tannen 1989: 12). In vlogs, interaction takes place not immediately, but via other channels, e.g. written comments or email messages, once the video has been uploaded. The creation of involvement is used, among other things, as a means to inspire communication through other channels. Vloggers must apply strategies to accomplish such involvement.

The strategy focused on in this chapter is repetition. Tannen (1989: 80) observes that "repetition is pervasive and functional in conversation". While most studies on repetition use interactional data, the present study tries to demonstrate that in monologues repetition is also pervasive and functional. With this research, I hope to contribute to our understanding of the vlogging community and its linguistic practices. Lange (2007: no pages) characterizes its members as sharing "a commitment to video as a crucial means of expressing and understanding issues that the video blogger wishes to share". So even though vlogs represent monologues, they are used as means of

interpersonal exchange. My research aims at revealing interactional elements in a genre that initially depends on one participant only. The following sections give an overview of previous research on repetition and involvement, and analyze instances of repetition in vlogs towards revealing their function as means creating involvement in spontaneous monologues.

5.2 Repetition as involvement strategy

Involvement as an analytical tool has so far mainly been applied to conversational activity. Exceptions are Tannen's (1989) treatment of public speaking and oratory, Frank's (1989) discussion of (written) advertising material, and Chafe's (1982: 50) comment on ritualized performance in religious contexts, for instance. However, there has been no work on spontaneous monologues regarding involvement strategies. In the following, I review research on involvement, extracting and adapting those points relevant to vlogs.

Gumperz's (1982: 1) notion of involvement, that is successful participation in a verbal exchange, requires "knowledge and abilities which go considerably beyond the grammatical competence we need to decode short isolated messages". This includes perceiving and interpreting "particular constellations of cues in reacting to others and pursuing their communicative ends" (4-5). Furthermore, conversationalists must actively signal their involvement in conversation, e.g. through nodding, gaze, or verbal participation. Gumperz makes involvement a prerequisite to understanding (2). In sum, to achieve and maintain involvement, participants in a conversation must infer the meaning of individual utterances, as well as their own role as a participant in the ongoing conversation, and consequently what is expected of them. While the inference of meaning and speaker roles can be achieved by vloggers and their viewers, active signaling of the viewers' participant status is impossible. In consequence, the vlogger has to assume that the viewers understand their status based on their experience of both the medium and regular face-to-face conversation. This is feasible insofar as the

written comments underneath vlogs testify to viewers' understanding of the vlog content.

Chafe (1985: 105), in an article on the difference between spoken and written English, notes that the prototypical spoken genre is characterized by fragmentation and involvement, and the prototypical written genre is characterized by integration and detachment. About the causes of this difference, Chafe (1982: 45) writes that an author's

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

Vloggers face a similar situation: their discourse is displaced, and they do not know who their audience is. Still, these factors cannot be the only ones in determining the level of involvement, and it is the aim of this chapter to demonstrate how vlogs achieve a high level of involvement. Chafe (1985: 117) distinguishes three types of involvement: 1) self-involvement of speaker, 2) interpersonal involvement between speaker and hearer, 3) speaker involvement with what is talked about. The first is expressed through the use of first person singular pronoun "I" and reference to thought processes, e.g. "I think". The second is expressed, among other ways, through second person pronoun "you", or asking for confirmation, e.g. "right?". The third is expressed through the use of expressive vocabulary, e.g. "really". These three types can all be found in vlogs. My research on monologues is predominantly concerned with vloggers' involvement with the absent viewer, therefore this chapter centers around Chafe's second type of involvement, interpersonal involvement between speaker and hearer. I approach this type of involvement primarily through the analysis of repetition, identified as a major involvement strategy by Deborah Tannen (1989).

Involvement, for Tannen (1989: 12), entails an "internal, even emotional connection, individuals feel which binds them to other people as well as to places, things, activities, ideas, memories, and words". The activities of "speaking and listening in conversation include elements and traces of the other" (12). The emotional connection vloggers feel towards their viewers while talking into a camera is difficult to trace and describe systematically, especially in terms of involvement. There is, however, ample evidence that it exists, for example when vloggers state that their main motivation for vlogging is receiving comments from viewers or staying in contact through other channels, or when they address their viewers saying "I love you". It seems clear though that when vloggers speak, there is self-monitoring (or traces of listening) in operation, and this becomes evident for example in self-corrections.

Besnier (1994) assesses the literature on involvement critically, pointing out that generally its scope is too broad to be a meaningful analytical tool, and that it has mainly been established in terms of a culturally narrow (i.e. Western) frame, minimizing its general applicability. While this criticism is warranted, particularly in cross-cultural contexts, I believe my use of it is justified. My analysis focuses on just one phenomenon related to involvement, namely repetition. The data under analysis is English, mostly by native speakers, and it was created for a very particular CMC context. This suggests a fairly uniform data set, to which the notion of involvement, as outlined above, seems applicable.

Tannen (1989: 48) lists a number of functions of repetition in discourse, which can be grouped into the categories of production, comprehension, connection, and interaction. Their overarching function is to establish coherence and interpersonal involvement. She thus identifies repetition as a major instrument for the creation of involvement. Similarly, Johnstone (1987: 207) considers repetition a cohesive device, a way "to keep conversation and monologue glued together". Furthermore, it is a way of creating categories, where repeated frames in discourse can group new, unfamiliar items with old ones (207). Tannen (1987: 216) points out that all discourse, including monologic discourse, is interactionally developed, in that "any utterance echoes prior utterances",

which constitutes the observation that, in a sense, all language use consists of repetition. She further claims that "nonconversational discourse types … use repetition strategies such as those observed in conversation" (1989: 80).

The concrete functions repetition fulfills in discourse, among others, are as follows: showing participatory listenership, ratifying listenership, humor, savoring, stalling, expanding, participating, evaluating, and bounding episodes (Tannen 1989: 59). Some of these can only be found in multi-party interaction, others can also be found in monologues. The first two, display of participatory listenership and ratification of listenership, can technically not occur in monologues, since they appear in the form of allo-repetition. As there is only one speaker present, there is no immediate repetition of another speaker's words or intonation possible. Practically, a vlogger can retrieve another's words from memory, or he or she can represent allo-repetition in constructed dialogue. Alternatively, he or she can read out viewer comments from previous videos. The remaining functions of repetition, that is humor, savoring, expanding, participating, evaluating, stalling, and bounding episodes, can be realized in monologues. However, the difference in participation structure between monologue and multi-party interaction necessarily results in a different realization of these functions of repetition. Especially participating as a function of repetition has a special role, since anything a speaker says or does in a monologue instantiates participation, whether repeated or not. Possibly, participating in, and thus delivering, a monologue entails, in its ideal form, a constant flow of speech without gaps. While in conversation, the task of "filling silence" is shared by several people, in a monologue, a single speaker must shoulder it alone. Thus, a significant function of repetition in conversation, namely signaling active involvement in interaction, perhaps translates into a slightly different function in monologue: here the participant role is already fixed, and repetition helps to fill it without too much effort on the production side. The following discussion of vlog excerpts will reveal when vloggers use repetition, how they use repetition, and to what ends they use it.

5.3 Repetition across vlogs

The first two examples (5.1 and 5.2), structurally different from the ones that follow (5.3 – 5.13), contains fixed phrases and expressions that are coined by the vlogger and occur in the same position in several or all of his or her vlogs. Thus, the original that the repetition replicates is not necessarily (but potentially, cf. Example 5.2 below) found within the same vlog, but is part of previous vlogs. Especially openings and closings of vlogs are often realized with the help of characteristic, fixed phrases (cf. Frobenius 2011, chapters 2 and 3). The two passages below, for example, are the opening lines of two vlogs by the same vlogger:

Example 5.1

    1   hello I am Yvonne Smith,

    2   and I am the travelling vegetarian.

    3   I hope you are doing well?

    1   hello my name is Yvonne Smith,

    2   and I am the travelling vegetarian.

    3   first things first,

One can assume that this is a signature opening that recurs in most, if not all, of her vlogs. The first two lines contain a greeting and two instances of self-identification, one by first and last name "Yvonne Smith" (l. 1), and one by a description reminiscent of an occupational title "the travelling vegetarian" (l. 2). The fixed opening facilitates the production of a monologue, in that it reduces the task of on-line speech production to retrieval from memory. At the same time, the speaker fills the exposed initial position of the speech event with information that is essential to both her sense of identity and (as the content of the vlogs shows) to the major themes she discusses. Furthermore, the salient repetition of these elements introduces formality. A regular viewer will recognize the opening as belonging to this vlogger, and thus feel a sense of familiarity conducive to the creation of involvement.

5.4 The influence of editing

The following examples illustrate the impact editing has on repetition in vlogs. The example passages below are taken from the same vlog. The first four lines represent the opening, and lines 20-22 are from the vlog body.

Example 5.2

      1   what's up YouTube,

      2   it's your boy PBone coming at you again,

      3   and let's take a ride with me?

      4   right?

- - - - lines 5-19 not reproduced here - - - -

    20 CUT

    21 {caption "later that night"}

    22 what's up YouTube, {in the car, music playing}

    23 I'm on my way,

- - - - lines 23-85 not reproduced here - - - -

    86 CUT

    87 {caption "I was the first person at bestbuy"}

    88 what's up YouTube,

    89 I just got back home.

As indicated by the vlogger's written message in the caption following a cut, there was a temporal shift between the recording of the opening and that of the vlog body (l. 21), and another later in the vlog (l. 88). The vlogger had to accomplish the task of opening a monologue three times. He does this by recycling the phrase from the actual opening (l. 1; l. 22; l. 88), an informal greeting "what's up" combined with the term of address "YouTube", a website specific, generalizing metonymy addressing his viewers. Their function differs insofar as in the first instance, it opens the vlog, and in the second and third, it reopens the vlog or opens a new episode. While this phrase does not seem quite as ritualized as that in Example 5.1, because it does not reflect the vlogger's agenda as

much, it still serves the major function of facilitating the production of a monologue as it does not require full on-line speech production capacity. As this example demonstrates, the asynchronous medium video offers vloggers editorial options that enable them to introduce temporal or spatial shifts, sometimes resulting in the repetition of the same phrase in a similar, but different, functional slot.

The repetition of the greeting "what's up" can be related to the function of creating involvement. As a greeting, it is directly addressed to the viewers, informing them that the speaker is establishing a communicative channel between them. The two repetitions invoke the memory of the actual opening (as well as, perhaps, openings of previous vlogs), and they re-enact parts of the communicative history that vlogger and viewers share. According to Johnstone (1987: 212): "Interlocutors in any situation create a shared language as they talk, by evoking shared memory and fitting what is not shared into what is". For vlogs, this allows the conclusion that speakers use repetition not only to facilitate their speech production, but they are providing the viewer with a recognizable frame, something known to them that they can relate to, before approaching a new topic. They create coherence and familiarity across vlogs and therefore involve their audience.

The following instance of repetition is also due to editing options. In Example 5.3, a young woman announces the next thematic section of her vlog, that is, it represents the transition from the topic of make-up products to that of clothing items.

Example 5.3

     149   okay now,=

           [{picks up make-up items and moves them to the side}]

     150   =[those of you interested in the c:lothing... .h p:art of the video?.. {gaze into camera, smile}

     151   okay] now those of you interested in the clothing part of the video,

     152   {changes facial expression, snaps fingers, points into the camera}

     153   watch right now.

154   okay.

Lexically, line 151 is an exact replication of lines 149 and 150. It consists of boundary marker "okay now" and noun phrase "those of you interested in the clothing part of the video,", which singles out a part of her audience as addressees. The message directed to them follows in line 153, "watch right now.", a verb phrase in the imperative. It is delivered with the support of several multimodal features: 1) a shift in facial expression from casual to determined, 2) a pointing motion into the camera carried out with both index fingers, which starts with the lifting of the hands up to the chest, and then proceeds to a snapping of the fingers of both hands, leaving the index fingers stretched out in the direction of the camera and the thumbs directed towards the ceiling. This elaborate pointing and the mimic action require both hands and apparently focused attention, which explains why the speaker aborts and repeats.

The 'performance' of line 150 contains many pauses and hesitations, possibly caused by the parallel action of moving small objects and the speaker's realization that snapping fingers and pointing cannot be carried out while her hands are occupied otherwise. Possibly, the speaker had planned that sort of transition for effect even before the recording, which would justify calling it a performance. But even if she made a spontaneous decision to include such a passage, she seems to be aware that the editing process allows cutting scenes as in the production of a movie. She does not use an abrupt cut-off or any other signs of self-correction, such as "I mean" etc., but calmly finishes the intonation unit while putting down the objects in her hands and repositions herself by shifting her gaze into the camera and smiling very briefly. I assume that the vlogger meant to edit the video accordingly, and thought of that option while delivering line 150, but neglected to do so for reasons unknown. While this is implicit evidence of the awareness of editing options, some vloggers also show this explicitly by addressing the topic. Another speaker, for example, comments "I apologize if that was too horrible, ((laughing)) if it's really bad I'll just cut it out". Both the noise referred to and the comment (obviously) remained in the footage, though.

The function of this repetition is only found on the production side because the viewers are only meant to see and hear the second rendition. For the speaker, line 150 then is retrospectively turned into a practice run for line 151, which seems to be the more satisfactory version. This passage is an instance of the speaker's taking the viewers' perspective in that she imagines the finished product as it would be seen by the viewers. The vlogger tries to optimize her output to have a certain effect on the viewers, which enhances the "internal, even emotional connection" Tannen (1989: 12) cites as constitutive of involvement.

Example 5.4 contrasts with Example 5.3 in that the repetition was not created in the speech production process, but entirely in the editing process. A man in his thirties is telling a story about recent online communication he participated in.

Example 5.4

    68   Jesus reigns is a great song.

    69   CUT

    70   we're probably gonna do it at () very soon

    71   CUT

    72   and the- CUT and the- CUT and then my tweet deck pops up.

    73   CUT

    74   my computer program that monitors twitter for me.

    75   CUT

    76   there is a tweet.

Line 72 contains two cuts, both producing identical versions of the first two words of the intonation unit. The intonation unit "and then my tweet deck pops up." serves as the complicating action of his narrative, told in the conversational historic present. The double repetition of "and the-" would possibly be considered an instance of stuttering in face-to-face conversation. Here, it adds a sense of dramatization accompanying the story line. Parallel to the way conversationalists make their narratives interesting to their interlocutors, this vlogger makes his story, which so far has little in the way of

excitement, more appealing. He does so by using editorial repetition rather than vocal repetition, which is an option unavailable in face-to-face talk. So while conversationalists could use multimodal features to achieve the same effect, the vlogger chooses a means from another production level. The considerable effort of cutting and editing the video, which is much greater than, for example, altering stress or intonation, is used to indicate to the viewer that the narrative is reaching an important passage. This shows that the vlogger is concerned about the absent viewers' involvement with the story, which in face-to-face conversation could be negotiated if necessary.

While Example 5.3 and Example 5.4 demonstrate speakers' awareness of editing options, Example 5.5 and Example 5.6 in the following section show how vloggers can exploit the medium and its particular settings, that is audio-visual mode and asynchronous transmission, for a specific purpose, such as humor and integrating viewers' comments.

5.5 Repeating viewers' comments

Vloggers can realize allo-repetition by reading or paraphrasing viewers' comments. Integrating viewers' actual questions and providing the answer shows concern for the viewers' interest in the interaction. Although this repetition can only be recognized as such from the context, because the original question or comment is not present, it is a means to represent the audience's recent involvement. In Example 5.5, a woman cites a question she has been asked repeatedly.

Example 5.5

136 uhm is there more I wanted to say,

137 oh I got a real load of comments telling me,

138 how the hell can you weigh 89 kilos?

139 uhm let me tell you,

The speaker indicates the fact that she is quoting from viewers' comments: "oh I got a real load of comments telling me," (l. 137). The actual quote follows in line 138, "how the hell can you weigh 89 kilos?". The vlogger subsequently begins to answer the question: "uhm let me tell you," (l. 139). The question is quoted using direct speech, using the pronoun "you" to address her, rather than reported speech with a shift in personal pronoun and backshift of tenses. Thus, she represents the question in the form it would occur in face-to-face conversation, even emphasized with taboo language "how the hell" to stress the viewers' incredulous stance. A question, as the first pair part of an adjacency pair, makes an answer conditionally relevant. Thus, the speaker sets herself up to make her statement a required contribution by acting as if it belonged to a synchronous exchange. This passage shows that by acting out the viewers' part as interlocutors, especially when integrating real questions from the audience, a vlogger can involve the viewers up to a point where monologue and conversation become very similar.

5.6 Play with visual mode, humor

A man who has just prepared dinner for himself is getting ready to eat in front of the television. The camera is focused on the TV set, possibly on a tripod, when he gets between TV and camera, resulting in a close-up of his face.

Example 5.6

1   uhm my dinner's ready,= {speaker not in the picture}

2   =I'll just show that to you, {gets in the picture, resulting in extreme close up}

3   over here, {gaze shift away from camera, then back}

4   this way, {index finger point, apparently directed to the side, only very briefly visible}

5   THIS way. {index finger point, apparently downwards, only very briefly visible}

6   alright.

7   [lady on TV: did you guys know that liquid paint thinner,]

8   [{camera shift to food}]

In this passage, the vlogger plays with the fact that the viewers' vision of his surroundings is restricted to the static scope of the camera. In lines 3-5 he asks his viewers to turn their attention to his food. He does so not only verbally, but also multimodally with gaze shifts and index finger points. The camera remains fixed in one position, and the food is not visible to the viewer, which makes it impossible to follow his request. Likewise, it is not possible for him to check the viewers' reaction. This turns his repeated and increasingly urgent request nonsensical: rather than acknowledging that the viewers' visual scope is static and restricted, he pretends to be participating in regular, non-mediated face-to-face communication and acts as if his interlocutors need additional help with finding the location of the referent. Thus, the excerpt represents an attempt at creating humor, using repetition. Another humorous use of repetition can be seen in Example 5.7.

Example 5.7

87   okay well,

88   I'm not gonna make a membership there,

89   because I'm coming BACK from Edmonton,

90   CUT

91   oh well if you're coming back,

92   well in that area there's a place here.. and here,

93   and here,

94   and here,

95   and here,

96   and here,

97   oh don't forget the one over there.

98   and uhm,

99   we have one right here too.

100   CUT

101   basically the whole time,

102   I was just like, …

103   can I just give you my membership and leave please?

104   don't you hate that when people try to-,

105   try to convince you to: .. give money to them.

106   I don't know,

107   gyms are like that,

This excerpt is part of a narrative: a woman in her early twenties is relating her experience of canceling a gym membership. Constructed dialogue is delivered with the help of cuts to indicate speaker shift. Lines 87-89 are spoken in a slightly lower pitch, representing herself as the customer; lines 91-99, spoken in a slightly higher pitch, represent the gym employee's words. As of line 101, she is evaluating the exchange as her present self. Thus, the vlogger uses editorial and multimodal features, that is, cuts and pitch variation, to replace quotatives indicating speaker shift. A fairly obvious and supposedly exaggerated instance of repetition occurs in the representation of the gym employee's speech: the clause "there's a place here" (92) is repeated in a shorter version "and here" (92 – 96) five times, and twice more: "over there" and "right here" (97, 99). Her speech is accompanied by points with a pen at places on a clipboard she had been holding in her role as gym employee. In this excerpt, it functions as an imaginary map that shows the location of other gyms operated by the same chain. The vlogger repeatedly shifts gaze between the clipboard and the camera, imitating the gym employee as he or she was supposedly seen by the vlogger while the conversation took place. The gym employee probably did point out several locations, but it seems unlikely that it was done in such a brief manner and that it involved that many locations. This list-like repetition is used to characterize the gym employee as persistent and nagging in the attempt to convince the customer to keep her membership. The exaggerated depiction of this incident as time-consuming and annoying adds a humorous note to the narration of an otherwise not overly exciting subject matter, inviting the viewer to

evaluate it in terms of their own experience with that sort of service encounter. This happens more explicitly in lines 104 and 105, "don't you hate that when people try to-, try to convince you to :.. give money to them.", which is a question asking for agreement.

5.7 Savoring

Another function of repetition identified in conversation is savoring, which can mark appreciation of humor (Tannen 1989: 64). One interlocutor repeats another interlocutor's (humorous) remark, thereby hinting at the enjoyment gained from it. There is no example of humor appreciation through repetition in my data set, possibly because speakers preferably show appreciation of another person's humor rather than their own to avoid appearing conceited. However, there is savoring through repetition serving other functions. The following excerpt contains an example of savoring that has a didactic function. The speaker is a young woman who is stressing the importance of drinking sufficient amounts of water generally, and especially when dieting.

Example 5.8

57	literally I did not do anything wrong yesterday.
58	I drank uhm four bottles of water,
59	which translates to two liters.
60	which you need.
61	you need two liters of water every day.
62	two liters. {all syllables stressed, changed pronunciation, speaker holds two fingers up}
63	so I drank my two liters of water,
64	and then uhm,
65	uh last night I took uhm two glasses of uhm Fanta.
66	so that's- Fanta light.
67	okay so diet Fanta.

68   uhm and,

Lines 59-61 contain a chiastic structure: "two liters. which you need – you need two liters". The following repetition of "two liters" (l. 62) is pronounced in a different manner than the surrounding context, with stress on all syllables, the /t/ realized as [t] rather than a flap, and a vowel sound /a/. Thus, the pronunciation of "two liters" in line 62 is /litarz/, whereas in all other instances the speaker uses a General American pronunciation /ˈliɾɚz/. Line 62 is not only set off through its pronunciation, but there is multimodal emphasis by gesturing: the speaker holds up two fingers to underscore the amount. She appears to derive enjoyment from pronouncing the phrase, and savors it by making it a heavily stressed element through threefold repetition. The chiasm of lines 59-61 and the isolated repetition of "two liters" in line 62 are a digression from her narrative-like report of her dietary efforts of the previous day. The digression serves as a didactic element to profess her point, with the savoring pronunciation as a device to make it memorable to the audience. Lines 58 and 63 (also repetitive) mark the last element of the narrative before and the first after the digression and thus frame it. The switch back from didactic to narrative mode is also instantiated with the discourse marker "so" in line 63. Spoken language with such an overt didactic purpose is particularly addressee directed, inviting hearers to make sense of it and therefore co-create involvement. The viewers are quite obviously supposed to not only take in her story about her diet, but the vlogger wants her viewers to be active by understanding her point and how important it is to her. This teaching mode is possibly owed to the fact that, as far as the vlogger's comments on viewer feedback allow this judgment, she is primarily speaking to other people who are trying to lose weight through dieting. Thus she is appealing to a common interest she and her viewers share.

5.8 Expanding

Speakers in monologues use repetition to expand a topic, as they do in conversation. In Example 5.9, a young woman takes up a topic that she had announced, among others, at

the beginning of the vlog. The "ears" in line 69 therefore is not the first token of the word in the vlog, but represents a repetition itself. The first instance occurs in Example 5.9, which serves as a structural device similar to the outline of a talk.

Example 5.9

35	I have lots of things to say,
36	and I wrote four of them,
37	on my hand,...
38	ears,
39	Tokio Hotel.
40	hair,
41	Avery.

- - - - lines 42-66 not reproduced here

67	CUT
68	um.. so yeah.
69	ears.
70	you're probably wondering why I said.. ears.
71	to begin with,
72	uh,
73	I said ears becau:se,
74	I'm seventeen,
75	and.. I got my ears pierced,
76	at last.
78	you can't- I guess you can't see them.

Line 69 contains an isolated "ears.", which functions as a topic opener (l. 1, "um.. so yeah." seems to mark a transition between two topics). It reinstates the earlier announcement of the topic in line 38, and it is the basis for the metalinguistic comment that follows in lines 70-73 "you're probably wondering why I said.. ears. to begin with, uh, I said ears becau:se,". This comment invokes the viewers' perspective in that it

formulates a question to the vlogger. The vlogger employs this as a rhetorical device to create interest in her topic, allowing her to elaborate it in disguise of an answer. In line 73, the speaker recycles the phrase "I said ears" from line 70, using it as a scaffold to construct the actual information she wants to transmit in the answer to her question (cf. Tannen 1989: 65). Imitating a question answer sequence of a multi-party interaction, she increases tellability. Similarly, Norrick (1987: 260) analyzes same-speaker repetition in the form of adjacency pairs as follows: Speakers "use the felt bond between the question and answer to suggest a special status for their repeat". This perspective shift epitomizes Tannen's (1989) characterization of involvement as a state in which speaking and listening contain traces of one another.

5.9 Evaluation

Speakers repeat items from foregoing talk to evaluate their appropriateness in that context. Example 5.10 contains instances of negative evaluation, or distancing, from the word choice in a previous intonation unit. In this vlog, a young woman talks about the prevention of illness.

Example 5.10

     324   and ju- just feel gross and yucky,

     325   but I eat my medicine,

     326   "I eat my medicine", {eyebrows are lowered and raised again, smiles}

     327   I TAKE.. uh my vitamins,

     328   I eat good food,

     329   I drink a ton of water,

     330   I exercise,

     331   I don't know what I'm doing wrong.

     332   um so if you guy:s are the kind of person who never gets sick,

Line 325, "I eat my medicine," is lexically awkward, since "medicine" usually collocates with "take". The speaker notices this immediately and, with lowered voice and disapproving expression, repeats the phrase. Her repetition and the change in intonation and facial expression draw attention to a flaw in her wording, signify that she has noticed it, and express disapproval. The subsequent "I TAKE" (l. 327) stresses the preferred collocate and remedies the previous mismatch. The vlogger does not follow through with a complete self-correction, but continues her enumeration with another item which allows her to demonstrate that she knows what the preferred collocate is. The negative evaluation represents a shift to the perspective of the viewers in that it anticipates potential criticism of her language use.

This passage exemplifies another function of repetition, namely what Tannen (1989: 67) calls "evaluation through patterned rhythm". Lines 325 and 327-330 are delivered in listing intonation. Except for the pronoun "I" there is no lexical repetition. Instead, the rhythmic pattern is the same in each line. The meaning added through this device, beyond the mere lexical meaning, is that she does everything considered beneficial for her well-being to stay healthy. This relates to involvement in that the speaker lists four elements of the paradigm of healthy living, but she expects the viewer to understand that she is really referring to all of the things that are part of healthy living. Thus, there is extra interpretational work to be done by the viewer, namely applying their own knowledge about what is considered part of that paradigm.

5.10 Stalling

The production of a monologue requires a single speaker to do all the talking. This poses a greater challenge than participating in a conversation with regard to the production of speech. In conversation, speakers take turns, which means that during one speaker's contribution, the other only has minimal production efforts when back-channeling or signaling attention otherwise. Thus, there are breaks that allow a conversationalist to prepare the next contribution, which includes retrieving relevant items from memory. In monologues, there is no other speaker whose contribution serves as a break, which

means that the speaker is constantly expected to produce talk to fill the silence. Example 5.11 shows how a vlogger in his twenties employs repetition as a strategy to continue talking while searching for further elements belonging to the list of names he enumerates.

Example 5.11

67	so.
68	that's my new team,
69	it's Typhlosion,
70	Gengar,
71	Cloyster,
72	°let's see Typhlosion,°
73	°Gengar,°
74	°Cloyster,°
75	Nidoking,
76	Umbreon,
77	>what was the other one,<
78	Crobat.
79	alright,

The vlogger is listing the names of members of his team of Pokémon (characters in a video game of the same name). He can name three without hesitation: "Typhlosion, Gengar, Cloyster," (l. 69-71). Subsequently, he lowers his voice slightly and repeats those three names after which he names two more characters: "Nidoking, Umbreon,". The function of this repetition is to gain time: repeating the three preceding intonation units does not require any online production effort, it just requires a mere retrieval of a ready-made chunk from very recent memory. While this talk is being produced, the speaker can search for and prepare what to say next. In conversation, this stalling strategy is useful to keep the floor while making a longer contribution. In a monologue, on the other hand, there is no need to defend the floor, on the contrary, there is an

obligation to fill it. (The question in line 77 ">what was the other one,<" is another such strategy to stall the ongoing speech production.) This repetition thus shows the speaker's twofold awareness of the audience: 1) the speaker is quick to fill gaps, that is, there is someone else 'present' who could perceive silence as uncomfortable; 2) the speaker knows the viewers cannot help fill the gaps.

5.11 Episode bounding

Tannen (1989: 69) demonstrates that repetition in conversation occurs at the beginning of episodes, where it operates as a kind of theme-setting, and at the end of episodes, forming a kind of coda. Thus, repetition can function as bounding device in multi-party interaction, and this constitutes involvement insofar as it takes two parties to jointly negotiate when a topic is finished, when a new one can start, and what the new one is going to be. Theme-setting and coda are accomplished with allo-repetition of adjacent intonation units, making them the product of co-construction by several speakers. Repetition has a similar structural function in the bounding of episodes in monologues. However, of course it cannot occur as allo-repetition, but has to be instantiated as same-speaker repetition. In Example 5.12, a man in his twenties asks his viewers to contact him if they are interested in working with him on his project.

Example 5.12

        270   CUT/blending over

        271   one quick thing I forgot to mention,

        272   about my side project?

        273   (there are) lots of people commenting to me,

        274   lots of people are interested in what I'm doing,

        275   and lots of people excited about what I'm doing.

        276   however there are some common questions?

        277   that I'm going to quickly address right here.

        278   one?

279   doesn't matter how old you are,

280   two,

281   doesn't matter what country you live in,

282   and three,

283   doesn't have anything to do with their video.

284   so if you're interested at all what I work-

285   what I'm doing,

286   what I'm working on,

287   please send me a private message,

288   we'll talk about the details there,

289   like I said,

290   there is some money in it,

291   it does take a little bit of effort,

292   you have to sit down and talk with me,

293   and learn some stuff,

294   however it's not hard at all,

295   anyone can really do it,

296   so,

297   if you're intereste- interested,

298   please message me,

299   and we'll talk about the details THERE.

This episode is introduced by an announcement of a new topic: "one quick thing I forgot to mention, about my side project?" (l. 271-272). The beginning is then marked by lexical repetition of the phrase "lots of people" (273-275). The same three intonation units are delivered with listing intonation, creating the impression of a very large number of people. These lines represent a well-chosen thematic opener, since the episode is predominantly concerned with these 'people', that is, viewers, who are potential collaborators for the vlogger's project. The end of the episode is marked by the repetition of elements that are not adjacent: lines 284-288 are replicated in lines

296-299. They contain the main information, namely that work on the project will be negotiated through a different channel ("messages") with interested viewers. This is not an exact repetition, as there are certain changes. Lines 284-286 contain a cut-off and self-correction: "so if you're interested at all what I work- what I'm doing, what I'm working on,". In the repetition, this is reduced to "so, if you're intereste- interested,", resulting in a more elegant performance. Line 287 "please send me a private message," is also reduced in the repetition: "please message me," (l. 298). Although there is a considerable amount of talk that takes place in between the two renditions, the repetition in 297-299 has the function of closing the episode: the speaker stresses the main points in a more concise way and thus makes the most important parts of his talk more memorable. As pointed out by Norrick (1987: 256), "restatement of a key phrase can summarize any sort of text" and this enhances textual coherence.

Especially since the repeated passages concern future interaction between vlogger and viewers, and thus are directly relevant to the viewers, this repetition represents an involvement strategy. The crucial information interested viewers need is presented twice, making the vlogger's request for interaction more urgent. The summarizing function indicates to the viewers that the vlogger has covered the topic, and will proceed with something else in the remainder of the vlog.

5.12 Conclusion

The research presented in this chapter has illustrated how repetition is used in CMC monologues to create involvement. Several models of the concept of involvement were reviewed in terms of their applicability to vlogs. Subsequent analysis was predominantly orientated towards Tannen's (1989) model, specifically focusing on repetition as the main strategy.

Repetition in vlogs can either occur in the form of same-speaker repetition, or in the form of cited viewer comments from earlier videos. For vlogs, this rules out two functions of repetition in conversation: participatory listenership and ratification of

listenership, which are based on allo-repetition. Other functions of repetition in conversation, however, can be found in monologues, reflecting in a very basic way that monologic talk has parallels to dialogue.

Some vloggers employ repetition across several or all of their vlogs. Openings and closings, in particular, are subject to ritualized behaviors, including repeated phrases and gestures. These create familiarity and, depending on word choice, set the level of formality. The current analysis has revealed that the previously identified conversational functions savoring, evaluation, expanding, stalling, and episode bounding are all present in vlogs.

Some of these functions are realized differently than in conversation due to influences of the CMC context. Others are structurally quite similar to instantiations in conversation. The influences of the CMC context can be grouped into three categories: 1) there is no turn-taking; 2) editing; 3) video as medium. Same-speaker repetition in form of an adjacency pair occurs in conversation as well as in monologues, and in both cases it enhances the interactional force. However in conversation, speakers are interacting with a present audience that can actually provide the missing part of the adjacency pair. Because of the lack of turn-taking, speakers in vlogs have no choice: they have to play the audience's part in an adjacency pair and thus characterize their imagined viewers in some way. This is not only true for adjacency pairs, but generally for statement – response sequences, which influences the function of episode bounding in that repetition does not necessarily occur in adjacent intonation units, but with units in between, resembling a coda.

Editing can play a role even in the speech production/taping part of vlog making, as some speakers show explicitly or implicitly an awareness of their editing options. With this kind of awareness, repetition is assigned a function that can be associated with language acquisition, but not spontaneous adult conversation: repetition can be used to practice talk, especially when it is delivered as part of a performance that is supposed to create a certain effect. Many cuts in vlogs seem to result from a desire to remove material judged imperfect. On the other hand, editing can create exact repetition of

sequences from a vlog. This has the effect of emphasizing the content of the sequence and drawing attention to it through its similarity to a stutter or a broken record that repeats the same passage over and over.

Finally, the affordances of the medium video are exploited in combination with repetition, e.g. for the effect of playfulness and humor. This reflects creativity on the vloggers' part, which is evidence that the vlog format has by no means strict conventions, but fosters experimentation. Further research on repetition in monologues could fruitfully investigate other genres such as lectures, speeches or answering machine messages. The investigation of repetition in TED Talks presented later in this study demonstrates the influence the different kinds of editing options have on speech production.

6 Pointing gestures in video blogs[9]

6.1 Introduction

This chapter describes the forms and functions of pointing gestures that occur in video blogs. Vlogs are a monologic, audio-visual CMC genre of spoken language with frequent occurrences of multimodal elements such as gaze shifts, shifts in posture, shifts in facial expression, as well as shifts in voice quality, pitch, and pointing. Vloggers use pointing as a means of establishing multimodal reference between elements of the website by pointing at objects on the website situated outside the video screen, thus projecting a gesture created in a three-dimensional space for its future two-dimensional environment. This chapter describes and analyzes the forms pointing actions assume, what targets they reference, and how they accompany the vloggers' verbal output. Further, the analysis demonstrates how pointing actions are adapted to the environment of the particular website, and how that adaptation is conducted by the speaker.

As with all forms of talk and written text produced and recorded for an audience receiving it at a later time, vlogs represent language that is produced in a different semiotic environment than it is received in, forcing either one or all interlocutors involved to mentally bridge the gap (cf. Baron et al. 2005). Answering machine messages, for example, sometimes contain date and time of the call (provided by the caller or automatically by the machine), to facilitate the hearer's creation of the context. Similarly, written notes may contain evidence of the writer's awareness of a temporal shift: "When you read this, I'll be gone". Thus, interlocutors may need to make a deictic shift between the actual situation of reception and the projected reference world. The present data differs insofar as the speakers know in advance the surrounding scene (website) of language reception, and can therefore at least partially predict the future location of the objects on it. A detailed description of the surrounding website is given below (Section 6.3: The framework of the website).

---

[9] Based on: Text&Talk 33(1), Maximiliane Frobenius, Pointing gestures in video blogs, 1-23, Copyright 2013, with permission from De Gruyter.

6.2 Pointing, context and CMC

Pointing, or deictic gestures, constitutes a subclass of gestures used to indicate concrete or abstract referents (Krauss et al 2000, Efron 1941 (1972)). This section provides an overview of literature on pointing in different contexts relevant to the present study.

Pointing is a situated, interactive activity which contextualizes semiotic resources, that is, the interactants' bodies, talk, the properties of the surrounding space, the interactants' orientation toward each other and the space, and the larger activity etc. (Goodwin 2003; Kendon and Versante 2003; see Jones 2008 for a discussion of multimodal communication in a CMC environment). The current study adopts the approach of accounting for these resources not as self-contained wholes, but as complementary elements juxtaposed by interactive practices. As the data examined here was recorded with no interactant physically present, several of the semiotic resources (e.g. the audience's posture and gaze, their gestures, and eye contact with the audience) are not available to the speaker. Moreover, speakers talk into a camera, which influences their own body posture. Finally, the notion of space is significantly different in the current data than in the face-to-face interaction making up the data of Goodwin (2003; see below: The framework of the website). Thus, analysis will reveal what interactive practices speakers in vlogs employ as they combine several semiotic resources, and whether they differ from those in the kind of interaction Goodwin (2003) studies. The investigation of gestures in TED Talks, where the audience is present, shows what impact the audience has on the interactive potential of the communicative setting (cf. chapter 8).

Regarding the forms of pointing gestures, several distinctions are described in the literature. Clark (2003: 244) contrasts two basic types of reference: 'placing for' and 'directing to'. The first describes a situation where an interlocutor moves a referent into the visual field of the interlocutor, and the second refers to a pointing gesture that directs the visual field/attention of the interlocutor to the referent. This distinction applies to vlog data as well, where 'placing for' can be realized with means not available in face-to-face interaction, that is, a shift of the camera position. The limited focal range

of the camera restricts a vlogger's options regarding the form of his or her gestures, as for example described in Enfield et al. (2007). Working on data from Laos, they distinguish B-points, which are big in form and therefore require a whole arm length of space, and S-points, which are small in form. The two gesture types have different pragmatic functions: B-points provide informationally foregrounded information about the target ("where" or "which one"), S-points provide informationally backgrounded information. Their study uses face-to-face interaction as data (1737). The forms pointing gestures take on in vlogs are influenced by the frame the camera imposes. Depending on the distance (and zoom) between vlogger and camera, an outstretched arm may not be visible on the video footage. The current study reveals what forms of pointing exist in vlogs as opposed to face-to-face interaction, and what pragmatic functions can be assigned to them. A taxonomy of the forms of pointing gestures beyond those occurring in face-to-face communication represents a gap in the research on gestures and on mediated communication, which this study attempts to fill.

Some related research has been conducted on sign language (Keating et al. 2008). They discuss data where deaf people communicate via video telephony, using a setting where the interlocutors not only see the other participant, but at the same time have an image of themselves on the screen, allowing them to monitor their signs. They observe that "the ability to see oneself enhances the ability to understand the perspective of the other" (1075). Keating and Mirus (2003) demonstrate how deaf signers adapt some of their signs to the two-dimensional, restricted space to contrast them with similar signs. Some of the data discussed in the present chapter shows very similar features, as some vloggers monitor their own image on screen. These studies, in addition to the present one, reflect the notion of the "'reciprocity of perspectives' whereby each party assumes that the other has a perspective and that if one adopted that perspective the world would look as it does to the other" (Hanks 2005: 196). Only an approach that includes this assumption can account for the fact that vloggers point into a space they do not see but that will be seen by the viewer.

Jones (2004) specifically addresses the notion of context in CMC. He suggests adapting and further developing that concept from something first applied to face-to-face communication and written text to fit a communicative situation that involves not just what happens on computer screens but also to what is happening in a speaker's physical surroundings (cf. Duranti 1992 re the dynamic nature of context). People communicate in different kinds of spaces simultaneously, with different degrees of attention. Vlogs are an ideal type of data to approach an analysis which acknowledges the polyfocal nature of communication, because the physical space the vloggers occupy (i.e. the producers, not the recipients) is partially inherent in the genre, due to its visual component. I have no access to the physical surroundings of the vlogger other than what is visible and audible on the vlog itself, therefore this chapter does not explore the impact of the different kinds of context on vloggers' behavior systematically.

Research on gestures in monologues revealed that the presence of an (imagined) interlocutor has effects on the frequency and type of gestures. Bavelas et al. (1992) classify gestures into topic gestures and interactive gestures. The latter two help maintain the dialogic character of the conversation and prevent it from becoming one speaker's monologue, whereas topic gestures depict semantic information. Their experimental research illustrates that speakers used a higher rate of interactional gestures when an interlocutor was present than speakers who talked in the presence of a camera only (Bavelas et al. 1992). The "alone condition" in their experimental setting is described as a condition toward, but not at, the extreme end of a continuum between a "person in complete isolation … and a dialogue with all of its systemic demands" (1992: 477). As this condition strongly resembles the vlog situation, an assessment of the interactive force of pointing gestures in vlogs provides a continuation of research on the effects of visibility of the interlocutor, only in a non-experimental setting.

Bavelas et al. (2008: 499), investigating the effect of the monologue condition as a variable, report that "when they [the participants of their experiment] were to imagine that the partner would see their videotape (vs. only hear their audiotape), they gestured at a higher rate and used more gestures that were not redundant with their speech".

Thus, even an implicit audience induces gestures, which explains why vloggers, who are talking to an invisible, non-present audience, use gestures.

6.3 The framework of the website

This section provides an introductory overview of the two-dimensional environment a vlog appears in, as opposed to the three-dimensional space it is produced in. The vlog-website contains multiple additional features, whose position on the page is fairly constant. There are minor changes due to the infrequent re-designing of some of the functions or due to further development of the features (e.g. new video formats). To facilitate the reading of the transcriptions and their analyses in this chapter, this section describes the positioning of the website elements as it appeared when the data was collected (March 2010). Meanwhile, the design has changed, resulting in a different positioning of some of the elements. The setup is the same for every user and every video, and it is designed by the programmers of the YouTube website (see Figure 6.1).

The most salient property of the website space is its two-dimensionality, and the resulting arrangement of its elements. The title of the video appears in the upper left hand corner, above the video screen. It appears in black writing on white background (1). The video screen (2) appears underneath the title, also in the upper left hand corner. Next to the video screen in the upper right hand corner is the video description (3). It contains information about the video by the vlogger, and data provided automatically by the website, including a 'subscribe button' and the user's icon. Writing appears in black (and blue) on grey background. The rating system (4) appears underneath the video screen on the left. It is represented by red and grey stars, which the viewers can click on to submit a rating. In the bottom left hand corner there are two sections that host viewers' responses: (5) shows thumbnail images of video responses to the vlog in question, and (6), the comments section, lists written comments by viewers chronologically, with the latest one on top. Users who are members of the "YouTube Partner Program" have a variant setup in that there is a box containing advertisements above the video description, so that everything on the right hand side moves down.
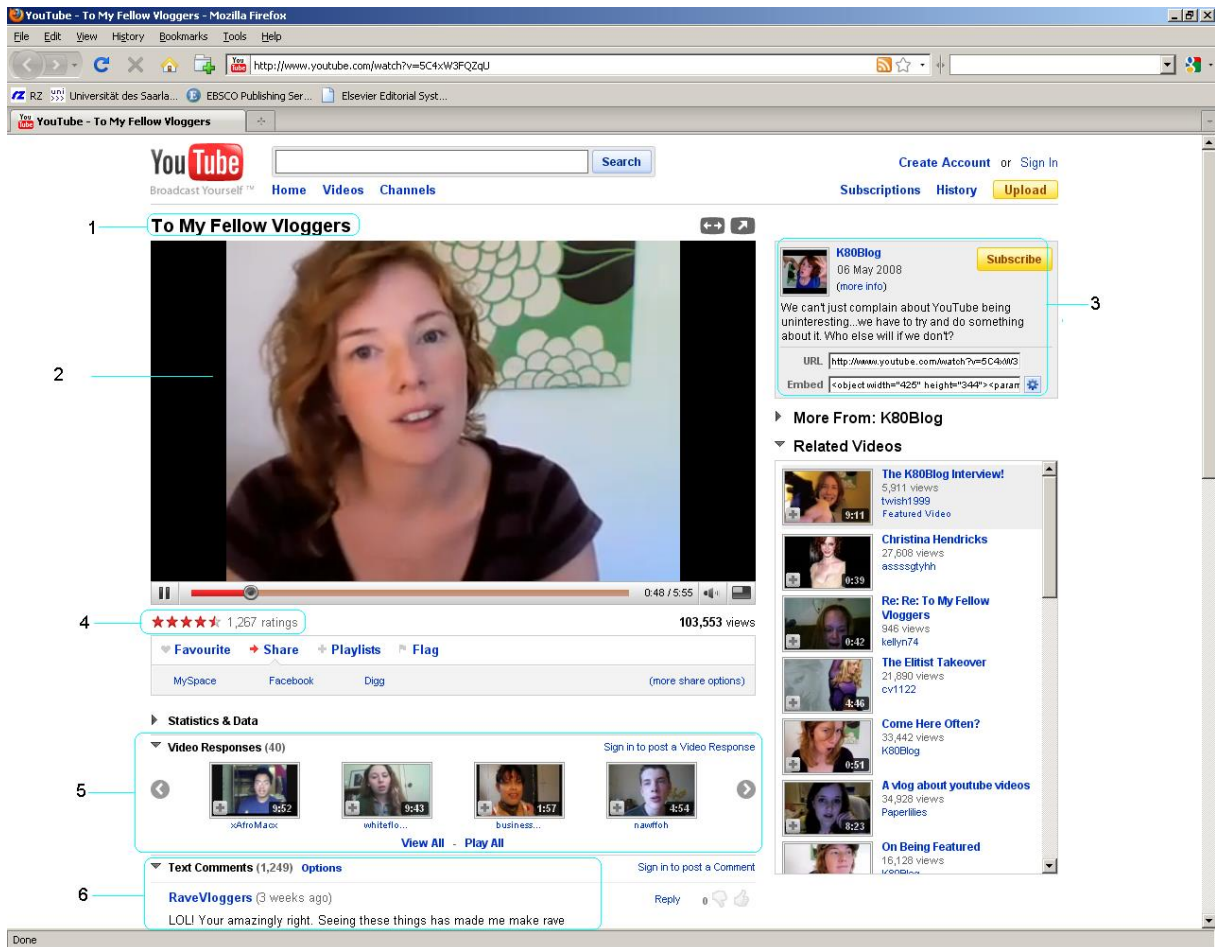
Figure 6.1, The setup of the website

As most vloggers sit up close to the camera (cf. the vlogger in Figure 6.1), the space that will later on be visible to the viewer is limited. As opposed to the situation in face-to-face conversation, viewers usually only see the head, shoulders, and part of the vlogger's chest (presumably vloggers film themselves close to the camera because sound quality deteriorates with increased distance). The limitation of space results in two immediate effects on the vloggers' pointing options. First, any gestural action that the viewer is supposed to see must be carried out within the scope of the camera; and secondly, there are only few referents within the scope of the camera.

6.4 Data collection and methodology

The approach chosen in this chapter aims first at revealing what pointing practices are commonly found in vlogs. Many of these pointing actions also occur in face-to-face communication. Secondly, the kind of pointing that is genre specific and thus possible only in the vlog setting are described. Therefore, my discussion also involves examples of practices that occur infrequently in my corpus, which are nonetheless of interest in that they illustrate the resources vloggers deploy to create reference. Considering that the conventions of the genre might not have fully developed yet, it is possible that some of these infrequent practices will become part of the standard repertoire at a later stage.

For each video, I noted down whether the vlogger used any pointing gestures or 'placing for' in the sense of Clark et al. (2003), and if so, at what frequency. I distinguished between the following categories: 1. pointing with the index finger, thumb, or whole hand towards a referent that is located in the speaker's surroundings, including immaterial referents and pointing at the camera when referring to the viewer; 2. pointing towards a referent that is located on the surrounding website, but outside the video screen; 3. holding a referent into the camera's focal range ('placing for'), including emblematic gestures as a subcategory (3.1); 4. camera shift towards a referent. Additionally, I counted the number of cuts a video contained; this number is somewhat problematic though, as not all cuts involve a change of referent, e.g. when they are used to edit out dysfluencies. Therefore, cuts are not investigated systematically in this work.

The 40 vlogs yielded the following statistical data: 18 were done by female, 22 by male vloggers, the shortest vlog was 1 min 07 sec long, the longest 10 min 50 sec. Out of 40 vloggers,

1:	28	vloggers point to a referent in their surrounding at least once	(16);
2:	10	vloggers point to a referent on the website at least once	(4);
3:	19	vloggers hold a referent into the camera at least once	(37);
3.1:	17	vloggers make an emblematic gesture/iconic symbol at least once	(12);
4:	17	vloggers use camera shift at least once	(27);

25 vloggers use at least one cut to edit their video.

Numbers in parentheses represent the highest count in a single vlog.

Referents in category 1 range from objects such as video games, clothes, the speaker's body parts, which are visible for the viewer, to objects and places that are outside the viewer's (and sometimes also the vlogger's) view. Finally, they include vague referents, such as a non-present group of people referred to as "they", pointed at with the thumb in the direction of the speaker's back, or even the viewer, referred to as "you" with an index-finger point towards the camera.

Category 2 contains points to a restricted number of referents, the most frequent being the space below the video where viewers can post written comments. This is followed by points to the video description, where vloggers leave links. Other possible referents are the rating system, video responses, the title, advertisement, etc. The kinds of points in this category cannot occur in face-to-face communication, as it involves referents only available in online communication.

Categories 3 and 3.1 are specific to communication that is mediated with the help of a camera, which has a limited and relatively static visual field. In category 3, objects such as clothes, cosmetic products, food etc. are held into the view of the camera, usually accompanied by the vlogger's comments about the respective referent. Category 3.1 differs insofar as the referent is another gesture which is not talked about; in some cases in my data it seems as if the speaker is not even particularly careful to ensure that the gesture is visible to the viewer. Thus, the level of consciousness with which the gestures are made sometimes varies between the two categories. Gestures often used are the peace sign, 'thumbs up' and number gestures. Categories 1 and 3 frequently occur in combination, such as when a speaker holds an object into the view of the camera and then points at it (or a particular part of it) with the index finger.

Category 4 contains another genre-specific phenomenon, that is, a version of 'placing for', where referents are not held into the camera, but the camera is shifted so the referent is visible. I consider this 'placing for' because, when the vlog is watched online,

the viewer's visual field, that is, the direction of his or her gaze, remains stable. A small subsection of the visual field, the video frame, however, is subject to change, making this kind of reference a hybrid of 'directing to' and 'placing for', which does not involve the viewer's gaze shift. Thus, these are pointing actions that cannot occur in face-to-face interaction.

Based on this categorization, I now discuss selected examples to illustrate common pointing behavior in vlogs. In Section 6.6, I present examples of how vloggers reflect the design of the website in their pointing gestures and the accompanying speech, and demonstrate how vloggers adapt to an environment that at the time of speaking does not exist.

6.5 Common pointing strategies in vlogs

This section gives an overview of common pointing practices in vlogs, as identified in the previous section. Spoken language is represented in the left column, and gestures in the right of the transcription tables. Overlap is indicated with (numbered) square brackets.

The first example illustrates categories 1 and 3, which often co-occur. A woman in her thirties shows and talks about pieces of fabric sewn together by children.

Example 6.1

1	so uhm,	
2	I wanted to show you,	
3	some of the backs of these,	
4	because they are...	
5	they're all different,	
6	so like this one,	
7		holds nine patch close to camera, not covering her face

8	has little [stitches][1],	[turns nine patch 90° so she can look at it][1]
9		turns nine patch back to camera, with index finger of left hand pointing at a seam on the patch
10	and not-	
11	they're pretty spaced [apart,	[turns nine patch 90° so she can look at it, as above][2]
12	you know,][2]	
13	some of these were probably [stitched][3] by four year olds,	[gaze shifts from nine patch back into camera][3]
14	[some][4] of them are stitched by older [kids][5],	[turns nine patch back to camera][4] [turns around nine patch 180°][5]
15	so here's the front of this little nine [patch][6],	[puts nine patch down][6]
16	and this one has,	

Figure 6.2, (silence after "little stitches")
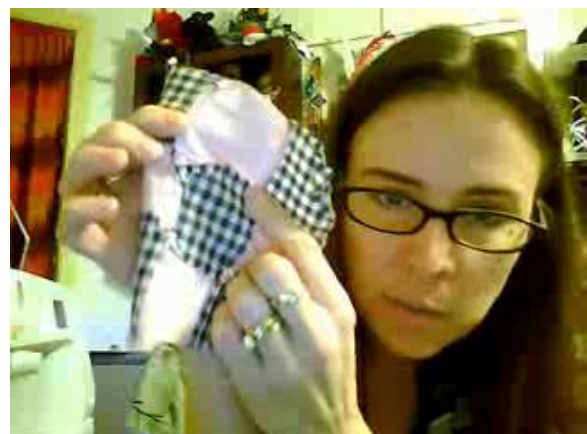
Figure 6.3, "spaced apart"

The speaker does not, unlike other vloggers, monitor the footage she tapes while shooting it, which partly explains why her gaze is not constantly fixed on the camera. The first instance of 'placing for' occurs in lines 6 and 7: the vlogger uses the proximal deictic term "this" and pronominal "one" to refer to the fabric in question. As this is not enough information for a viewer to identify the referent, the vlogger places the referent into the visual field, that is, the camera range. These two acts of reference do not occur simultaneously, but sequentially. This might be a consequence of the speaker's effort to place the patch carefully so as to make it sufficiently visible and not block the view of her face.

Lines 8 and 9 display a similar structure: first, the vlogger talks about the referent "stitches", then she refers to them with an index finger point. This time, the sequentiality is clearly a product of the complex task the vlogger sets herself. She has to co-ordinate speech, gaze, and gestures so that it is coherent to a viewer who sees this through the immobile view of a camera, without any chance of co-constructing an interaction. Therefore she has to regain orientation in the newly developed frame of reference: the nine patch. She does this by turning it by 90°, looking at it, and placing her index finger directly on the new referent (cf. Figures 6.2 and 6.3). With the finger positioned that way, she completes the point by holding the whole frame of reference back into the camera. If she were monitoring herself on the computer screen while taping, orientation could have taken place without moving the patch and radically shifting gaze (cf. Example 6.6).

In lines 11-13, the vlogger's gaze is mainly directed at the referent again, though it is unclear what the motivation is for this. As she is talking about the quality of the stitching, I presume she is double checking her evaluation and/or is planning what to say next. Lines 14 and 15, finally, show a pointing/speech reference combination in reverse order. This time, the speaker first places the front of the nine patch for the viewer into the camera (by turning it around), then she mentions it verbally, using the proximal spatial deictic term "here".

The two categories 1 and 3, which the previous example illustrates, are the most frequent types of pointing actions in vlogs. Especially placing objects for the viewer receives very high counts in single vlogs, which is a consequence of the fact that some vlogs are devoted to e.g. reviewing or unboxing new video games, or to presenting recently acquired clothes or cosmetic products ('hauls').

The next example illustrates category 4, where camera shifts are used to place referents into the viewer's visual field. A 25 year old woman, shooting her vlog outside, is talking about an ice cream man.

Figure 6.4, "and"   Figure 6.5   Figure 6.6, "ice cream man"

Example 6.2

1	we're gonna go see some bats?	
2	and we also (just got some) free ice cream?	
3	right here, (2)	camera shift away from speaker to ice cream van for ~ 2 seconds, then back to speaker
4	it's the ice cream man,	
5	he gives awa:y free ice cream?	

Pointing and verbal communication are sequentially ordered: there is only minimal overlap of the camera shift and the speaker's intonation units in lines 3 and 4. The biggest part of the camera shift is not accompanied by words at all. The proximal deictic

"here" (l. 3) induces the camera shift to the referent, which is only subsequently identified more specifically as the ice cream man. There is a clear distinction between 'placing for' in the original sense (according to Clark 2003), and 'placing for' as it is illustrated in this example. In this example, not only is a specific referent placed into an interlocutor's frame of reference, but the whole frame of reference undergoes a change (cf. Figures 6.4, 6.5, 6.6). In the example, the image is changed from a close up of the speaker to a scene showing an ice cream van, including the ice cream man. That allows for a certain ambiguity as to what the referent is until it is resolved by further reference (cf. l. 4). In this case one could argue that "ice cream man" is a metonymy representing not just the person but also the van with its contents, leaving it unclear what exactly the referent is.

The two previous examples both illustrated how vloggers import referents into the viewer's frame of reference. The next example belongs to category 2, as it features an index finger point to a referent that is outside the video frame, on the website.

Example 6.3

1	and um the Dickens Inn is on Saint Catherine's DOCK.	
2	you can look it all up,	speaker raises hand so it's visible; speaker's gaze shifts from camera to speaker's left + point to the left begins; gaze shifts back to camera + point continues (cf screenshots below)
3	I'll put a link over there for you okay,	point to the left continues, is repeated 6 times
4	so (you're all) gonna meet there SATURday,	
5	and um,	

Figure 6.7, "you can"

Figure 6.8, "look it"

Figure 6.9, "all up,"

Figure 6.10 "I'll put a link over there for you"

The referent of this short sequence is a hyperlink that the speaker is planning to put into the video description which will appear when he uploads the video to the internet. As he is pointing to a referent that is present when the speech/gesture segment is received, I do not consider this 'abstract pointing' (McNeill 1992, 2003, 2005, McNeill et al 1993), or "creating a virtual location" (Haviland 2004: 205). But it is clear evidence of the fact that the speaker understands that his present context and the viewer's context differ fundamentally. His knowledge of the website's organization enables him to adopt a viewer's perspective.

His own context, however, imposes restrictions on the the way the gesture is made. As the pragmatic function of this point is to identify the location of a referent, which is informationally foregrounded information, one would expect a B-point (Enfield et al.

2007)[10], which itself would entail a gaze shift towards the referent. The gesture under analysis, however, is small in form, and the gaze shift is extremely short (cf. Figures 6.7, 6.8, 6.9, 6.10). A B-point would be impossible and inappropriate since the viewer would not be able to see it as it would be outside the camera range. There is repeated back- and forward motion, which might be a compensation for the reduction in form. The gaze shift is not permanent for two possible reasons: one, the referent is not visible to the speaker since it does not yet exist; two, the speaker cannot interact with the viewer to check whether he or she has aligned body orientation/gaze to the referent as well. This puts the speaker at risk of talking and looking into the direction of a future space, while the viewer is still looking at the speaker. That would result in a mismatch of gaze and speaker orientation, which could induce the viewer to lose interest in the communication altogether. In Goodwin's (2000) terms, in this asynchronous type of communication the contextual configuration does not permit visible co-orientation. Why the speaker shifts his gaze in the first place is unclear; it might be an automatically triggered impulse.[11] Clearly, the present example demonstrates an adaptation of the pointing action which is specific to vlogs.

The first three examples served to give a general impression of the kinds of referents that are targeted in vlogs, and how that targeting is accomplished. The passages examined represent common pointing practices, including the adaptations necessary in this environment. The following section goes beyond describing pointing sequences found frequently in that it features an example where there are combinations of multiple strategies as well as shifts between the three-dimensional space the speaker occupies, and the two-dimensional space of the website. Furthermore, there are two examples by the same vlogger, which demonstrate how a pointing gesture can gain pragmatic load.

---

[10] Enfield et al.'s work is based on data collected in Laos; cross-cultural variation of gestures suggests that these findings might not be one-to-one applicable to American English data (cf. Kita 2009)

[11] In the subcorpus, the most common referent of this category is the comments section below the video (also cf. next example), which is pointed at nine times. In only two out of these nine instances, the speaker does not keep his or her gaze fixed into the camera, but it is slightly lowered. To make any substantiated claims about gaze in relation to the location of the referent (next to or below the video frame), a close look at much more data would be required; but there does seem to be a tendency for gaze not to shift if the referent is below the video.

6.6 Viewer adapted pointing

One aim of this study is to show how speakers of monologues interact with their non-present audience, how elements of their communicative effort are clear signs of audience design. While there is debate about gestures in general whether they mainly help the speaker express him- or herself, or help the interlocutor understand the speaker, or both, I work under the assumption that pointing gestures are strongly (though not necessarily solely) motivated by the desire to help another identify a referent. Definite reference, including deictic gestures, requires mutual knowledge to be felicitous (Clark and Marshall 1981), which, in the case of vlogs, is secured through the particular knowledge members of a community have, and through "potential physical copresence" (38) of the interlocutors and the referent involved in this online speech event. Co-presence in vlogs can only take place virtually for two reasons: (1) when the speaker points at a referent outside the video frame, that referent does not yet exist; (2) when the speaker points at a referent, the viewer is not yet present to direct his or her attention to the referent. Thus, virtual copresence involves a temporal and spatial shift that the genre vlog can compensate due to its setting in a multi-channel website. The analysis of the following examples focuses on how gestures act as interactive devices in various ways.

The two examples that follow (6.4 and 6.5) are both by the same vlogger; they show a very similar pointing gesture, an index finger point with the right hand, pointing downwards at the comments section below the video (category 2). In both cases, the vlogger (female, 20-30) asks her viewers to leave written comments concerning particular topics.

Example 6.4

1		claps hands
2	but I've been watching *true blood* back to back,	hands stay within camera range, palms held together, move in that position from left to right

3	I'm I'm terrified	hands clasped
4	at the fact that I am but,	hands apart, held up at chin level, backs facing camera, fingers slightly spread apart
5	let me know what your shame show is,	hands below camera range, seem to be open, palm facing up, fingers towards camera
6	leave it in the comments,	index finger point downward, one motion starting at about chin level to halfway below camera range
7	I'm very curious to find out	hands folded
8	and um	

Figure 6.11 + Figure 6.12, "let me"

Figures 6.13, 6.14, 6.15, "leave"

The speaker employs various gestures during the sequence presented above, two of which are accompanied by speech that addresses the viewer directly. The first one (l. 5) has elements of a conduit gesture, but also of a deictic toward the camera (cf. Figures 6.11, 6.12); thus the gesture captures both the address of the viewer (expressed in speech via the directive "let me know" and the possessive pronoun "your") and the transfer of information she requests. The next gesture/speech unit is a continuation of this, in that the speaker specifies through what channel she expects the requested information (the website offers various alternatives of communication). This point, like all category 2 points, demonstrates the speaker's ability to imagine the viewer's perspective: while the referent is invisible to her, she actually directs her index finger towards the floor (cf. Figures 6.13, 6.14, 6.15).

Unlike the vlogger in the previous example, she does not prepare the finger point with a gaze shift. Strictly speaking, these gestures would not have been necessary for the speaker to be sufficiently comprehensible; but the expression of one thought through two semiotic systems simultaneously turns her utterances into a rich, multimodal code that expresses a desire for interaction with the viewer, and not just regarding the content, but also the manner of communication.

Goodwin (2000a: 1498) shows that "a quite different kind of work, involving the precise deployment of semiotic resources with properties quite unlike the structure of speech, is required in order to build social action with the gesturing hand", explaining congruent content in talk and gesture in talk-in-interaction. A concept that is understood to encompass an "interactively organized process" (1492), action is not easily applied to the data discussed here. Clearly, the semiotic resources of the recipient, crucial in the negotiation of the unfolding of social action, are imagined and reconstructed by the speaker.

A speculative explanation for this apparent redundancy of speech and gesture is the speaker's experience of the building of social action in talk-in-interaction. If there were an interlocutor present, this gesture would induce co-orientation to the referent, and thus enhance the sincerity of the request for information. This, in turn, would

emphasize the interest she takes in her viewers, an inherently social behavior. If experience is the reason for this (and other) points, then that would suggest that monologues are, to some extent, modelled on talk-in-interaction, rather than representing a completely original mode with underived sets of conventions. This confirms the hypothesis stated in the introduction of this investigation. It is further supported by Schegloff (1987: 222), who claims that "speech exchange systems, and their turn-taking organizations, are the product of transformations or modifications of the one for conversation, which is the the primordial organization for talk-in-interaction". Likewise, Haviland (2007: 150) states: "Interaction is a compelling model for talk, even apparently monologic talk".

The next example features a very similar point (downward deictic), but it occurs in a different gestural surrounding, and it is coupled with very different speech. The four gestures preceding the point are beats carried out with both hands (cf. Figures 6.16, 6.17, 6.18). (beats occur at bolded syllables)

Example 6.5

1	and **you** will be	both hands move down to below video range, palm facing down, come back up
2	a **hun**dred percent fine	both hands move down to below video range, palm facing down, come back up
3	so **good** luck	both hands move below video range, palm facing down,
4	**let** me know how you- how you go.	come back up, move down
5	so	index finger point down
6	keep me up to date	hands below camera range
7	uhm	

Figures 6.16, 6.17, 6.18, "good"

Figure 6.19, 6.20, 6.21, "so"

In line 5, the vlogger makes the same downward deictic gesture towards the comments section, coupled with the discourse marker "so" (cf. Figures 6.19, 6.20, 6.21). In contrast to the previous example, there is no verbal mention of the referent. The gesture/speech unit is, however, preceded and followed by a directive ("let me know", l. 3, and "keep me up to date", l. 6), both of which function as requests for the transfer of information. Thus, in this example, the gesture conveys necessary information that is not supplied verbally and without which the meaning would not be completely inferrable.

"So", generally used as a marker of result (Schiffrin 1987), seems to fulfill several functions here. With regard to speech only, it brackets the preceding and the following intonation unit. There is no significant logical shift that would suggest a causal relation between 4 and 6 ("4 so 6"), however. With regard to the gesture, though, "so" seems very appropriate as it seems to express roughly the same as it expressed in the previous example: "leave it in the comments," (l. 6). That meaning would induce a shift between two elements with a causal relation, one a unit of speech, the other a gesture. As this of course does not fit the original definition of discourse marker, and especially because "so" occurs simultaneously with the second unit of meaning, this

interpretation is still somewhat speculative and would require more similar data for substantiation.

In comparison to Example 6.4, the gesture in Figures 6.19, 6.20, and 6.21 carries more meaning independent of speech and thus has a higher pragmatic load. The speaker can only use it in the way that she does because she relies on her viewers' experience of countless previous vlogs by the same user that feature this gesture, which helps them recognize this particular practice. She can assume her reference to be felicitous for those viewers who share this mutual knowledge, which in this case is based on the membership of the community of regular viewers of this vlogger's videos who are familiar with the standard design of the website and the location of the comments section underneath the video window.

Another feature of this user's vlogs makes the request for comments an interactive element: the vlogger incorporates selected comments posted below the previous video into the new vlog in written form, which she then comments on in a voice over. For many viewers (as can be concluded from some of the comments) this is an incentive to post comments themselves. Thus, a single gesture by this vlogger can trigger the activation of a communicative history, as well as the expectation of future communication following the same pattern.

The next example contains clear traces of audience design in a less obvious way than the request for interaction in the two previous examples. There is a combination of categories 3 and 1: the speaker combines a camera shift (l. 1-7) and carries out several pointing gestures during this shift.

Example 6.6

| 1 | I am eventually going to [ha:ng], | [moves camera up, top third of head is visible] |
| 2 | a curtain from and I can't reach, | speaker's left arm moves above head, points behind speaker |

3	uh- there we go.	switches to right arm, index finger points up
4	right along here.	right arm still points up
5	spanning this who:le.. width.	whole hand sweeping motion from left to right (speaker's perspective)
6	((coughs))	
7	so I can separate my kitchen and my living room,	camera shift back
8	and I can put some of my big heavy duty appliances like this,	points with her left hand (index finger) behind herself, searches for referent, arrives at blender
9	cause I'm really short on counterspace.	stable point at blender
10	I can run those along the kitchen side of the curtain.	pointing finger moves from right to left (speaker's perspective)
11	and still have room to prepare meals.	switches from index finger to thumb point, hand stays behind speaker

Figure 6.22 "a curtain"

Figure 6.23, "there we go"

Figure 6.24, "like this"

During the whole passage, the speaker keeps her gaze fixed on the camera/computer unit. As an alternative to checking on the direction of her points in the original space they occur in, she follows them as they appear in the video footage. This can be concluded from her meta-comments about the success of her gestures (l. 2/3, cf. Figures 6.22, 6.23). As a result of this, she adopts – at least in terms of the gestures – a viewer's perspective. Her pointing gestures are transformed to fit not the three-dimensional space they originate in, but the two-dimensional space they will be received in. In contrast to the previous examples, where this shift occurred because of a shift between the video frame and elements of the surrounding website, here, the same happens within one and the same frame of reference, the video. Knoblauch (2008: 84) observes a similar pointing technique in power point presentations: the speaker can "point into the rays projected by the beamer so that the shadow may serve as a two-dimensional indicator".

The speaker herself indicates her adaptation verbally. Reaching/touching a referent turns a point into a proximal gesture (Bangerter and Oppenheimer 2006), but is not required for a point to be successful in three-dimensional space; interlocutors usually understand that the (imaginary) extension of the pointing device identifies the referent. "I can't reach" (l. 2) is evidence that, in her adopted view (three dimensions mapped onto two dimensions), pointing works differently. The switching of the pointing arm represents a false start and subsequent self-correction.

More evidence of the perspective change can be found in lines 8 and 9, where the speaker points at a blender that is located behind her, thus again the third dimension (depth) is relevant (cf. Figure 6.24). However the speaker still opts for a two-dimensional point that produces overlap of referent and index finger. In the three-dimensional space of her living room, the vlogger's point would miss its target.

This example demonstrates vloggers' ability to master complex pointing tasks while speaking. Here, a speaker monitors herself on screen, thus adopting a viewer's perspective. The resulting adaptation of the form of the gestures can be considered a sign of accomodation to the viewer. The fact that for some viewers, the vlogger's points may be less helpful than desired does not diminish the effort made. However, successful communication requires, in turn, the viewers' efforts to understand what frame of reference the vlogger is currently in.

6.7 Conclusion

This chapter has pursued two main purposes. The first part was designed to reflect the kinds and frequencies of pointing gestures (including 'placing for') in vlogs. There are four main categories that were focused on in this chapter: (index) finger pointing inside the vlogger's physical space (1), (index) finger pointing into the two-dimensional virtual space of the website (2), placing objects into the camera's visual field ('placing for') (3), camera shifts (as a version of 'placing for') (4).

The spatial constraints of the medium video have an impact on the pointing gestures in vlogs. Pointing gestures must be made within the video frame if the vlogger wants the viewer to see them, thus they usually are small in form. The video frame does not allow for very many referents, and it is static (unless the vlogger actively moves it). Thus, a crucial component of face-to-face interaction is missing: co-orientation of gaze. Therefore, placing referents into the visual field of the camera/the viewer (which are distinct subcategories of 'placing for') is a common practice in vlogs. Often, vloggers

employ a combination of the 'placing for' and 'directing to' when talking about particular aspects of the referent.

Speaker's gaze differs from that in face-to-face interaction, since mutual co-construction is impossible. The spatial situation of this genre also saw the creation of a unique kind of pointing, that is points that transcend the boundaries between various spaces this communication is situated in. As a result, vloggers point from a three-dimensional space to referents in a two-dimensional space, long before these referents even exist. Thus, these gestures create both spatial and temporal deixis.

The second part of this chapter discussed how pointing gestures can contribute to making a monologue a viewer-directed performance. Pointing gestures, if they draw on the experience of a viewer community, can gain in pragmatic load so that the surrounding talk does not need to be entirely explicit. In the two cases discussed, these gestures helped emphasize the speaker's interest in her viewers and her request for information from them. They specified what particular communicative channel the speaker expected that information to come through. Therefore, this kind of gesture represents relational work. The fact that the speaker points at referents invisible to her, but present to the viewer, signifies accommodation.

Vloggers also adapt their gestures to the website environment with different means, that is, they adopt a viewer's point of view by monitoring their output on screen while shooting the vlog. This results in gestures that transform three-dimensional space into website, or two-dimensional space: they lack the depth vector. This is compensated for by overlap of gesture and referent on the screen ('reaching'). Although virtual online space is, of course, not a real, physical space, these kinds of gestures are behavior that indicates people's attempts to inhabit such a space, at least temporarily, as vloggers abandon the meaning these gestures have in their physical surrounding. Such an adaptation to virtual space, which is the space vloggers share with their viewers, is evidence (among many other factors) that vloggers try to interact with their viewers. Possibly, these gestures represent an attempt at social action adapted to the monologue situation.

7 Commenting practices in vlogs: How written interaction is tied to spoken monologues

7.1 Introduction

Social organization and the accomplishment of social order is a joint achievement enacted in great part through conversational practices (Garfinkel and Sacks 1970). In conversation, the turn-taking mechanism allows speakers to demonstrate behavior that offers interlocutors instructions for subsequent actions, and at the same time this behavior reflects how prior contributions were understood as part of the ongoing interaction. In other words, a turn at talk provides guidance as to how a next possible turn can look so as to be a sensible contribution to the joint communicative effort. Any subsequent turn at talk alludes to prior turns and the guidance contained in them to produce an ensemble of orderly activities over time. One technique that allows speakers to bring about order in conversation is 'tying', which acts as a display of understanding: "Tying an utterance to an utterance is the basic means of showing that you understood that utterance" (Sacks 1995: 718).

These observations are based on conversational interaction. Investigations of practices that allow the creation of social order in other settings would be a reasonable next step towards broadening our understanding of the organization of human (communicative) behavior. The communication mediated through YouTube's commenting function is analyzed in this chapter. Here, vlog viewers respond to vloggers by posting written comments underneath the video frame of a vlog in reaction to the video. This chapter seeks to describe such comments as orderly practices. Certainly, a complete account of the interaction initiated by vloggers using the particular affordances the genre video blog offers them, as is done in this investigation, must include a section devoted to what actually moves viewers to respond as expressed in their written comments.

Obviously, viewers' responses can take various forms, and they are mediated through all kinds of channels (see Lange 2008). Only one of these channels is readily available

for analysis, however, and therefore my only evidence consists of those responses that are manifest in the comments section that is part of a YouTube vlog. Thus, I cannot report on the most immediate reactions that viewers display, such as laughter, gestures, interjections and/or other verbal contributions that occur during or right after watching the video. I also cannot say anything about reactions in the form of offline (i.e. oral, face-to-face) communication between several viewers, e.g. evaluation of the video. And thirdly, I cannot comment on responses to the vlogger that are mediated through private channels, for example email, phone calls, IM (Instant Messaging) etc. In other words, the data investigated in this section consists of instances of written communication that are (supposedly knowingly) posted to a public forum, accessible not only to the vlogger, but to all other viewers. It represents but a fraction of the whole set of reactions to a vlog, but to maintain a methodologically consistent approach it is necessary – and considering the consequent narrowing of the scope of this section – even desirable to focus on one specific subset of data. Since the written comments appear on the screen below the video, and thus constitute part of this particular genre, they are an obvious choice for the present analytic considerations.

The methodology this chapter is based on is in line with conversation analytic (CA) and ethnomethodological research in that it seeks to make arguments derived from members' methods. Thus, analysis is carried out in such a way as to reveal interactants' knowledge and application of behavioral conventions as evidenced in their actions, rather than an outside analyst's understanding. In this vein, the spoken, monologic vlog is seen as an initial act, the starting point of interaction, which is followed (chronologically and spatially) by a potentially indefinite set of written comments. Sequential analysis of selected passages reveals which relevance structures exist between vlog/comments and comments/comments in an attempt to gain an understanding of how this communication is organized. This should yield particularly interesting and novel insights, because this investigation is concerned with interaction that encompasses two modes of communication, the spoken and the written, in a fairly new contextual configuration, that is, YouTube. This chapter therefore presents

pioneering research into how people employ these affordances to organize their interaction in general, and how they tie written language to recorded audio-visual data in particular. It is fundamentally derived from observations of human interactive behavior, such as the following remark about writing: "Another local social dimension is how a given writing shows a particular 'recipient design,' i.e. how it is tailored for the person receiving or reading that item, in just those circumstances" (Watson 2009: 19).

7.2 Research on YouTube comments

Communication on YouTube in general, and comments in particular, have been investigated from various angles, both quantitatively and qualitatively. Studies have treated the functions of commenting on YouTube regarding the community and as a system for feedback and support (Chau 2010), the effect of sentiment contained in comments on the rating of a comment (Siersdorfer, Chelaru and Nejdl 2010); the reception of YouTube videos according to varying audiences (O'Donnell et al. 2008); antagonism and 'flaming' as a function of social or culture dynamics, rather than anonymity (Lange 2007).

The notion of 'affinity spaces' (Lindgren 2012) is at the heart of the theoretical approach chosen in this chapter in that it represents a concept that can be applied to contexts such as that of YouTube comments sections, explaining that they are a "space for structured activity, no matter the number or heterogeneity of the actors" (164). Further, Lindgren states that "the relatively disinhibited climate in the threads allows for extreme discourse but that regardless of this, the negotiation of social rules as well as the formation of coherent discursive patterns still take place" (155). The "structured activity" and the "formation of coherent discourse patterns" are exactly what the present study focuses on, aiming to uncover in much more detail how this negotiation is executed visibly and accountably. In a similar vein, Jones and Schieffelin (2009) characterize YouTube comments as "tend[ing] to retain a high degree of topical coherence, if not a cumulative progression or structure of responsive turn taking" (1062f.). While turn-taking is a concept developed from and for the description of

spoken interaction, and thus I consider the mechanism it represents and the attached terminology only partially applicable to written interaction, there certainly are features in YouTube comments sections that are similar to those of conversation.

The *cumulative progression*, that is in case of the present chapter the progression and coherence created from the video to a comment left by a viewer/ created between the video and a comment left by a viewer, and how these elements together are demonstrably considered orderly action by the interactants forms the main theoretical contribution of this study. Walther et al. (2010), characterizing commenting as *reactive exchanges* (as opposed to *interactive* ones), observe that "YouTube viewers tend to post comments about videos; they seldom respond to other comments and almost never comment on comments about comments" (471). However, it is not entirely clear from their study whether this observation is empirically based on a corpus of YouTube comments, and if so, if this is a representative sample. Some videos are surely more likely to foster extended interactions in their comments sections than others. At the same time I agree with this observation comparatively speaking: other platforms, e.g. Facebook, seem to encourage longer exchanges despite the fact that – at the time of writing (Jan 2013) – comments sections on Facebook do not offer nested comments (or a reply function within comments), whereas the YouTube comments sections do.

In CMC in general, disrupted adjacency is a problem that interactants solve using various strategies (Herring 1999). The concept of adjacency here is borrowed from research in conversation, where it describes a sequence of two spoken turns where the second one immediately follows the first – thus it refers to organization of speech through time. In written CMC, adjacency is adapted to mean spatial immediacy, that is two written contributions (such as a comment or posting in a forum) occurring next to or below one another. In multiparty CMC, adjacency is much more at risk than in spoken conversation, since the visual and permanent nature of the exchange allows for contributions to visibly occupy the immediately adjacent slot to a first contribution that another user wants to post a reaction to – even though the contribution that actually occupies the slot might not be in reference to the immediately prior one.

The strategies to circumvent this problem and thereby maintain cohesion include the following: 'addressivity', which describes the practice of prefacing a contribution with the intended addressee's user name (Herring 1999; Werry 1996; also see Honeycutt and Herring 2009 for the use of the @-sign on twitter); 'linking', which refers to an explicit reference to a previous contribution; 'quoting', which creates the illusion of adjacency (Herring 1999, Bublitz and Hoffmann 2011).

Regarding the quality of the link between comments and videos, Jones and Schieffelin (2009: 1063) have observed that "contributors create an aesthetic or stylistic coherence by mirroring the content and genre of the commercials themselves". This, however, was based on reactions to TV commercials that were produced professionally, and in addition to being aired on TV, they subsequently uploaded to YouTube.

Video blogs, on the other hand, represent user (amateur/nonprofessionally) generated content, showing a private persona who, supposedly, also manages the YouTube channel where the vlog appears, and thus can be considered a person "known" to the viewers and potentially responsive to text comments. I assume (based on the data viewed so far) that this creates quite a different communicative atmosphere, making vlogs a site of interest for the investigation of coherence across modes. In other words, while a vlogger, as the creator of the video content, can be addressed and responded to as such, the uploader of a commercial is often not the initiator of interaction, but a mediator. Therefore, I suspect that comments sections of commercials contain fewer, if any, direct responses to issues raised in the video addressed to the uploader, whereas issues in vlogs, such as questions or opinions are possibly discussed more frequently with the vlogger him- or herself, making the vlog the initial act of communication.

7.3 The YouTube comments section – design and affordances

In the following, I describe the current (2012/2013) design of the YouTube comments section, explaining the affordances offered to users. Comments left underneath videos appear in a specially designated area under the video, its title and a partly collapsible text area that contains the upload date and a video description, as well as information on the uploader's channel and the video category. In principle, comments appear in chronologically reverse order, with the newest one at the top. Thus, viewers can see the newest comments without much effort, while they have to scroll down and click on a link to see older comments. This has a strong impact because viewers apparently often do not go to the extra effort of reading previous comments, as has been observed by Jones and Schieffelin (2009: 1063): "The redundancy of the comments suggests that posters do not read more than the most recent comments, as the same queries and opinions recur". While I partially agree with this observation, some of the data I have screened so far shows such a high occurrence of redundancies that there must be other reasons than the disappearance of comments from the initial comments page over time.

A comment is headed by the author's user name and the thumbnail of his or her profile picture appears to the left. These two elements are added automatically and invariably, making self-identification among interactants unnecessary.[12] An indication of how long ago the comment was posted (not particularly accurate, cf. "two years ago" in Figure 7.2 below) and a rating system in the form of two hand-shaped icons ('thumbs-up' and 'thumbs-down') and their tally, if applicable, are also included automatically in each comment. Finally, there is a 'reply' link, which allows users to

---

[12] Of course, YouTube users can and do sign up without giving their real names, and without uploading a picture, making it possible to conduct anonymous interactions on the site. And in this respect, a YouTube user who does want to disclose their real identity but operates under an alias would have to do extra work to do so. However, when I say self-identification is unnecessary I mean that, for the present purpose of assigning a comment to its author, nothing else needs to be done, even if the author's identity remains opaque. Compare, for example, phone calls (before the advent of caller-ID), where the parties involved had to do extra work to identify each other. This type of work is done automatically in this setting.

express through visual means that their own, pending contribution is a reaction to this specific previous comment.

A comment visibly left as a reaction to another is automatically marked as such either by two spatial alterations, namely (1) the reply appears directly underneath the comments it corresponds to and (2) it is indented (cf. screenshot); or it appears as other comments do, too, with additional text and hyperlink "in reply to *username* (Show the comment)". Clicking the "Show the comment" link will then change the appearance to the first version, where the indented reply is shown directly underneath the comment that received the reply. Note that while the general appearance of YouTube comments, that is, chronologically reverse, is counterintuitive at least for readers and writers of English texts, in that we are used to reading, and trying to make sense of, texts from top to bottom (and left to right), this particular design feature flips the order to the familiar chronological appearance. This radical break within a subsection of the larger comments section indicates that YouTube's web design is partially adapted to our established spatial grammar, while, at the same time, it reflects the consideration of other factors. So instead of asking why some comments appear in chronological order, one should probably ask why not all comments appear in this way which helps readers recognize coherence. Apparently, there is a preference among the site designers for certain comments, that is, the latest comments, to be visible on the screen in proximity to the video, which is greater than the desire for a facilitation of the reading of a comment exchange.

The reverse chronological order is also affected by several other factors. The comments section is further subdivided into 'Uploader Comments', 'Top Comments', and 'All Comments'. The top one, 'Uploader Comments', contains comments (both original ones and replies to others) by the user who uploaded the video, as well as those comments by others that the uploader has replied to (cf. Figure 7.1). The section underneath, 'Top Comments', presumably contains those comments that received the most 'thumbs-up' ratings. Finally, 'All Comments' contains those comments that do not stand out by virtue of authorship or rating.

America's TOP TEN?

Zipster08 · 1,100 videos

▶ **Subscribe** ⟨ 30,262

1,167

👍 117   👎 3

👍 **Like**   👎

About   Share   Add to   🗃   ılıl   🏴

Published on Nov 1, 2012
I'm losing touch! Are the Rolling Stones really worth an $800.00+
ticket?

————————— Show more —————————

Uploader Comments (Zipster08)

1 month ago
Just wondering, how old ARE you?
Reply · 👍 👎

    Zipster08   1 month ago
    Old enough to be your dad and young enough to kick your ass!! ;-)
    Reply · 👍 👎   in reply to Donny Potenza

1 month ago
Fuck Obama.
Reply · 👍 👎   in playlist Uploaded videos

    Zipster08   1 month ago
    Too late!!
    Reply · 👍 👎   in reply to DrewMcIntyreBrand

1 month ago

Kiss was on the list? Really? For what? lol Love you Zippy!

Reply ·

Zipster08   1 month ago

They have a new cd in the top 10!!

Reply ·   in reply to dawnten

Top Comments

1 month ago

About a year after my parents were married (1971 or 72), my mother found my father's stash of Hustler magazines. She sat down with a pack of construction paper, a pair of scissors and a glue stick. Over the course of about two days, while he was at work, she made bikinis for every naked woman in every magazine and glued them in place. She put the magazines back in their hiding spot and waited. It took him about a week to find them, but when he did... he was totally embarrassed.

Reply ·   4

All Comments (101)

click to leave a comment

weeks ago

On another subject, Have you ever listened to the Flat Duo Jets? They wrote music highly inspired by musicians from the 50s and 60s. Their early songs and performances were quite energetic and they mellowed out a bit later on.

Reply ·

182

Figure 7.1, Comments section design (2012/2013)

This particular order assigns high value to the contributions of the uploader, as they are immediately visible on the screen, whereas with descending order of assigned 'importance' and increasing age, more work needs to be done to retrieve a comment. Assuming that visibility of a comment is of value to a commenter, the site design seems to encourage contributions that attract the uploader's reactions and contributions that prompt others to assign it a positive rating. Whether the design actually has an impact as an incentive for comment design or not, however, cannot be determined in this study due to lack of evidence (and space).

One resource that is particular to YouTube is the so called "deep link", which allows commenters to create a hyperlink to a particular place in the video. Another viewer who clicks on it will be shown the video starting at/from exactly that place (cf. Figure 7.2).
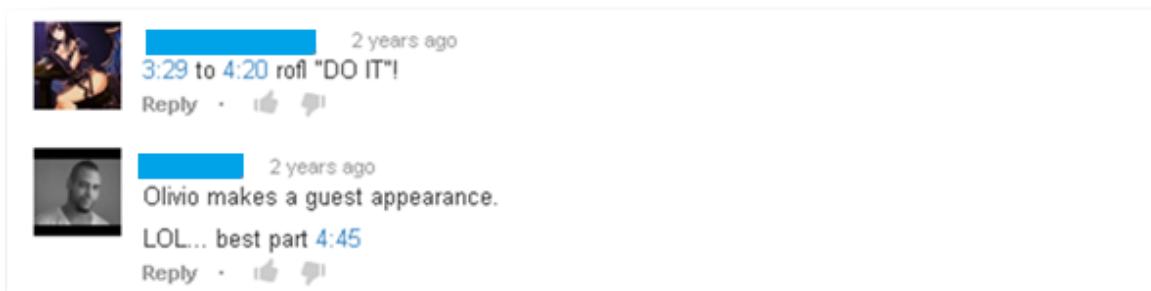
Figure 7.2, Deep link design (2012)

Thus, commenters have a non-verbal tool to establish reference to an element of the video which they specifically want to comment on. This is discussed in more detail in the analysis section of this chapter (7.3), together with other means of tying written language to recorded spoken language occurring in vlog comments.

7.3 Analyzing speech to comment coherence

For the analysis of the data I deliberately chose a qualitative approach for several reasons: the corpus of comments belonging to the vlog corpus that I collected, which would allow for a quantitative analysis, is a record of interaction from 2010 when I systematically assembled the data. In the meantime, YouTube's design has changed – so the actual context for which the contributions were created is no longer available. I suspect, however, that the surroundings in which a contribution is situated have a major impact on users' design of these contributions. Thus, for the purposes of the present chapter, that is, to identify coherence and sense-making strategies employed by users of a certain type of CMC, I deem the CA premise that instances of language use should only be analyzed in the context they occurred in particularly applicable.

Before I move to the sequential analysis of selected examples, I explain what type of interactional work a seemingly simple act such as commenting on YouTube involves.

The existence of interaction through commenting on vlogs is a reflection of a certain unspoken agreement between those who self-select to participate: the publically accessible space allocated by the site to written comments is treated as a forum where it is considered appropriate behavior to voice one's opinion about any aspect of the video that is playing above it, and thereby to mark them as reactions to this video. The vlogger, as the initiator and creator of this space, signals that he or she is receptive to such reactions (this is especially true as there is the option to disable comments).

The act of commenting then is an acknowledgement of the existence of a video and of the fact that any contribution within the commenting space below will be understood as relating to the video and/or another contribution within that space. This convention is not free of violations. There is even a function that lets users submit a decision as to whether they find a contribution appropriate for the space it is in: when hovering over a comment with the mouse cursor, the option to "Flag for spam" appears (cf. Figure 7.3 below).
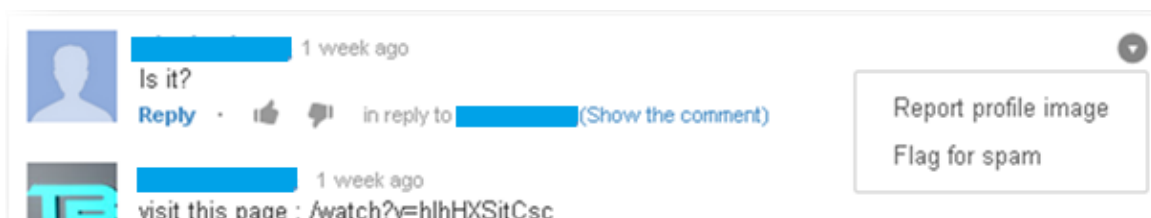

Figure 7.3

Once "enough" (YouTube n.d.) users have marked a comment as spam, it is hidden and can be retrieved by clicking on a link (cf. Figure 7.4 below).

Figure 7.4

The site designers have, therefore, implemented a tool to help negotiate what is considered appropriate behavior, and what is not. Needless to say, this tool can be used to flag any comment, so contributions that some might consider perfectly appropriate might be hidden, as well. In practice, many of the thus hidden comments seem to contain links to other videos or advertisement (cf. Figure 7.5 below).
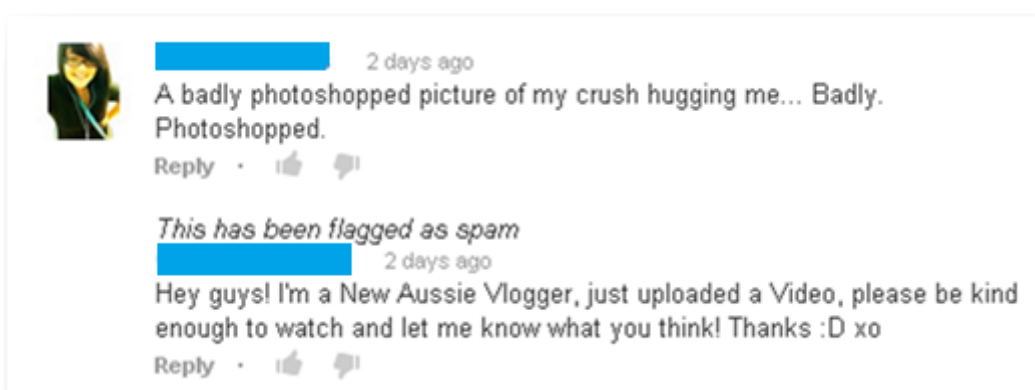

Figure 7.5

Initially, I analyze a very basic technique frequently found in vlog comments: repetition of words or phrases from the vlog. The example consists of an excerpt from a transcript of a vlog and a screenshot of three comments left underneath this vlog by viewers. The speaker (Zipster08) is telling a story about an incident involving his dog and a dog trainer at a pet store.

Example 7.1

BUTCH The KNOWLEDGE Bandit!

Figure 7.6

1   and everybody,

2   one by one,

3   all of the people who were there to have their dog trained,

4   they started laughing,

5   she's like you are disrupting my class,

6   I'm like "so so:rry:"

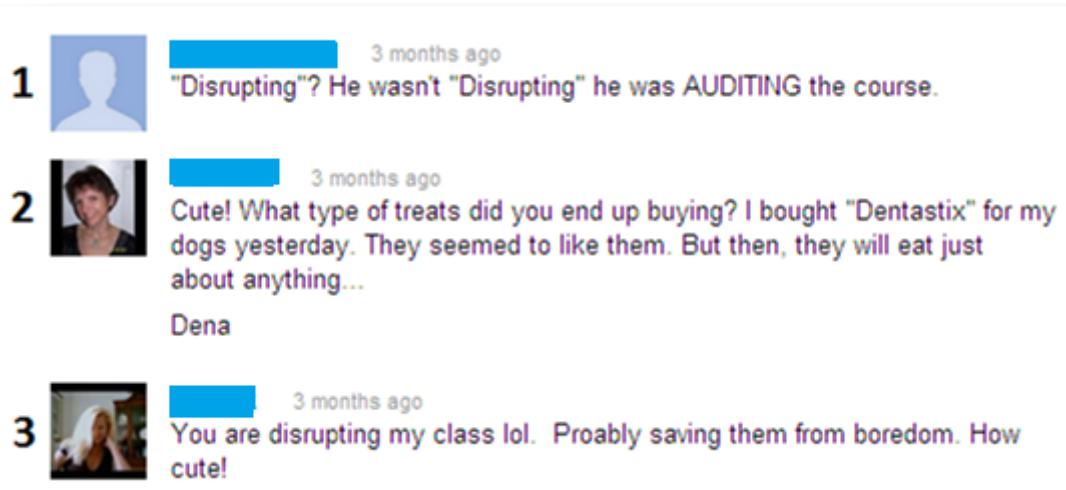7   anyway,

8   so yeah,

9   you were cute today,

10 Butch,

Figure 7.7

Zipster's video contribution created a space for interaction, with his audio-visual recording being the initial act. In order to access this space, a viewer has to click on the video (they can then pause it or mute it, of course, but it is assumed that most viewers will play it until it is finished or they leave the space). The three comments shown here were specifically placed in a context where they can be understood as coherent actions, making sense, maintaining social order (note, for example, that they could not easily appear underneath any other video on YouTube, or any other space on- or offline without creating some confusion as to what they might be referring to). And by posting these comments, the viewers signal that they understand the vlog as part of an interaction that they, too, can choose to be part of.

The tying technique that all three comments make use of is that of repetition/quoting (see Chapter 5 of this study on repetition as involvement technique; Tannen 2007 (1989); Herring 1999). Comments 1 and 3 repeat the direct speech Zipster quotes from the dog trainer, which apparently functions as part of the complicating action, or the funny bit, of his narrative. By repeating Zipster's words, the commenters signal understanding of these words, and more specifically, they signal that they consider them in some way essential to the story. The significance of the word "disrupting" seems to lie in its humorous nature in this context. Comment 1 reflects this by extending the inappropriate use of the phrase "disrupting my class" when addressing a

dog to the equally inappropriate term "AUDITING". Comment 3 expresses the humorous uptake of the phrase even more directly through the well-established acronym "lol" (laughing out loud). Comment 2 repeats the assessment "cute", which expresses agreement with Zipster's evaluation. The commenter then ties back to the orientation section of the narrative through a question relating to why Zipster went to a pet store in the first place.

What needs to be emphasized is that these three comments (which appeared in this sequence, as depicted in the screenshot, within the larger comments section of 70 comments) manage to form a tie between two modes, the spoken and the written. This is done through means that are well established as tying techniques in synchronous conversation, namely repetition. However, the situation here is different in that both parts of the interaction, the spoken and the written, are asynchronous: the spoken is recorded, and is thus perceived with a delay, and the written likewise leaves a permanent (digital) record, which can be accessed later. Thus, the problem that commenters have to overcome is a complex task. Not only do they have to make their comment relevant despite a temporal delay, both on the recording and after the video has been uploaded, they also have to bridge a spatial gap. In other words, as the spoken language is a continuous monologue, a commenter cannot interrupt or take a turn to tie their own contribution to a relevant utterance when it occurs; by the time a commenter sees the video, Zipster has already continued talking. In fact, by that time he has already stopped talking and completed uploading the video, so both time on the video as well as regular, everyday life-time has passed – and thus a commenter must (necessarily) 'skip-tie' (Sacks 1995), that is, tie two non-adjacent units, to a particular utterance, as adjacency is out of the question. At the same time, commenting takes place in a visual context that translates time into a particular spatial grammar: in the case of YouTube, the latest comment appears at the top, from where it will be replaced ('pushed down', if you will) by the next one. A commenter has to keep in mind that spatial immediacy to the video is not (permanently) possible, and one cannot predict what will appear above one's own comment. Consequently, the author of a comment X has to provide for the possibility that another comment Y will

appear in the space above, and that there is a comment Z below it, both of which comment X does not relate to. If the author of X wants to comment accountably on the video, this intention has to be signaled in some way to avoid confusion stemming from the availability of two potential contexts a comment can tie to: another comment, thus staying in the written mode[13], or the video, thus switching to audio-visual mode.

7.4 The deep link and its implications for mutual understanding

Two comments in the following example make use of the design feature that allows users to tie comments specifically and accountably to a place in the video, the so called deep link. Again, the simple use of this deep link function signals that a comment does not refer to any other comment in the spatial surroundings, but to a place in the video, which, if the reader of the comment is interested, can be accessed by clicking on the link. This presupposes a lot of extra work on the reader's part: a reader must be interested and willing to click on the link, which can have a disruptive effect on two levels. For one, the current activity of reading comments is interrupted by directing the reader's/viewer's attention (back) to the video; and for another, if the video itself is still playing, it will be interrupted and moved forward or backward to exactly the specified place, resulting in a discontinued audio-visual track. Unfortunately, to my knowledge, there are no statistics or further insights on the actual click-rates of these deep links available, nor have I been able to find reactions to comments with deep links that provide clues as to how they are received. Thus, I cannot comment on consumer behavior in this case. But it is my assumption that precisely because of the disruptive nature of these links, not many viewers/readers click on them. The following analysis presupposes the comment authors' expectation that their link will actually be clicked on.

---

[13] as mentioned before, there is a design feature that allows visibly staying within the written mode – the reply function

Example 2

Figure 7.8, Screenshot of vlog at 2:16

Transcript of vlog at 5:18

1   track number nine is xyz,

2   which are actually the last three letters in the English alphabet,

3   if you did not know,

Figure 9, comments to I GOT A BOY ALBUM & MV REVIEW – Soshi Vlog #43

Comment 1 (cf. screenshot) ties to a place in the video where the vlogger lip-synchs a song. At 2:16, the song (which seems to be interpreted by a woman) is at a particularly

high pitch. Comment 2 ties to a place in the video where the vlogger is talking about music.

One assumption one might make about the deep-link function is that commenters could use it to approach a conversation-like state, in that they could tie to, say, a particular spoken question, e.g. "what's your favorite color", and then provide the answer in a comment, e.g. "blue", thereby creating adjacency. This rarely seems to be the case, though, and I assume that this is so because while a deep link provides the starting point for a relevant referent in the video, it does not automatically stop. This makes it more complex to precisely point out what the comment ties to, in that it could potentially be any stretch from the deep link to the end of the video[14].

Both comments 1 and 2 do not approach this hypothetical conversational scenario. Both report on an immediate reaction that supposedly took place in temporal immediacy to the watching of the relevant scene. Comment 1 reports on a feeling of anxiety in reaction to the lip-synching (cf. screenshot of video above), and comment 2 reports on laughter in response to an apparently humorous remark (cf. transcript above). However, the comments do not themselves represent the respective reactions directly, and thus adjacency is not created. What is presented in the two comments is a report of a situation where two temporally adjacent contributions occurred: "u scared the shit outta me" and "I laughed at 5:18 waayyy harder than I should have". Both of these comments are marked as a reaction to the previous contribution that can be retrieved independently of the reported incident by clicking on the deep link.

As previously mentioned, a reader/viewer who wants to make sense of these comments has to do considerable work. A time stamp as such is simply a string of numbers which, in itself, cannot create coherence. It only becomes relevant as a context creating device when it is clicked on. However, it simultaneously becomes a context changing device in that it directs attention away from the written comments to

---

[14] I have, in fact, seen a commenter use two deep links in one comment, similar to "best part: 2.15 to 3.03", cf. Figure 7.2. While this pointed out to a reader both beginning and end, it was not useful in terms of the function of the second deep link, in that this one, too, would be treated as a starting point if one clicked on it. A speculative explanation of this 'misuse' could be that the commenter did not actually want to use a link, but his or her simple text was automatically converted.

the recorded video. A viewer has to be prepared to concur with this change of context and the resulting change of mode. Subsequently, they have to actively search for the element of the video that is actually meant, for it might be any number of things, either verbal/audio, visual, or a combination; and it could be only a fraction of a second long, or much longer. Thus, the placement of such a deep link in a comment reflects a certain confidence of the commenter that whatever it is that is referred to in the video will be identifiable by the reader/viewer with the help of this link, and it reflects a presupposition by the commenter that a reader/viewer who clicks on the link will assume that this link will help them find the relevant referent. In other words, there is a mutual agreement between commenter and reader/viewer that, by inserting and clicking on the deep link, both parties collaborate to create coherent interaction and therefore make sense of each other's actions.

7.4 Comments and reactions to them

So far I have shown that spoken vlogs, or parts of them, can be considered the initial act of an interaction which is continued in the comments section, mediated through written words and links. The evidence presented consists of these written replies: comments that demonstrably tie to the initial act through devices such as quoting or the deep link. Thus, I have demonstrated that the interactants themselves – the members – orient to vlogs as initiations of such communication. What remains to be shown, then, is that these comments left in the space below the video are, likewise, considered replies to that specific video by the members. To that end, the available reactions to such comments need to be assessed regarding their status as evidence for the assumption that there is a coherent, interactive exchange being conducted. Consequently, the following example will be used to discuss demonstrable coherence structures in comments and their replies. The first screenshot below (Figure 7.10) shows a point in the video that is part of a photo slide show, with music playing and no talk. This is the time linked to the deep link in comment 3 in the second screenshot (Figure 7.11).

Example 7.3

Figure 7.10, TTMT #170 - Anime Overload 2012

Figure 7.11, Comments to TTMT #170 - Anime Overload 2012

Comments 1 and 3 were left as reactions to the video, more precisely to pictures shown in the video of people in costumes ("Eridan" is the name of an anime character). Each comment received a response by the uploader of the video. As previously mentioned, the comments section has a reply function, which automatically alters the appearance of a comment left in reply to another, in that it is placed directly underneath the one it replies to, and it is indented, with a smaller thumbnail image of the commenter's profile picture. Most importantly, though, there is an automatic notice "in reply to *username*". These design inherent resources, which are available to any user simply by clicking on the "reply" link, make the status of a comment as a reply to another structurally unambiguous. Similar to the function of the deep link discussed in the previous section, this reply function can be used to visually signal a tie between two contributions, but the signal itself does not create coherence. It should rather be seen as an invitation by the user posting the reply to search for the relevance of the reply to the comment replied to, a signpost helping another reader to recognize what structures are being created throughout the interaction. In other words, the reply function helps users to react to a comment accountably. This is not to say that one can only reply to a comment using this function – it is feasible that users leave replies in the form of regular comments. However, that means that they will have to find other means to demonstrate that they are posting a reply to a comment, rather than a comment to the video. Research on Facebook comments, where this reply function does not currently exist, has shown that one way to signal reply status in the absence of spatial immediacy is using names of other users as terms of address (Frobenius and Harper forthcoming).

Regarding example 3, it is noteworthy that, both times, it is the video uploader who leaves the reply, rather than another viewer. Nowhere in the comments is she directly addressed, and there are no further signs of her being the sole recipient of the comments. While it is possible (and is often the case) that a three-part interaction *vlog-comment-reply* takes place between three users, in this particular case (Comments section design 2012/2013), the vlogger as creator of the space and initiator of the interaction self-selects to contribute the third part. Both uploader

replies also seem to be friendly acknowledgements of the comments, which, rather than express necessary[15] additional information, function as a signal of uptake, understanding and gratitude, perhaps more in the spirit of phatic communion. This is in line with anthropological research on vlogging (e.g. Lange 2007b), which provides the insight that people use vlogs to share their life experience and to participate in the community. Thus, an answer that is strictly speaking not necessary because it would not be considered officially absent might then be a show of availability and interest in the communication.

Returning to coherence and the issue of evidence that these examples do, in fact, represent a three-part interaction rather than three separate units, we can say that the first of the two exchanges (comments 1 and 2) are more strongly coherent than the second one (3 and 4). The mention of the character "Eridan" in the comment and the noun phrase "his costume" in the reply form the verbal tie between the two contributions. The reply thus demonstrates (1) the understanding that the comment was left with respect to a certain element of the video in that it is placed in this specific comments section and picks up a particular topic from the video; (2) the comment is a contribution that can, itself, be replied to; (3) the understanding of the comment itself.

In the second exchange, only the demonstrative pronoun "that" in comment 4 ties to the whole of comment 3. "That" is relatively unspecific in its tying properties, so comment 4 could be a reply to any comment. This weak tying character makes the reply particularly interesting because it can be used to show that a user can confidently construct a comment as a reply, reducing the interactional work of tying to a minimum, as the spatial design of the reply function will compensate for missing clues that help a reader recognize relevant structures.

---

15 By "necessary" I mean, for example, the status of a second pair part of an adjacency pair, such as question/answer or summons/answer, where official absence of the second pair part allows for certain inferences. The notion of conditional relevance, which expresses exactly this necessity of the second pair part (cf. Schegloff 1968), seems hardly applicable to the present context of written comments – however, there is, to date, no systematic research on this. Thus, it is questionable whether in this context official absence can even occur.

7.5 Coherence across modes

I have presented the kinds of commenting practices that are facilitated by the affordances of the comments sections on YouTube. However, these are not the only reactions one can find on the site. In fact, YouTubers employ various means to encourage and respond to comments. They range from direct requests to leave comments (spoken/gestural ones in the vlog, see also Chapter 6, or written ones in the video description) to reflections on comments to previous vlogs.

The following sequence of screenshots and transcripts represents a fairly complex chain of sense-making in vlogs, in that we find at least two instances of mode switches within the same interaction. The example starts with an excerpt from a vlog, where a speaker asks a specific, vocabulary related question that she would like to see answered in the comments to the video. The screenshots below show some of the several thousand comments that were left underneath the video. Next, there are screenshots of a second vlog by the same vlogger, in which she shows some of these comments while simultaneously responding to them. This is represented in a table containing images and transcript of talk (screens 1-5). So we have:

a) a question (vlog talk)

b) answers (written comments)

c) responses to the written comments (voiceover in following vlog)

Example 7.4

a) Transcript of excerpt from "STARTS WITH A B…"

1 but anyway,

2 sorry,

3 can I also ask,

4 if you can think of any wo:rds.. that are similar to,

5 in the sphere of,

6 doesn't have to be a synonym,

7   for the word crass?

8   that start with a b.

9   because I ran out while I was filming.

10   CUT {to herself, acting as one of her characters}

11   brazen?

12   braggingly?

13   CUT {to herself as vlogger-character}

14   and looking back,

15   I had complete control,

16   why didn't I just change the script,

This transcript contains an act, namely a question, that makes it obvious that also from the vlogger's perspective, a vlog can be considered the initiating element of an interaction, with the space below serving as a platform where reactions can be expressed in a certain format. The question that extends from line 3 to line 8 asking for synonyms of "crass" that start with a "b" clearly indicates that this interaction is not meant to be unilateral, but that there is a desire for feedback. More specifically, this type of question demonstrates awareness of the kind of format this feedback can be conveyed in, namely (written) words in the comments section below. The vlogger presupposes tacit understanding of this, as she does not explicitly point to the comments section to viewers who want to answer her question.

b) comments:

Figure 7.12, Comments from "STARTS WITH A B...", uploaded August 21, 2012

As can be seen in the screenshot above, several commenters do answer in relation to the question raised in the vlog by providing synonyms (6, 7), other words starting with "b" (2)[16], or by stating that they cannot answer the question (1). These four examples clearly demonstrate understanding of the question as the opening of an interaction that can be continued in the comments section. Other comments do not relate to any

16 "bvideo bupload" is a reference to the infrequency with which the vlogger posts new videos, despite her regular promise that the next video will be uploaded in a few days, which is a recurring topic in comments underneath her videos

specific part of the video (3 and 9), and yet others tie to other parts of the video (5 and 8), and one comment evokes communication that took place on a different site (4). Regarding the larger context of this interaction, it needs to be added that the vlogger is fairly successful in terms of view count and the number of comments her vlogs regularly attract. Furthermore, she regularly asks questions or raises issues as suggestions for comment topics. But most importantly, she has a history of including selected comments in their written form in her next video, sometimes with spoken remarks that often ridicule herself or the commenter, or that allow them to appear particularly witty. I assume that this practice is an incentive for many viewers to comment. This integration of comments into the next video is the next step of my analysis. I have divided the transcript into five sections according to the still image that is shown while the vlogger is talking. For the sake of readability, the analysis follows immediately underneath each section.

c) Transcript of voiceover from "LOOK ME IN THE EYES", uploaded August 25, 2012

screen 1
bobocomments Nat there is a word that starts with a B and its Cuddle... as in that's what we need to do Ditre bvideo bupload gameruler789 Whenever someone does the "it starts with a B routine" I say things really loud to block out my opponent.

1	okay,=
2	=it's porno music slash comment time,
3	and I have no idea why,
4	but I didn't think of the word brash even though it rhymes,=

5	=or boorish,=
6	= I'm glad

Screen 1 shows parts of comments 2 and comment 5 from the screenshot of selected comments above, as well as another one. The appearance is slightly altered though still recognizable: Each comment is headed by a username, which appears in blue and underlined, suggesting a hyperlink to a profile; and the comments themselves appear in black print. There is no thumbnail image of the commenter's profile picture, and no time stamp. The middle comment by a user named Ditre has clearly been edited in that it has been reduced to "bvideo bupload" from a string of six terms in total starting with "b" (as shown in the screenshot above in section b). The accompanying speech by the vlogger introduces this part of the vlog in lines 1 and 2 as "porno music slash comment time", which is her standardized phrase in each of her vlogs to mark the beginning of the final part of the vlog in which selected comments from the previous video are shown. The act of copying these contributions and their recontextualization in the video represent a similar act to that discussed in example 1, where commenters quoted speech from the vlog in their comments. It emphasizes the relevance of certain contributions to the present context. Interestingly, the mode shift is not effected completely in this case: The quoted comments retain their written form, and they are usually not read out. So while these comments are subject to the ephemeral nature of the video (and thus acquire one characteristic of spoken language), they do remain written language that needs to be seen and read, not heard. One can only speculate as to why the vlogger does not integrate relevant comments fully into the spoken mode of the vlog: This visual integration preserves the character of the comments and makes them recognizable as such. This might strengthen the vlogger's appeal to commenters, as their comments will eventually feed back into her vlogging.
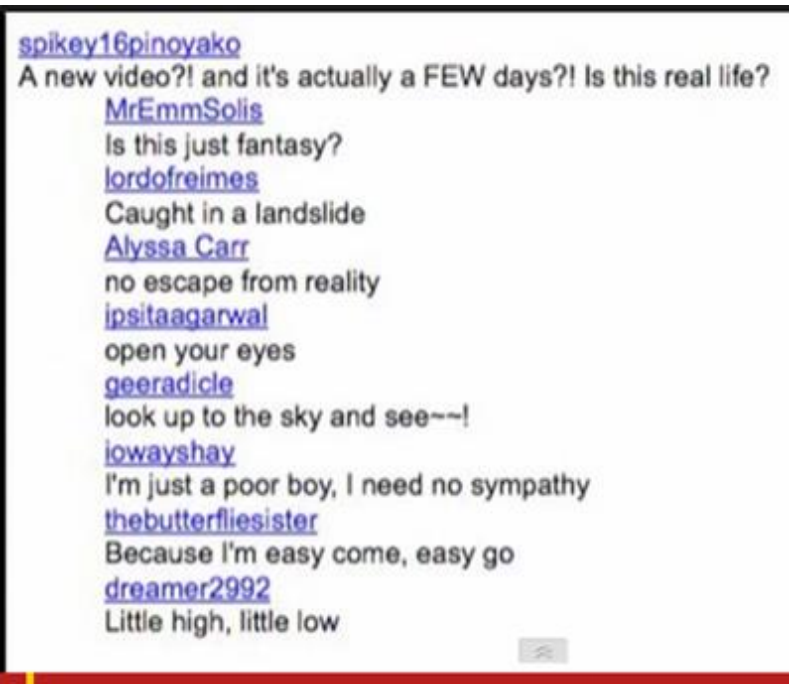
Note that the vlogger does not specifically refer to any of the three comments shown in screen 1, as she is talking about particular vocabulary items that are not the subject of the comments. So the tying that the vlogger achieves is solely based on the visual quote of the comments. She thus signals that she understands that these comments

were left as a reaction to her previous vlog, and that as such they are relevant to such a degree that she extends that relevancy to her own visual contribution through such an endorsement. This visual quoting thus turns out to be a remarkable practice, in that it makes use of the option to separate the audio stream and the visual stream in terms of the messages that are being communicated. The visual stream is therefore used to mediate a reaction that does not add any new content explicitly, while the audio stream adds completely new information that cannot immediately be tied to the visual stream, but must be seen as coherent with the general context of the comments section to the previous video.

	screen 2
	Sarah Sachs
	Lol, I've been overly eager to give someone a pen before - then I spend the next few hours watching them with MY pen until they return it.
	StickmanQueen
	Singlet: A sleeveless garment worn under or instead of a shirt.
	calebmacc
	You can tell you're lonely cause the other versions of you are waayyy to excited to be your friend.
7	to know you all have good vocab,=
8	it's beautiful.
9	and Sarah that's hilarious,
10	I hate lending people my pens as well,
11	I always panic about not getting them back,=
12	=and I get all

As opposed to the comments in screen 1, the first one in the following screen is specifically tied to in the vlogger's speech. The comment itself refers to an enacted scene from the vlog which showed the vlogger as one of her characters lending another one of her characters a pen. The strongest tying device used here is part of the

commenter's user name "Sarah" (l. 9), which serves as term of address and as such disambiguates which of the three comments the vlogger is referring to. The subsequent account of her own feelings regarding lending pens to others (l. 10-14), which almost reads as a follow-up story that has been reshaped from a concrete narrative event to an abstract, generalized account, creates topical cohesion. This results in the coherent organization of visual and audio stream, such that the audio content can be recognized as a reaction to the visual content in which the vlogger retrieves the original contribution by another participant.

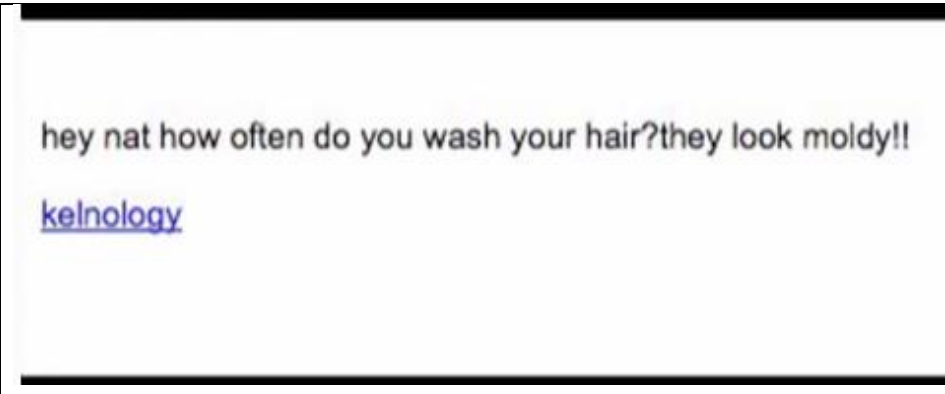|	
	screen 3
13	"are you done with that" at the end of it,
14	it's very awkward.
15	and yes it's true,
16	I have the best subscribers in the world,
17	so thrilled you

The string of comments shown in screen 3 is of particular interest since they are not separate comments left independently, but they form a coherent unit. The initial comment masquerades as a reaction to the exceptionally short time between the

uploading of the previous video and the one commented on (cf. the remark "bvideo bupload" above, which revolves around that same topic). But predominantly, it acts as the initiation of a joint line by line quoting of the rock band Queen's *Bohemian Rhapsody*, which eight subsequent commenters continue. This joint endeavor is no small achievement, as it presupposes that among those who happen to be attending the comments section at a point in time, there are several who not only know the lyrics of the song well enough to recognize and quote it. These people must also understand that there is the particular action of joint quoting happening, which is organized through the affordances of the comments section.

The vlogger demonstrates recognition of these efforts: she understands that these are not random comments that happen to appear as adjacent contributions, but the incorporation of them as a co-ordinated unit reflects her uptake of the inter-commenter coherence. She makes this more explicit through her voice-over that expresses her appreciation of the commenters' joint effort: "so thrilled you guys could band together on this." (lines 17/18). Similar to the tying in the previous screen, the vlogger co-ordinates audio and visual stream to produce a particular kind of adjacency which helps viewers recognize that what she is saying is a reaction to these comments in particular. This can neither be termed temporal adjacency, because the presentation of the written comments is simultaneous with her speech, nor can it be called spatial immediacy, for her speech is not mediated through space.[17] Perhaps, for lack of better terminology, *multimodal adjacency* describes this phenomenon most appropriately.

screen 4	

---

17 This simultaneity causes problems for viewers: commenters occasionally state that they have to pause the video to read the comments and then continue listening to it, thus resulting in temporal adjacency.

	hey nat how often do you wash your hair?they look moldy!! kelnology
18	guys could band together on this.
19	and my hairs are moldy?
20	that's just the

	screen 5 YOUR MUM RATES ME
21	worst news I've heard all week,
22	I wash them I promise.
23	all my hairs.

Screens 4 and 5 need to be discussed together because the voice-over in the audio track that starts in screen 4 encompasses the visual track of both screens. Screen 4 contains a similar tying procedure as screen 2: A comment is shown and there is a verbal reaction to it in the voice-over, which ties to it. In this case the tying is done

through partial repetition and reformulation of the comment so that the quote is mediated both visually as well as through talk. During the spoken reaction to this, the visual track switches to show the final screen, which has the four words "YOUR MUM RATES ME" sequentially appear in big bold letters. This is the standard screen this vlogger uses to end her vlogs, and it appears to be an allusion to a motif in contemporary youth (online) culture which involves sexual relations with other people's mothers. As such it cannot be shown to tie to anything specifically,[18] but broadly speaking it refers back to all previous uses of this screen and the frequent occurrence of jokes and allusions pertaining to the "your mum" motif on the internet or in this vlogger's vlogs.

7.6 Conclusion

Conducting coherent exchanges is an organized form of joint human interaction. This is true for synchronous conversation as well as for communication that is asynchronous, mediated and enacted through several modes, such as the interaction that takes place in vlogs and their comments sections. One device that facilitates the structuring of such exchanges is tying, which is an essential way to add coherence to an interaction. Through such practices, interactants give and receive clues regarding what appropriate behavior, and specifically verbal contributions, consist of. The present chapter represents an initial investigation into the tying practices used in the comments sections of vlogs.

Tying a comment to a vlog is a specific problem as it involves a mode switch: Vlogs are predominantly mediated through spoken language and gestures, whereas comments are, by design, mediated through written language. Thus, any comment must be designed in such a way that a fellow viewer/reader recognizes precisely what it refers to. This entails signaling that the spatial grammar imposed by our reading practices,

---

18 One might speculate about the sexual innuendo that is present in the "your mum" reference and the potentially ambiguous comment about hair, but there does not seem to be any evidence to substantiate that.

that is, reading from top to bottom, is suspended in favor of *reading from video to comment* and in reverse chronological order. It also entails providing for the fact that the referent will have to be detected in the audio-visual track and that a reader seeking to make sense of a comment must bridge the discontinuity of modes.

The YouTube comments section provides two affordances that help signal whether a comment is a reaction to another comment, or to a particular place in the video: the reply function and the deep link. The reply function is less problematic than the deep link in two respects: it reinstates the established top to bottom reading order; and there is no tying between speech and writing involved, so that the tying occurs within the same mode. The deep link, on the other hand, ties a written comment to a certain position in the audio-visual context of the video, making a mode switch inevitable. Furthermore, it is imprecise as a tool in various respects: it only marks the beginning point of a scene (however long it may be) that is tied to; and it leaves open whether the comment ties to the audio track, the visual track, or both. As a result, users of these functions must do additional tying work to provide guidance as to how to make sense of their comments.

The data presented in this chapter has demonstrated a range of practices that YouTube commenters and vloggers successfully employ, and it has revealed that interactions can extend considerably, not only in terms of the number of contributions, but also regarding modes. To achieve this, the concept of adjacency needs to be adapted from its original application to two consecutive conversational turns to a more complex model for the description of 'spatial adjacency' in the case of written language, and 'multimodal adjacency' in the case of written language as a reaction to spoken language and vice versa. The present study has only touched upon this in passing, but more thorough research on this, based on more varied data, could further advance our understanding of how people use media sites combining a multitude of interactional modes.

8 Vlogs and TED Talks – a comparison

8.1 Introduction

This chapter goes beyond the discussion of vlog data: the comparative investigation of another monologic genre serves as a backdrop and control setting, so to speak, to determine what is specific to vlogs in the findings my analysis has produced, and what is common to at least one other genre. This comparison is a first step toward developing an agenda for the study of monologue, which receives considerable attention in the conclusion.

The data I use for comparison of vlogs against non-vlogs comes from *TED*[19] *Talks*. TED Talks are monologues that are given in the form of lectures at conferences (TED Conference, cf. http://www.ted.com/). That is, there is an audience physically present, whose audible, and sometimes visible, reactions are captured in the video footage of these talks. There are some considerable differences between vlogs and TED Talks, which I describe in this section. However, as a genre, these talks are also sufficiently similar to vlogs to draw comparisons that are more meaningful and to the point than fairly general observations one might make when comparing vlogs to, say, dinner table conversation.

The most significant difference between vlogs and TED Talks is that, from an editorial viewpoint, the talking and filming/editing are separate steps not only in a temporal linear sense, but most likely they are carried out by different people. Thus, we have a first step in the production, the actual delivery of the lecture in a lecture theater, and there is the filming of it, usually with several cameras from different angles, and the subsequent editing. In the editing process, the footage from up to six cameras as well as a screen capture of the presentation slides are combined to result in one seamless audio stream (i.e. nothing is cut out of the talk) and one stream of visual documentation that several sources feed into.

---

[19] TED stands for *Technology, Engineering, and Design*

Even more drastically different from vlogs is the reception situation of TED Talks. The speakers are on a stage and talk into a microphone **to a present audience**. This means that TED Talks are received at the same time as they are produced, and audience feedback is possible (e.g. laughter and clapping), if not ouright encouraged. The conference has been held every year in at least two places since 1984, however, only since 2006 have talks been uploaded to the internet (the oldest talk posted is, indeed, from 1984). Thus, my assumption is that speakers predominantly talk to the audience which is physically present with them in the room while not consciously attending to a later online audience, as opposed to vloggers, who do just the opposite. For more recent talks, where speakers knew their lecture would be put online, this might be different.

The speaking situation of TED Talks, in addition to the stage and the audience, bears a number of traits that distinguish it from vlogs. The speakers are invited by the TED organization; they do not self-select the way vloggers do. As there is some sort of selection process, I assume that a person's presenting skills are essential to go through this process successfully. In fact, the introductory paragraph to TED's FAQ sections says:

> TED isn't a typical conference. The TED audience has high expectations of the speakers; the TED speaker team works with speakers well in advance of the conference to help shape a presentation that will succeed on the TED stage. TED is the place to give the talk of your life (TED Conferences » About » Speaking at TED. *http://www.ted.com/pages/speakingatted* (06.05.2013) [no author, no publishing year]).

The original production of these talks seems to be the result of meticulous preparation according to this passage, and also according to a 'behind the scenes' video (cf. Go behind the scenes of a TED talk. *http://blog.ted.com/2009/07/15/go_behind_the_s/*. (06.05.2013) (McManus 2009). The actual delivery of these talks not only often attests to impressive factual knowledge, but also to considerable stage presence and online language production capacity. This can be seen when following the transcripts of the

talks which are available on the website alongside the videos. Most of these show well-composed texts, delivered in complete sentences rather than syntactically fragmentary units typically found in spoken language. Indeed, one might wonder whether some of these talks were memorized by the speaker, and the transcript was more or less a copy of their script. TED Talks are restricted to a maximum of 18 minutes in length, and usually focus on one topic or idea (the slogan of the whole TED organization is "Ideas worth spreading"). By comparison, vlogs have a ten minute limit, and while some speakers focus on one topic, others move freely from one to the next. In general, TED Talks seem much less spontaneous and much more professionally planned than vlogs.

The editing, as mentioned above, is presumably done by someone other than the speaker. It has a significant impact on the genre TED Talk, if one can call it such, just as it does on vlogs. However, at this point, the status of TED Talks needs to be clearly defined as a genre fundamentally altered by its recontextualization in an online environment. In their version as video clips, they are removed from their immediately preceding and immediately following conference context. In none of the videos studied is there evidence of an introduction of the speaker by a host, nor is there evidence of a discussion following the talk. Instead, there are opening credits (including music and the TED logo and slogan) as well as closing credits. This means that the transition from non-lecture talk to lecture talk, which must have taken place at the conference, and the transition back after the talk is finished were deliberately edited out.

Investigating the openings and closings of these talks and comparing them to openings and closings of vlogs yields particularly interesting insights, as these passages seem much more uniform across the TED talk corpus than across the vlog corpus, hinting at stronger genre influences at play in the conference situation than in the vlog situation. As with YouTube, the TED website allows viewers to leave written comments on the talk. Additionally, and unlike YouTube, TED provides a script and subtitles (created by volunteers) in the video to many of its talks in a wide range of languages. Thus, several measures are taken by TED media professionals to make these videos not just videos

on a website, but distinct tools to spread the content of TED Talks to as many people as possible around the globe. Therefore, a thought worth elaborating is whether TED Talks in their online form actually constitute a separate online genre, a sibling to its offline conference version. As previously mentioned, editing, supplementing of transcripts, translations, and the recontextualization in an online environment support this notion. On the other hand, however, it is quite noticeable that TED Talks are not generic to the internet – especially the comparison of multimodal features (gestures and commenting) reveal that the performance of these talks is not adapted to this online environment. No speaker I have seen so far, for example, performs pointing gestures that make sense only in an online environment, as vloggers do.

8.2 The TED Talk corpus

This part of my data is made up of twenty TED Talks, collected on March 13, 2013 from www.ted.com. The interface of the TED website allows users to sort the talks according to various criteria, including e.g. length, event, and available subtitle languages (see Figure 8.1).
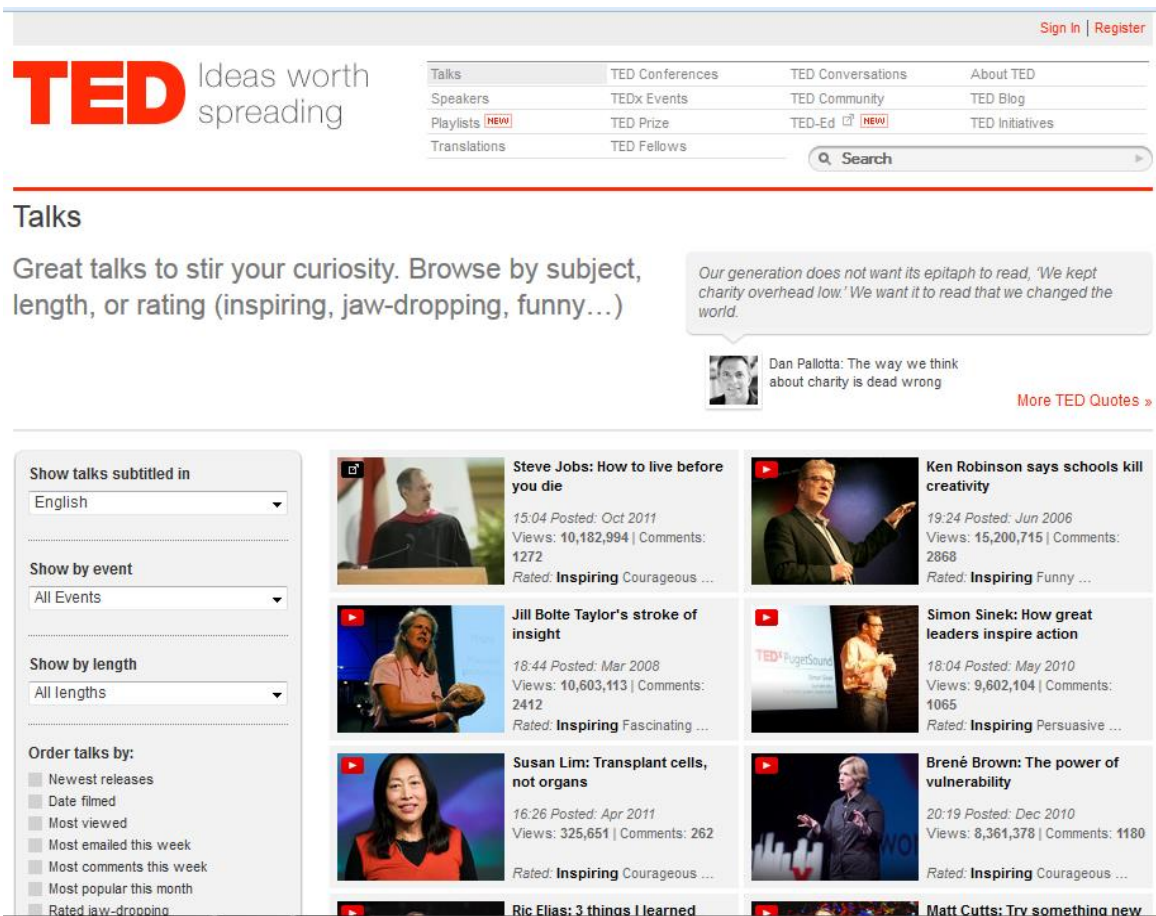
Figure 8.1, TED website interface

For the compilation of the corpus used in this study, I chose the settings 'all lengths', 'all events', 'all topics', 'subtitles English' (this was actually not a criterion I am interested in, but as there is no 'no subtitles' setting, I chose 'subtitles English' assuming that this would be the least restrictive choice in terms of search results), and 'most inspiring'. The 'most inspiring' setting, of course, requires explanation: most other settings yielded a strong imbalance towards a majority of male speakers. In an attempt to match the slight imbalance in my vlog corpus of approximately 40% women to 60% men in order to try to minimize any gender related differences (which are not a specific research objective in this study), I chose the setting 'most inspiring', because in the first twenty videos, this search offered eight by women and 12 by men. Two videos were not used, one by Steve Jobs and one by J. K. Rowling, as they were not original TED Talks filmed at a TED conference. And as they were only linked to the site, but hosted by YouTube, there were no transcripts available. For this very practical reason

and to maintain as much uniformity as possible, I chose the two videos that immediately followed the first twenty instead. The shortest of the videos (including opening and closing credits) thus collected was 3 minutes and 27 seconds long, the longest 21 minutes and 2 seconds. All talks were given in English, by – it seems – predominantly native speakers of various varieties, including British English, American English, Canadian English, and Indian English. Many of the speakers have an academic and/or entrepreneurial background, and even those speakers who do not seem to be native speakers have a good command of the English language. The earliest talk was performed in 2001, the newest ones in 2012. They are a collection from various TED events, with a varied audience size.

8.3 TED talk openings

All TED Talks are headed by credits showing a dynamic depiction of the TED logo, its slogan, and reference to the website where the talks are posted (cf. Figures 8.2, 8.3 and 8.4). These credits are approximately 15 seconds long, and they are the same for each talk (except for a minor variation in some, where a sponsor or the location of the conference is mentioned in writing).

Figures 8.2, 8.3, 8.4, opening credits

The blend-over from the credits to the first shot of the speaker on stage is accompanied by audience applause. Interestingly, this applause does not seem to be the actual applause that was given to the speaker upon walking on stage, for in each video there appears to be the same sound track, regardless of the sound quality of the rest of the video, which varies according to room size, audience size, quality of the recording equipment, etc. My assumption is further supported by the discrepancy of the sound of audience applause at the end of a TED talk and its sound at the beginning

of that same talk. In some cases, this difference is quite noticeable when the two are compared back to back. Consequently, when there is overlap of audience applause and the first words of the talk, this must be understood as a result of the editing efforts by TED media professionals, not as an on-site negotiation of floor holding rights between speaker and audience. This editorial freedom which constitutes an alteration of the actual events (albeit an admittedly small one) does move the online rendition of these talks towards instantiations of another, web-based genre. Thus, it is the auditory blend-over that represents a feature that replaces an interactionally negotiated transition at a conference in order to make it fit into its new online context. In other words, with regard to TED Talks, it is a genre specific feature only in the genre's online version.

Speakers at TED conferences employ some remarkably different strategies to those used in vlog openings. What is immediately striking is the lack of the greeting plus term of address combination that is present in so many vlogs. In all of the twenty talks in my corpus, only one begins in a remotely similar way to the vlogs studied:

Example 8.1
1   ((receding applause))
2   good morning,
3   how are you.
4   ((some audience laughter))
5   it's been great hasn't it?
6   it's been- I've been blown away by the whole thing.
7   in fact I'm leaving.
8   ((audience laughter))

This opening can almost be called conversational in that it uses a greeting "good morning," and a highly conventionalized formula "how are you." that can either have the pragmatic function of helping to open a conversation in a manner that signals politeness, or it can function as a request for knowledge. Either way, it requires an

answer from an interlocutor, which entails turn-taking which is a characteristic of western, non-confrontational, non-ritualized conversation. What makes this opening non-conversational is the setting which clearly fosters the expectation of a monologue – a speech situation where the audience is not expected to interrupt. While the audience in a vlog cannot interrupt, the TED conference audience can. What stops a physically present audience from participating in a more significant way other than laughter or clapping is the common understanding of their role as an audience, a *fait social* in Durkheim's (1894) sense, so to speak. Thus, when the speaker invokes a conversational script, he creates a script opposition to the monologue script the audience most likely has in mind. His main opening strategy thus appears to be the use of humor. In addition to the above example, the following strategies can be identified among the 20 TED Talk speakers:

3/20 say what they want to start with

7/20 say "I am/studied/trained" + profession

2/20 refer to an earlier TED Talk they gave

2/20 ask a question (rhetorical/non-rhetorical)

4/20 start immediately with an anecdote/story/observation that takes them to their
       topic

1/20 gives a fairly formal description of her situation and attitude as a speaker

Not one of them uses a form of self-identification that includes their name, as is common practice in vlogs. Nor do any of the speakers explicitly introduce their topic or the title of their talk, by way of a common formula such as "I'll be talking about…" or "the title of my talk is". This significant difference points to the elaborate planning that goes into a TED Talk as opposed to most vlogs.

A fairly common way to open a talk is a slightly altered self-identification: rather than giving their name (which may have been mentioned by a conference chair – this can be seen only once in a video that shows speakers preparing for their talks, cf. Wishnow et al. (08.05.2013) [no publishing year]), speakers explain their profession or occupation

when it is relevant to the topic of their talk. The following three openings are examples of this strategy.

Example 8.2

    1   ((receding applause))

    2   I'm a storyteller.

    3   and,…

    4   I would like to tell you a few personal stories about,..

    5   what I like to call the DANger of the single story.

Example 8.3

    1   ((receding applause))

    2   I grew up to study the brai:n,

    3   because I have a brother,

    4   who has been diagnosed with a brain disorder,

    5   schizophrenia.

Example 8.4

    1   ((receding [applause))]

    2   [back in New York,]

    3   I am the head of development for a non-profit called Robin Hood.

    4   when I'm not fighting poverty, I'm fighting fires,=

    5   =as the assistant captain of a volunteer fire company.

These three speakers all mention what they are trained to do or what they consider themselves to be professionally. The first (Example 8.2), by calling herself a storyteller, makes her profession as an author sound less like a business she makes a living with, but more like a fulfilling vocation that gives her satisfaction via the contact to those who hear her stories. This gives her the authority and expert status to further talk

about the topic of storytelling, which is what her talk revolves around. In Example 8.3, the speaker also uses her training as a neuroscientist to establish her expert status which allows her to continue lecturing about the brain, while at the same time giving a biographical reason for her career choice. The third speaker (Example 8.4), whose talk does not actually revolve around his professional area of expertise, still uses the work involved in his profession, "fighting poverty" (l. 4), to parallel the volunteer work he actually will talk about, namely "fighting fires" (l. 4). This fairly elegant parallel construction not only allows him to move smoothly to that topic while mentioning that he does this as a volunteer (a fact that is important for the unfolding of the story), but it hints at the professional approach he takes toward his work at the fire brigade. Again, this lends him authority and expert status.

Other speakers open their talks without any such introduction or display of expertise: they start telling a story *in medias res*, or they relate an anecdote. Consider these two examples:

Example 8.5

    1   ((receding[applause))]

    2   [imagine]… a big explosion.

    3   as you climb through 3,000 feet.

    4   imagine a plane..full of smoke.

    5   imagine an engine going claclaclaclaclaclacla,

Example 8.6

    1   ((receding applause))

    2   probably a lot of you know the story of the two salesmen who went down to .h Africa in the 1900s.

    3   they were sent down to find if there was any opportunity for selling shoe:s.

    4   and they wrote telegrams ba:ck..to Manchester.

    5   and one of them wrote situation hopeless.

    6   stop.

7 they don't wear shoes.

8 and the other one wrote,

9 glorious opportunity.

10 they don't have any shoes yet.

11 ((laughter))

Example 8.5 gives a description of various sensations the speaker experienced as a passenger on the plane that landed on the Hudson River. While the talk itself is about the various insights into life in general that the speaker gleaned from this incident, he uses a very graphic description of the situation that inspired existential human fears in him to convey a sense of the urgency of this situation to his audience. The other example (8.6) uses a much less dramatic introduction by telling an anecdote that illustrates that people often come to different conclusions about one and the same situation, depending on their attitude and viewpoint. While the story about shoe salesmen is not directly related to the topic of the talk, which is classical music, it illustrates these differences of opinion which the speaker also considers a valid observation in his area of expertise. And it does so while adding a humorous note. The following two examples illustrate speakers' direct acknowledgement of the task they have to master, namely starting a coherent monologue in front of an audience.

Example 8.7

1 ((receding applause))

2 so I wanna start by uhm offering you a free,

3 no-tech life hack,

4 uhm,

5 and all it requires of you is this.=

6 =that you change your posture...h for two minutes.

Example 8.8

1 ((receding applause))

2   what I thought I would do is I would start with a simple request.

3   uh I'd like all of you.. to to pause for a moment,

4   you wretched weaklings,

5   and take stock of your miserable existence.

6   ((audience laughter))

Both speakers use the word *start* to make explicit their understanding of the situation: it is their floor alone, and consequently they are in charge of managing it alone. Both speakers make the situation immediately relevant to the audience, however: the first by making an offer to them, the second by making a request. This strategy is of interest as an involvement strategy: as it is obvious to anyone that these first words in a talk mark the start of it, it is superfluous to mention this fact. Thus, these statements must have additional functions. I assume that there are two main reasons. For one, this reflects the cognitive process of planning speech production that precedes the speaking by a considerable time that allows for the speaker to be self-reflexive about it. This, in turn, helps involve the audience in that it is permitted to witness these processes rather than just be presented the results. For another, these speakers signal a part of the overall structure of the talk, namely the first unit – an offer and a request, respectively.

In this section (8.3), I have identified the strategies used by speakers at TED Conferences to open their talks. Furthermore, I have illustrated and analyzed the most important ones with the help of examples, pointing out the similarities and differences to vlog openings. The following section is concerned with TED Talk closings.

8.4 TED Talk closings

TED Talk closings are surprisingly uniform, and in their uniformity they are different from vlog closings. Like the openings, the closings have a recognizable edited credits

section, which is the same for each talk. It includes a single, dynamic sound and the TED logo (cf. screenshot below).

Figure 8.5, TED logo

For a video editor to decide where exactly to put this credits section, they must make a decision regarding when the end of the talk has been reached. As opposed to a vlog, this seems fairly easy for a talk where the audience is present: the talk is over once the monologue condition is no longer in effect, that is, when the speaker is not the sole floor owner anymore. Thus, as an editor, one only needs to look for the moment when someone other than the speaker makes a significant contribution, in most cases the whole audience starts to applaud. If this is treated by the speaker as a non-violation of his or her right to the floor, in other words if they are not waiting to resume their monologue, audience applause can be treated as a structural indicator of the end of a talk.

While this makes the editor's job fairly easy, as he or she can rely on the interactants' orientation to the unfolding interaction, the question becomes: how does the audience know when the talk is over and they can start to clap (note that it seems to go without saying that clapping is, indeed, the appropriate audience contribution, which in turn indicates that not just speaker and audience have certain expectations, or in other words, follow a script, but also the outside observer)? This question is at

the heart of the difference between vlogs and TED Talks because the closings of TED Talks are the product of joint negotiation between speaker and audience, and the closings of vlogs are not. Hence, there is a strong requirement for interaction. This section is concerned with how this interaction develops successfully so that all participants understand, signal understanding, and agree as to when they consider the talk to be over.

The strategies to achieve this are well adapted to this situation, in that they all, inevitably, are based on the speaker's unmistakable signal to the audience that they are handing over the floor. By far the most common way to do this is thanking the audience, which is done by 16 of the 20 speakers. Two speakers use the phrase "an idea worth spreading", which is the slogan of the TED organization. One says "that's all I have", and another one takes the appearance of someone else on stage as a cue to stop. The following three examples illustrate the most common case, where a speaker thanks the audience in general, or thanks them specifically for their attention.

Example 8.9

1   my story: has been about my journey from organs to cells.

2   a journey through controversy,

3   inspired by hope.

4   hope that as we age,

5   you and I may one day celebrate longevity.. with an improved quality of
        life,=

6   =thank you.

7   ((audience applause))

Example 8.10

1   if the divine cockeyed genius assigned to your case.

2   .h decides to le:t some sort of wonderment be glimpsed for just one
        moment through your efforts,

3   then olé.

4   and if not.. do your dance anyhow.

5   and olé to you nonetheless,=

6   =I believe this and I feel that we must teach it,

7   olé to you nonetheless.

8   just for having the sheer human love .h and stubbornness to keep showing
         up.

9   thank you.

10   [((audience applause))]

11   [thank you.]

12   June Cohen: olé

Example 8.11

1   and if we can chan- understand leadership like that,

2   I think if we can redefine leadership like that,

3   I think we can change everything.

4   and it's a simple idea,

5   but I don't think it's a small one.

6   and I want to thank you all so much for letting me [share it with you today.]

7   [((music))]

8   (audience applause drowned out by music)

What is common to all three of these examples (and all others where a speaker thanks the audience) is the sequence of final summarizing statement followed by an acknowledgement. And in each case, this sequence is followed by audience applause. This is a strong indicator that a simple "thank you", whether it is appended directly to the final statement as in Example 8.9, or added after a slight pause as Example 8.10, or even as part of a more elaborate phrase as in Example 8.11, has precisely these functions: to signal to the audience that the talk is over and to prompt them to display an appropriate reaction, namely applause.

This type of closing is radically different from that of (telephone) conversation or even vlogs. While telephone closings typically have a pre-closing plus closing structure, which is the scaffold along which interactants negotiate the ending of a phone call, and many vloggers adopt this structure in an adapted version, TED Talks do not parallel this closing structure. A pre-closing has the function of giving an interlocutor the chance to raise a yet unmentioned topic before the conversation is actually terminated. Thus, it seems that as opposed to vlogs, which have an inherent remnant of conversationality, TED Talks are not meant to be interactional for as long as the speaker has the floor. So the interaction mentioned above consists in dictation rather than negotiation.

The reason for this is easily explained: the end of a TED Talk does not mark the end of the speech event, as all interactants are still present in the room, the speaker is still on the stage, and this speaker still receives the bulk of the attention. What changes in the setting is the different organization of turn-taking that arises after the talk. The end of a vlog, on the other hand (and also the end of a phone call), at the same time marks the end of the speech event, the shutting down of the mediating channel. Synchronous interaction does not change, it stops completely. In other words, what is achieved by TED speakers by saying "thank you" is not a closing, but a transition.

An alternative to thanking the audience, at least for two speakers, is the phrase "an idea worth spreading". The following example clearly shows that it has the exact same function that thanking does (note that the speaker's "thank you" comes only after the audience has started to clap).

Example 8.12

    1   I believe .h that the more time we spend choosing,

    2   to run the deep inner-peace circuitry of our right hemispheres,

    3   the more peace we will project into the wor:ld,

    4   and the more peaceful our planet will be.

    5   and I thought THAT was an idea worth spreading.

    6   ((audience [applause))]

7   [thank you]

As in the previous examples, there is a summarizing statement (l. 1-4), followed by "and I thought THAT was an idea worth spreading." (l. 5), which prompts the audience to take over the floor and give applause. The other two cases that do not include an acknowledgement as audience prompt (Examples 8.13 and 8.14) resemble Example 8.12:

Example 8.13

1   because when we work from a place,

2   I believe,

3   that says I'm enough. (2)

4   the:n we stop screaming and start listening.

5   we're kinder and gentler to the people around us and we're kinder and gentler to ourselves.

6   that's all I have.

7   ((audience applause))

In this example, the functional slot of the acknowledgement is taken by "that's all I have." (l. 6). The original TED transcript, which is available on the TED website, actually includes a "thank you" by the speaker which does not seem to occur in the video clip. At best, the speaker silently mouths something similar to a "thank you" after audience applause has started, but it is not at all produced either visibly or audibly. Clearly, though, the transcriber seems to have been expecting this very common element so confidently that its actual occurrence was not necessary for him or her to have included it in the transcript. The following example is sufficiently different for the transcript to be a faithful record; it is structurally deviant in that the closing is constructed as imposed from the outside, rather than developed by the speaker, as in all other examples.

Example 8.14

1   so I consider them as to be as successful as uh,

2   .h as uh Lewis Alcindor or Bill Walton,

3   or uh many of the others that we had,=

4   =there was some outstanding uh outstanding uhum uh players.

5   I've- have I rambled enough?

6   I: I was told that when he makes his appearance I was supposed to shut up.

7   ((laughter, audience applause))

The speaker does not produce a summarizing statement of the whole talk, but is perhaps in the middle of an episode or wrapping one up (this is not entirely clear). What follows is not an acknowledgement, but a question that directly addresses the issue of ending the talk "I've- have I rambled enough?" (l. 5). The term "to ramble" initiates a slightly self-deprecatory mood. This is continued in line 6, which characterizes him as being controlled by others: "I: I was told that when he makes his appearance I was supposed to shut up.". This seems to be taken as humorous by the audience, and as such it does not represent a complaint, but it functions as the prompt for the present audience to show recognition of the end of the talk by clapping. So here, the sequence summarizing statement plus acknowledgement is replaced by the speaker's reaction to another person's appearance on the stage and a subsequent abrupt but humorous, apparently inevitable stop.

In light of the genre question previously discussed, the impact of editing becomes central. As discovered in this section, the end of the talk itself is easily identified. The task of the editor, though, could be construed as something more complex than so far assumed. If there really is a conscious effort by the TED organization to provide the online genre TED Talk, the decision as to when to cut a video to mark its end point must be made in conformity with genre standards. This, I find, is not done in a consistent manner. Example 8.11 illustrates that the closing credits music can overlap with the last words of the speaker, completely drowning out the audience reaction. In contrast to that, Example 8.10 shows the conference host coming back on stage, giving

individual feedback along with the audience's extensive clapping. Example 8.9 is midway between the two: the video shows some of the audience's clapping, and then blends over to the closing credits. This inconsistency in editing may well be the result of different kinds of organizing practices at the various conferences, or simply of different editors' influences. Or, quite possibly, there is no standard that editors knowingly adhere to other than the very obvious one that the following speaker's talk should not be part of the video.

8.5 Audience involvement

The comparison of vlogs with TED Talks in terms of audience involvement necessarily takes into account the fundamental difference in setting: TED Talks are given to a physically present audience and, while the audience generally does not have a speaking turn, it can nonetheless demonstrate its presence and uptake of what is being said on stage both audibly and visibly. Therefore, the following section mainly focuses on those strategies that impell the audience to react, which usually comes in the form of laughter, murmuring, applause, or loud whistling. A note on the presence of these reactions in both the recordings and the transcripts is in order: Most of the audiovisual recordings seem to capture the speaker's verbal output in exceptionally good quality, whereas with the audience's sounds it is not entirely clear how much detail the microphones have picked up. When there is overlap of speaker and audience, the recording usually favors the speaker, making it impossible in some cases to judge whether there was an audience reaction at all. This, of course, has an influence on the transcripts. The original transcripts available on the TED website include audience laughter or audience applause only occasionally (even when it is clearly noticeable on the sound track), and as can be expected from a non-linguistic transcript, laughter was not transcribed where it occurred, that is for example, in the middle of grammatical units, but after whole sentences or even what the transcriber had decided is a paragraph in the written representation of spoken language. Similarly,

overlap is not recorded in these transcripts. Thus, to give a fair representation of audience reactions, the transcripts had to be adapted and refined accordingly.

With this sort of evidence of immediate audience involvement available, it would be negligent to ignore it in favor of potential or assumed audience involvement, as was done in Chapter 4 with vlogs, where this type of audible audience reaction is not available (note, of course, that the actual visible audience involvement in the form of written comments has received ample attention in Chapter 7). The present section therefore goes beyond discussing matters related to participation framework and repetition in TED Talks and shifts to triggers in the talks that have laughter, applause or other reactions as a consequence.

Before beginning this analysis, there should be a general characterization of the TED audience, especially compared to the opaque vlog audience. All TED Talks are given at a conference where all talks are plenaries. Accordingly, some of the audiences are fairly sizable (one speaker mentions the figure of roughly 1600 people he is talking to). Attendees of a TED conference have to apply, and the website offers the following description:

> The TED Conference application is not a typical conference registration form. We're looking for people who are likely, in our judgment, to be a strong contributor to the TED community and/or the ideas discussed at TED-- and/or the TED Prize wishes and projects that come out of the conference.
> Before you apply, here are a few things to know:
> Set aside an uninterrupted hour to fill out the form. The questions touch on your goals, accomplishments and connection to your community. You definitely don't want to write this on the fly! (cf. TED Conferences » Attend » TED Conference » Apply for an invitation to TED. (06.05.2013) [no author, no publishing year]).

Judging from this passage, it seems that in order to be accepted into a TED conference one must make a considerable effort. Given the very heterogeneous nature of the talks

in terms of topic and speaker background, especially when compared to an academic conference where, generally speaking, all attendees have similar educational backgrounds, or are united by their interest in similar topics, there must be an interest in hearing these talks regardless of the topic. Apparently, the name "TED" generates enough trust regarding quality and relevance of the talks that individuals interested in attending are willing to make such an effort.

Not only is the TED audience literally a selected one, but as they are assembled in a closed off space, all of them hear each talk in its entirety. Neither of these two factors can be guaranteed for a vlog audience. A vlog audience is self-selected, and since there are no restrictions regarding access to a video, vlog viewers do not have to apply and do not have to be strong contributors to any community. At the same time, vlog viewers can, without disturbing the vlog production or anyone else's viewing pleasure, stop watching a video by navigating somewhere else online or offline. The organization of participation in the two settings is thus considerably different. In vlogs, both speakers and audience self-select, whereas at a TED conference, speakers and audience are selected and invited. In other words, vlog participation is governed by, if one were to use economic terminology, the 'invisible hand', while TED participation is governed by a very visible hand, namely a rather formal process of selection.

While this presumably leads to vast differences in the audience structure between the two settings, exactly what these differences are can only be speculated on. Based on the character of audience participation in vlogs, which encompasses a wide range from downright hostile comments (see Introduction for brief comments on flaming) to friendly, encouraging or otherwise positive remarks, one would assume that vlog viewers as a whole represent a fairly heterogeneous audience. TED audiences, in contrast, would have to be characterized as a homogeneously benevolent audience. When making these observations, however, it is paramount to keep in mind the differences in setting. The anonymity of online communication apparently fosters behavior where the preference for agreement among interactants (see for example Sacks 1987) is suspended, which is not the case at TED conferences. This is not to say

that anonymity is the decisive factor – the expression of disagreement in face-to-face encounters is easily imaginable at, for example, academic conferences. In the two settings under discussion, however, it seems to have a certain amount of influence. The cause for the seeming benevolence at TED conferences might be that someone's mere presence signals their status as worthy to be there according to the organizers, a somewhat elitist mechanism, which might also be operative at academic conferences. The variety of topics and speaker backgrounds at TED events, however, do not give rise to the competitive character that can make an academic conference confrontational.

As previously mentioned, audience reactions take the form of laughter and applause (very frequently), murmuring and gestures (rarely) and whistling and cheering (very infrequently). But neither the transcripts nor the recordings show any evidence of members of the audience taking any turns at talk in a conventional sense, that is, turns made up of spoken words, except for a single answer to a yes/no question in one of the twenty talks, which might have a rhetorical character anyway. In the following, I discuss humor and the elicitation of laughter as a speaker strategy that causes the most active audience involvement apart from opening and closing strategies already discussed above.

8.5.1 Humor and laughter in TED Talks

In every one of the twenty TED Talks, laughter represents a major means for the audience to signal uptake. This section specifically looks at this elicitation as a form of interaction in a stage monologue. While some speakers seem to have planned only a few, predictable occasions for humor while giving an otherwise somewhat serious talk, others expand on their humorous 'give-and-take' with the audience, comparable to a stand-up comedian. The following example shows one of the more pre-planned uses of humor that trigger laughter as an audience reaction.

Example 8.15

     1   a few years ago.

     2   I felt like I was stuck.. in a rut,

     3   so I decided to follow in the footsteps of the great American philosopher,

     4   Morgan Spurlock.

     5   ((audience [laughter))]

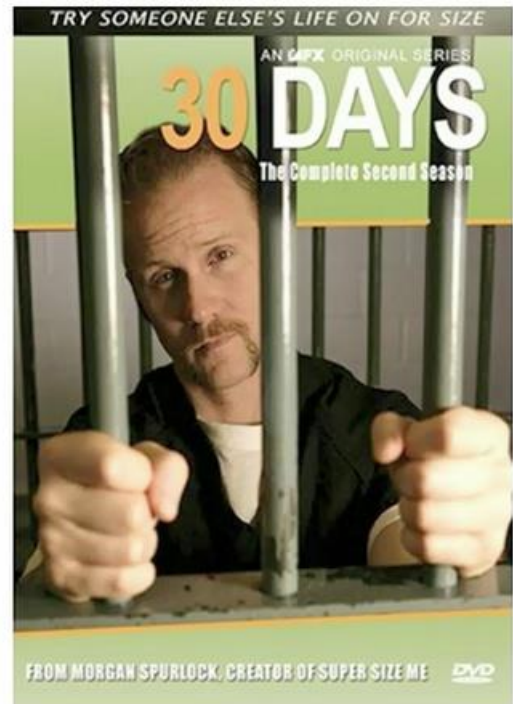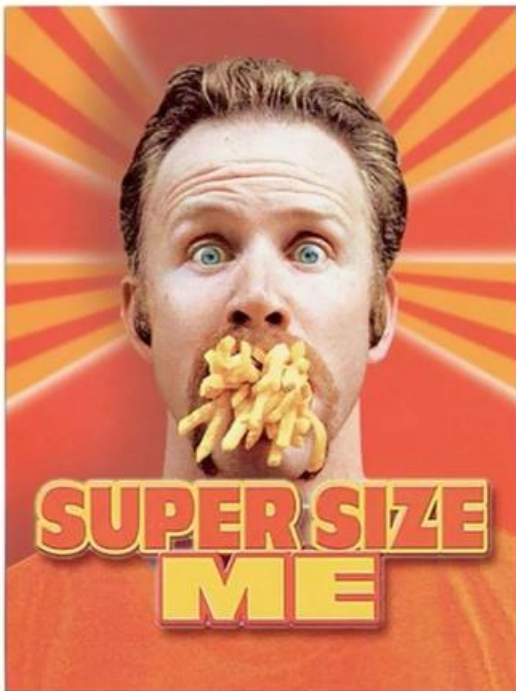   6   [and try something] new for 30 days.

Figure 6, media use during lines 4-6

The speaker creates two opposing scripts by describing a contemporary film maker, Morgan Spurlock, as "the great American philosopher,". The effect is enhanced by the use of visual media: Figure 6 shows the screen that is projected for the audience, showing two DVD covers of films produced by Spurlock (this screen makes up the entire video window from lines 4-6; it is unclear how long it is up in the lecture theater, as this is not visible in the subsequent video). The audience laughter is brief, and it seems more like a signal of uptake rather than a show of real amusement. Accordingly,

the speaker does not pause for long, but creates slight overlap by continuing his syntactically complete phrase.

A higher, more intricate level of speaker audience interaction through humor and laughter can be identified in Example 8.16. The speaker is talking about the numbing of both positive and negative emotions, so when using the first person pronoun in line 6, she is not necessarily talking for herself, but more generally for, in her own terms, the present 'US adult cohort', whereas in lines 10-12, she is solely speaking for herself as a social scientist.

Example 8.16

    1   here's vulnerability.

    2   here's grief.

    3   here's shame.

    4   here's fear.

    5   here's disappointment.

    6   I don't wanna feel these.

    7   I'm gonna have a couple of beers and a banana nut muffin.

    8   ((audience [laughter))]

    9   [I don't wanna] feel these.

   10 and I know that's- I know that's knowing laughter.

   11 I- I hack into your lives for a living.

   12 [I know that's "uehahaha, God", {mimicks intonation of someone who just had a personal flaw pointed out}

   13 UHM.]

   14 [((audience laughter))]

   15 ((loud audience laughter))

In this passage, laughter is triggered at least two separate times. The first instance could be, similar to the previous example, pre-planned, where the build-up of negative emotions via a repetitive structure in lines 1-6 is contrasted with an overly simplistic problem solution that is based on the consumption of high-calorie food and alcoholic

drinks in line 7. After giving the audience laughter some time to die down, the speaker repeats her last phrase from line 6 before the cause for laughter – apparently to go back to a serious keying. Figures 8.7 and 8.8 show the facial expression of the speaker shift from serious to humorous.

Figure 8.7, line 9 "I don't wanna feel these"   Figure 8.8, line 10 "and I know that's-"

Rather than continuing the serious key, the speaker actually takes up the laughter she has created in the audience to both thematize it and, by doing so, expand it. Her voicing the audience in line 12, "I know that's 'uehahaha, God,'" and her comment on their laughter signal back to the audience that the speaker is welcoming this sort of interaction, rather than accepting and passing over it, or even rejecting it as an interruption of her talk. Thus, the initial laughter (l. 8) does not seem to be reduced entirely, but it is incited again already before the mimicking starts (beginning of line 12), and is given new impetus by the parody. In sum, in this passage the speaker invites laughter and subsequently encourages it to grow and thereby makes it not just a by-product of her talk, but, it appears, an essential, interactive part. The following speaker does this at an even more extended level.

Example 8.17

 1 there's a raft of research,

 2 but I know it from my personal life.

 3 if my wife is cooking a meal at home,

 4 which is not often,

5   ((some audience laughter))

6   thankfully,

7   [((loud audience laughter))]

8   but you know [if she's doing. (1)

9   she's huh,

10   no she's good] at some things.

11 but if she's cooking,

12 you know,

13 she's dealing with people on the phone,

14 she's talking to the kids,

15 she's painting the ceiling,

16 you [know she's,…]

17 [((audience laughter))]

18 doing open-heart surgery over here,

19 ((audience laughter))

20 if I'm cooking the door is shut,

21 the kids are out,

22 the phone's on the hook,

23 if she comes in I get annoyed,

24 ((some audience sounds, possibly coughing))

25 I say Terry please,

26 I'm trying to fry an egg in here,

27 [((extended, loud audience laughter))]

28 [you know, (2)

29 g- (2)

30 give (3)

31 give me a break,]

32 ((audience laughter))

33 actually,

34 you know that old philosophical thing,

35   if a tree falls in a forest and nobody hears it did it happen.

36   remember that old chestnut.

37   I saw a great t-shirt really recently which said,

38   if a man speaks his mind in a forest and no woman hears him,

39   is he still wrong.

40   ((loud audience laughter))

41   ((speaker laughs))

This type of humorous build-up is even more geared towards creating an interactive space between speaker and audience than the dynamics of the previous excerpt (Example 8.17). This whole passage is remarkable in that it represents a progression from suggestive and subtle humor, which is still related to the topic at hand, in line 4, to a tangent that seems only to fulfill the purpose of extending the humorous interaction (l. 33-39).

Lines 3 and 4 "if my wife is cooking a meal at home, which is not often," evoke a common scenario of family life and qualify it in terms of its frequency of occurrence. The laughter of a few members of the audience (l. 5), however, retrospectively adds another facet, namely that of an assessment in terms of the quality of the food the speaker's wife prepares. Apparently the mere mention of her cooking triggers in some of the audience members a frame that leads to them expect an assessment of this food, and more specifically, they seem to expect a negative assessment of the food, which would clash with the expectation of spouses protecting each other from criticism rather than exposing each other to it. What is prepared in subtlety in line 4 and merits a little laughter is fleshed out and brought to the fore in line 6, where the speaker confirms the interpretation of his previous utterance as negative evaluation when he adds a positive assessment of the infrequency of his wife's cooking, "thankfully,". Accordingly, the majority of the audience now seems to be participating in the laughter (l. 7).

In terms of timing, lines 3-10 expose not just the speaker's awareness of the audience's reactions, but a control element as well. The feeble and dispersed laughter in reaction to his subtle allusion to a potentially humorous frame is given ample space – in the representation of this transcript, one might argue it is given the space of a turn. In other words, there is a pause in the monologue encouraging an audience reaction, which is the signal to the audience that there is an ambiguous frame at work. Thus, this pause could be assigned the status of frame instantiator, while the following, openly humorous remark in line 6 acts as the main carrier of humor. The laughter triggered in this second instance is also given space, but here, if one wants to continue the turn-taking terminology, the turn is not granted without challenge. The speaker does not pause after the humorous remark, but continues on, apparently with the initial thought from line 3. This overlaps with the audience's loud laughter, causing the speaker to pause briefly, after which he tries again (l. 9), still not ceding the turn entirely. Line 9, "she's huh,", finally includes a laugh pulse by himself, where previously he retained a serious keying. Figures 9 and 10 show the speaker's facial expression at "thankfully," and right after his own laugh pulse.

Figure 9, "thankfully,"   Figure 10, "huh,"

This brief change in keying is an acknowledgement of the audience's participation, but the speaker does not seem to want to dwell on it for too long, as, in line 10, he takes over the floor again. "no she's good at some things." reflects that he did not meet the expectation of mutual support in a marriage, and that this is the cause for laughter.

Therefore while the subtle, preparatory humor is given a rather long time to sink in, the actual 'punch line' is almost skipped over by the speaker at first and only subsequently being oriented to, and even then he hesitates to give in to the humorous frame that the audience has detected. Why he does not pause when he can be fairly certain of audience laughter occurring is not entirely clear. To determine whether this works to emphasize the script opposition or whether the speaker specifically seeks to elicit an audience interruption to signal appreciation of his joking would require further, systematic investigation of this type of speaker audience interplay.[20] In sum, these two instances of laughter alone demonstrate that floor management in a monologue can be achieved through humor and, vice-versa, humor and audience reactions can be shaped through floor management techniques.

The following instances of humor and laughter in the transcript follow a two part structure, where the first part is a clear punch line that triggers an audience reaction and the second part builds on and extends the first part by supplying another punch line-like utterance similar to the first that is met with some more laughter. In both cases (l. 15-19 and 26-32), the speaker continues speaking immediately after the punch line, forcing the audience to interrupt his intonation unit. Both times, the speaker continues with the discourse marker "you know", which is semantically so unspecific that it can easily be used as a filler to facilitate online speech production and make speech appear fluent. This supports the notion that continuous talk after a punch line until laughter starts may, indeed, fulfill functions quite distinct from a simple transfer of information. It might simply function as a pause-filler until the audience has had time to react appropriately to the joke, or it may be a more complex arrangement of script opposition that is constructed, or, of course a combination of both. However, the frequency of this phenomenon in a comparatively small piece of data indicates that there are functions to this overlap that have yet to be explored systematically.

Both passages build on the discussion of multi-tasking in terms of gender stereotypes. The first uses a list of things the speaker's wife allegedly habitually does while cooking,

---

[20] See Schwarz (2010: 94), who maintains that stand-up comedians "address them [the audience] directly and try to keep their attention and earn their appreciation, but they do not wait for their response"– clearly, this is not the case in this passage, which can be labeled a comic monologue.

which quickly moves from regular, script-conform activities to "painting the ceiling" (l. 15). This first punch line receives laughter which at first overlaps with the speaker's subsequent intonation unit but he is then quickly granted sole 'floor rights' as he pauses (l. 16-17). The following intonation unit proposes an equally inappropriate activity regarding the cooking process, namely "open-heart surgery" (l. 18), and thus can be described as a second punch line to the same build-up, exploiting the humorous potential of opposing scripts without actually adding anything new. I assume that this triggers more laughter in the audience because there seems to be a 'humorous credibility' the speaker has created for himself with this audience.

The humorous potential of the second passage derives from the opposition to the first, as well as from its internal structure. Here, the speaker is describing his own abilities in the kitchen with regard to doing things simultaneously: while he characterizes his wife as exceptionally capable of juggling various activities at once, he describes himself as feeling challenged by just two parallel processes. This culminates in the exaggerated use of constructed dialogue, "Terry please, I'm trying to fry an egg in here," (l. 25-26), which contrasts a relatively simple cooking process with an accomplished adult's ability to deal with this process as anything other than a highly complex task that requires all of his attention. Again, the following loud audience laughter (l. 27) is not given the sole floor immediately, but the speaker continues with "you know," (l. 28). Subsequently, the speaker tries to regain control over the floor (l. 29-30), but the considerable pauses in between attest to the volume and dominance of the laughter, which forces him to abandon these attempts. However, the structure of the previously discussed passage is paralleled: the first punch line triggers audience laughter, which is only admitted as an interruption of the speaker's next utterance. Then follows a second punch line, in the present example "give me a break,", which draws more laughter. To determine whether this strategy is a speaker idiosyncrasy or a common strategy in humorous monologues would require further research and is beyond the scope if this investigation.

The three passages I have discussed regarding humor, laughter, and the possibility for interaction that these phenomena represent demonstrate that speakers can use their dominance as floor-holders to initiate an interaction with a monologue audience that gives them the opportunity to signal their interest and appreciation. In short, although a monologue audience supposedly has a predominantly passive role, they are given the opportunity to be actively involved despite the limitations of participatory means. It is obvious that a vlog audience has no way of showing similar involvement so that this type of immediate, interactive involvement cannot occur.

Repetition has been shown to be an omnipresent element in conversation (Tannen 2007 (1989)), and Chapter 5 revealed that vloggers, too, use repetition as a strategy to achieve various interactional goals. What the comparison of vlogs with TED Talks has demonstrated so far (and, indeed, continues to do in the following sections) is that the splitting off of the editing process to result in two sibling genres is responsible for a good part of the differences one can trace between vlogs and TED Talks. This is also partially true for repetition: those instances of repetition brought about by or geared towards the editorial process are very unlikely to occur in TED Talks, for the speakers have a present audience in the room, whose influence on the talk is presumably much greater than that of an imaginary future online audience – if this exists at the time of speaking at all. To illustrate: none of the twenty TED speakers ever allude to the editing of the video (nor any other speaker I have seen outside the corpus), and none of them seem to talk directly into a camera; instead they all direct their attention towards the audience in the room. Therefore, TED speakers will most likely not begin their monologue with the attitude that what is not produced correctly the first time can be edited or re-shot, as the synchronous interaction with the audience does not allow for this. Vloggers do not speak to a physically present audience in a synchronous interaction, therefore they can rely on the editorial options to delete or alter whatever they do not wish to be part of their communicative efforts.

The point of interest in the use of repetition in TED Talks compared to vlogs lies in those instances where repetition is used when active involvement of the audience is

taking place. The example previously discussed in the treatment of the use of humor to elicit laughter (8.16) contains such a passage:

Example 8.18

1   here's vulnerability.

2   here's grief.

3   here's shame.

4   here's fear.

5   here's disappointment.

6   I don't wanna feel these.

7   I'm gonna have a couple of beers and a banana nut muffin.

8   ((audience [laughter))]

9   [I don't wanna] feel these.

10   and I know that's- I know that's knowing laughter.

11   I- I hack into your lives for a living.

12   [I know that's "uehahaha, God", {mimicks intonation of someone who just had a personal flaw pointed out}

13   UHM.]

14   [((audience laughter))]

15   ((loud audience laughter))

The repetition in this passage occurs in line 9 as a lexically identical rendition of line 6 "I don't wanna feel these.". Though semantically entirely congruous, the pragmatic functions of the two are very different. The first, line 6, is a negative assessment regarding all of the elements of the list leading up to it (l. 1-5). As previously explained, this principally problematic topic of negative emotions is exploited for humorous purposes by its script opposition with a simple (but obviously misled) solution (l. 7). It successfully elicits laughter from the audience (l. 8). While the previous analysis of this passage explained the elicitation of laughter, this analysis aims at pointing out the pragmatic function of the repetition in line 9, namely floor management. The speaker

interrupts the audience laughter, cutting it short to retrieve sole ownership of the floor and signaling that the interaction just elicited is to be discontinued in favor of her own elaboration on the topic she has started. Her own subsequent comment on the laughter she has just dismissed must be attributed to either a change of mind or even a spontaneously planned build-up of laughter, whose full impact can develop after a drawn-out phase of a pretense of seriousness. The point this passage illustrates about repetition in monologues is, however, that it can be an effective tool for floor management when the otherwise unchallenged dominance of the speaker is being negotiated. Clearly, no such pragmatic function is necessary in vlogs.

Another type of repetition that is characteristic of one TED Talk in particular gives the talk an almost poetic tone. The pattern consists of sentences ending in a noun or a noun phrase, which is then repeated to be the head of a new noun phrase. This pattern is actually recognizably marked in the transcript available from the TED website:

Example 8.19

1   But my joyful period was marred by solemn and morbid thoughts -- thoughts of walking through the prison's high-security death row, as this was the only route to take me to the makeshift operating room.

2   But, in my life, the one fulfilling skill that I had was now invoking feelings of conflict -- conflict ranging from extreme sorrow and doubt at dawn to celebratory joy at engrafting the gift of life at dusk.

3   It made me wonder if there could be a better way -- a way to circumvent death and yet deliver the gift of life that might exponentially impact millions of patients worldwide.

4   And so guess what, scientists around the world and in the labs are racing to convert aging adult cells -- aging adult cells from you and me -- they are racing to reprogram these cells back into more useful IPS cells.

5   And in our lab, we are focused on taking fat and reprogramming mounds of fat into fountains of youthful cells - ((audience laughter)) - cells that we may

use to then form other, more specialized, cells, which one day may be used as cell transplants.

6   My story has been about my journey from organs to cells -- a journey through controversy, inspired by hope -- hope that, as we age, you and I may one day celebrate longevity with an improved quality of life.

With the exception of the addition of audience laughter in excerpt 5, these are taken from the transcript in their original form. They are obviously not organized into intonation units, and there are no symbols to mark prosodic shifts. But what is striking (and thereby the reason for my using these excerpts rather than a transcript that follows the conventions used otherwise) is the use of "--" to mark this particular pattern of repetition. Thus, the TED transcriber considered it noteworthy and consequently chose a way to emphasize visually that is more marked than, say, a comma.

Although the TED transcripts are obviously not designed to represent spoken language faithfully, that is, ready for linguistic investigation, the transcript of this particular talk comes very close in some respects. That is not to say that the transcript was (possibly) produced by a linguist. Rather, the speaker gave a performance that, when transcribed, strongly resembles language that is composed as written language. The complexity of the sentences produced throughout the entirety of the 18 minutes the talk lasts indicates elaborate preparations for this talk, perhaps even including a rigorous learning by heart of the whole speech. So what this pattern of repetition, alongside the generally elaborate grammatical structures, hints at is the pressure to give a convincing performance without subsequent editing, which does not exist in vlogs. Hence, a fairly general statement, which would certainly benefit from further investigation, can be made: the efforts necessary to give a successful talk, where successful means that the speaker is entirely comfortable with having an audience hear and see all of it, have to be made before the actual performance in the case of TED Talks, but in the case of vlogs, some of them can be made after the talk in the

editing process. In other words, the presence of a live audience has the effect of forcing a speaker to an 'on-time delivery'.

8.6 Visual media in TED Talks

Gestural communication, for linguistic research, is a curious feature in vlogs, as there appears to be evidence of the capacity for adaptation to the environment that humans have. Especially the virtual space that this interaction takes place in demands quite complex pointing gestures, which speakers are capable of producing quite effortlessly. TED Talks are given on a stage in a lecture theater, and while few of the speakers (three of the twenty) show awareness of the fact that the footage might be uploaded to the internet, this does not seem to affect the speakers' gestures at all. In other words, no TED speaker points at the comments section underneath their video, nor at any other part of the TED website outside the video screen. Also, no TED speaker adapts their three-dimensional pointing to the two-dimensional website space.

In this section, the two part production process (1 the actual talk, 2 editing and uploading) receive attention, as the editor's job is fairly complex compared to vlog editing since there is a choice of footing from six cameras and the screen, which need to be streamlined into one short film. What happens to the footage when becoming an online TED Talk should be particularly interesting in light of the genre/sibling genre discussion.

As mentioned before, vloggers face certain restrictions because of their status as laypeople: their camera and sound equipment, generally speaking, is much less sophisticated than that which is available to the TED conference organizers. As a result, vloggers mainly rely on the use of one camera, which is usually static, filming a relatively small space in which the vlogger sits close to the camera to be within the microphone's pick up range. Compared to this, the recording equipment at conferences can cover a much larger space from various angles and with consistently good sound quality. This includes images of the audience.

However, the large differences derive not only from the perspective of technology, but from the setting itself: TED speakers are on a stage where they are visible to everyone in the audience, and they know that everything else there is on stage, including the screen, is visible to everyone in the audience. This is the case without any additional efforts: it is a fact everyone in the room can take for granted. By contrast, vloggers have to be exceedingly careful, that is, set up and check their equipment and its proper functioning, and to ensure that they are visible and audible on the footage they are recording. In short, vlogs, as mediated discourse, depend on technological equipment, whereas TED Talks, as their stage talk genre rendition, do not. Their talk is, of course, supported by the amplification of their words through speakers, but for the basic function of establishing contact with the audience in the lecture theater they have no such essential and defining need for equipment as vloggers do. For a TED Talk to become an online video, this dependence exists – but the professional use of equipment facilitates this in such a fashion that TED speakers do not even have to be aware of their producing a stage talk and the raw material for an online video simultaneously. Thus, as pointed out above, in the first phase of production there is no indication that what is being produced is communication specifically geared towards an online environment – this is all done in the editing process. In vlogs, on the other hand, both the first and the second production phase are characterized by their online environment design.

One result of the editing process in TED Talks is the loss of visual information. When editors choose between several footing options, for example they might show the whole screen but not show the speaker at all, features such as facial expression, gestures etc. are not available to a viewer of the video. Figures 12-14 illustrate such editorial choices: Figure 12 is a close-up of the speaker, and her left hand is visible in the lower right hand corner. The part of this hand gesture that takes place below shoulder level is thus invisible to the viewer. Figure 13 is a mid-range shot of the speaker (taken just two seconds after the one in Figure 12). As the speaker is actually talking about the pose she is demonstrating, but only referring to it with the deictical "like this", the successful transmission of information depends on the visual channel.

Accordingly, the editor has to switch from close-up to mid-range. Figure 14 shows the image projected on the screen, but the speaker is not visible at all. In one talk, this screen only footage is up for incredible 24 seconds, during which the speaker continues her talk but cannot be seen. This demonstrates that these decisions in the second phase of production can be quite drastic. A visual element that is added in the editorial process is a caption that reads "For more talks visit TED.com" (see Figure 11).

Figure 11

As the present description and analysis of TED Talks is based on their online version, these editorial choices are part of the results.

Figure 12 (10:20)   Figure 13 (10:22)
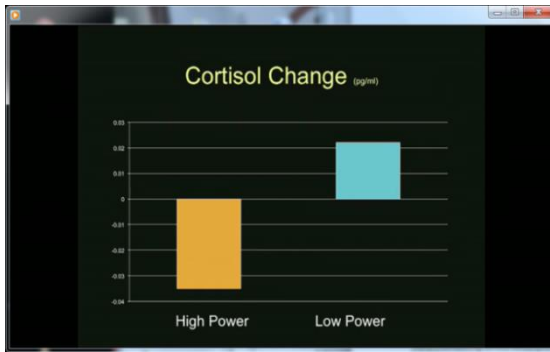
Figure 14 (12:17)   Figure 15

Out of the twenty TED Talks, ten are given without any additional visual media: the speaker is standing (or in one case sitting) on stage and all visual information emanates from him- or herself, rather than a screen. Nine speakers use a large screen behind them on stage to project images and/or text onto it, one speaker uses a flip chart to draw and write on during the talk. In some of the talks where speakers used a screen, the images of themselves or the audience recorded by the cameras were projected live onto the screen so that the audience could see the speaker in person and on screen simultaneously (cf. Figure 15). One speaker has, in addition to the screen, an object on stage to demonstrate its structure, another uses a grand piano on stage.

8.7 Pointing at the audience

As mentioned previously, TED speakers do not point at the virtual space shared with their viewers online, which, as I have argued in chapter 6, is a multimodal device that enacts the drive to establish a connection. Thus, there is no meta-referential gesturing that indicates the twofold existence of this genre. TED speakers, however, use a whole range of gestures that originate and have referents in their three-dimensional physical surrounding, as one might expect. Covering them all, though a laudable topic, is beyond the scope of this comparative analysis. Consequently, a subset of these gestures is presented in more detail. In terms of establishing a connection with the audience, these gestures come closest to the meta-referential pointing gestures that occur in vlogs. The gestures focused on here are pointing gestures at the audience in general or direct physical contact with a member of it, and I demonstrate how the use

of media at the time of speech production, as well as at the editing stage, has an impact on language use.

Pointing at one or more audience members seems to be one of the most direct non-verbal strategies to address the TED audience. In the edited online version, the visual audience reaction, if there is one, is rarely shown (although in Examples 8.20 and 8.21, this information is available). In my data-set, there are only six talks where such audience directed pointing occurs. I distinguished these from beat gestures that were carried out with the index finger that roughly pointed in the direction of the audience – these latter ones are not discussed here. One clear-cut case is shown in Figures 16 and 17, where a speaker points at an audience member, while identifying her verbally (l. 5).

Figure 16

Figure 17

Example 8.20

    1   and you just stopped.

    2   and you stared.

    3   it was creepy.

    4   ((audience laughter))

    5   this girl right here knows exactly what I'm talking about. {points at audience}

    6   ((audience laughter))

Lines 1-3 are constructed dialogue addressed to the speaker's earlier self as part of a narrative, which in itself elicits laughter as a reaction from the audience. The pointing and verbal mention of a particular audience member in line 5 builds on this established humorous narrative by likening the audience member to the character of his narrative in terms of the experience of being stared at by him. As the audience member is not shown in the footage, and therefore both gesturing and gaze behavior cannot be described exactly in relation to their target, the following analysis is based on an assumption that a slight but noticeable gaze shift and the enactment of a certain participation framework suggest. During lines 1-4, the speaker's gaze is fixed in one direction; supposedly he is staring at the person he subsequently identifies as "this girl right here" (cf. Figure 16). With line 5, the speaker adopts a pointing gesture and simultaneously shifts his gaze slightly to the left (cf. Figure 17, viewer's perspective). If the pointing gesture is directed at the particular audience member, which is in line with the verbal reference "this girl", gaze and gesture have different targets. The participation framework that is employed in this little break from the narrative clearly excludes the audience member from the direct addressees of the statement and gesture as she is talked about rather than talked to. In sum, verbal communication in co-ordination with gaze and gesture behavior indicate a switch in addressee status, and thus make the single audience member not exactly the butt of a joke, but a subject of sympathetic laughter for sharing the allegedly "creepy" (l. 3) experience of enduring the speaker's stare. As previously mentioned, the audience is not shown so that any

more intricate communication than joint laughter cannot be taken into consideration here. This means that it is not clear whether the female audience member singled out actually displayed any kind of reaction to the speaker's stare. However, the joint audience laughter can serve as evidence that there is consent regarding the interpretation of the singling out of an audience member as funny, rather than hostile.

The whole passage certainly demonstrates that an editor can choose to not show the audience in cases where their behavior might have triggered something that occurred on stage. Likewise, the pointing and personal deixis might have caused other audience members to try to catch a glimpse of the one person to see whether there are traces of any foregoing triggers. This freedom to direct attention to anything in the room that the live audience members (and the speaker) have is not available to the video viewers. Thus, a crucial factor that distinguishes the online version of TED Talks from their live counterparts is the limited choice of semiotic systems, determined by the editor, as opposed to the rich, multivalent environment of meaning making that the conference attendees can participate in as they see fit.

The following passage (Example 8.21), which involves a pointing gesture, demonstrates the immense complexity caused by the choice of channels available to conference attendees, which subsequently has to be reduced to 'fit' into a video without appearing incoherent. In contrast to Example 8.20, where the occurrence of a possible trigger was not made available to the online viewers, the following passage offers footage that suggests a cause and effect relationship in a chain of communicative acts. Note that underlined passages in the transcript here indicate that the speaker is shown in the footage, while the audience is shown during the rest of the passage.

Figure 18

Figure 19

Example 8.21

1   maybe you're hunching,

2   uhm crossing your legs,

3   maybe wrapping your ankles,

4   sometimes we hold onto our arms like this,

5   uh,

6   [((audience laughter))]

7   sometimes we uh [spread out,

8   {turns around, presumably to screen}

9   I see you.] {points at screen, turns back around to look at audience}

10   um so I want you to pay attention to what you're doing right now,

During line 2, the image shown in the video switches from the speaker to the audience (cf. Figure 18). This continues until "out" in line 7. During this time, one audience member (circled white) looks up from the booklet he is reading and adjusts his posture to sit up more upright. He also begins to smile noticeably. While the speaker is continuing her talk, the audience interrupts her with laughter. As the talk does not seem to offer any obvious cause for this laughter, its trigger must be searched for elsewhere. The editor's decision to show footage of the audience, rather than the speaker on the stage, forces the viewer of the video to adopt the perspective of the speaker, who probably looks at a similar image of the audience. She herself must look for the cause of the laughter if she does not want to ignore it, as obviously she knows she was not trying to excite this reaction. The cut back to the speaker on stage comes with "out" (l. 7), so just after the audience laughter has started. This is just in time to capture the speaker's reaction to the laughter. Presumably influenced by audience members' gaze towards the screen behind the speaker, she turns around to look at it herself. While the speaker can now identify the cause for laughter visually, the viewers of the video cannot: the screen is actually never shown in the video during this whole passage. Thus, the viewer has to infer, based on the speaker's "I see you" (l. 9), that what is being shown on the screen is the footage of the audience which was just shown in the video. Hence, the cause for laughter seems to be the self-conscious adjustment of the audience member's posture, while the speaker on stage is talking about postures, and the audience can see themselves on the screen. Again, all attendees of the conference talk can freely direct their attention to any active or inactive semiotic mode, while online viewers fully depend on an editorial choice that compensates for the lack of this freedom.

The editorial decision just described shows that the audience-speaker interaction in lines 6-9 is understood by the editor as a multi-modal adjacency pair, where the laughter is perceived by the speaker as a first pair part that causes her to react by indicating her understanding of the laughter through a gesture. The editor provides not only visual coverage of this adjacency pair, but also of its lead-up, namely the

audience member's posture shift, thereby demonstrating recognition of the online viewers' dependency on a selective editing process.

What both pointing gestures discussed above (Examples 8.20 and 8.21) have shown is that speakers on the TED stage can use pointing not only to initiate complex processes of interaction with a predominantly passive audience, but also to react to such initiations. The discussion has also allowed a glimpse of the challenging task facing an editor to adequately capture such processes of meaning making for an online video audience. This stands in contrast to the production of vlogs: vloggers only reach an online audience, so their communicative efforts are tailored to the mediated channels. Their subsequent editing is thus potentially facilitated, as it generally does not have to make offline processes of interaction accessible to an online audience (although see exceptions in chapter 4 on participation frameworks). In other words, TED editors are facing a different challenge of constructing complex multilayered representations (cf. Hutchins and Palen 1997) than vloggers do.

8.8 Online commenting on TED Talks

The previous sections on TED Talks have revealed strategies that speakers use to interact with their physically present audience in the lecture theater. It has become clear that the mere presence of an audience at the delivery of a monologue is a parameter that enables the adoption of strategies that cannot occur in vlogs. The present section is concerned with interaction that can and does occur in fairly similar form in vlogs, that is, written comments. The website design of the TED page is comparable to YouTube regarding certain major features: once a user has clicked on a video, he or she is taken to a page that contains a large video screen, a video description on the right of the video, the title of the talk above the video, and the comments section below the video (cf. Figure 20).

Figure 20, TED page layout

As is demonstrated in Figure 21, online visitors of a TED Talk can choose in what order they would like to read comments to the video they are watching. A drop down menu lists five options: "Newest first", "Oldest first", "Top rated", "Recently updated", "Most replies". "Newest first" is the default setting, with the most recent comment appearing at the top of the list (note that the metaphorical use of "first" when denoting a spatial

relation is indicative of our acquired spatial grammar that imposes a reading order whose direction is from top to bottom).
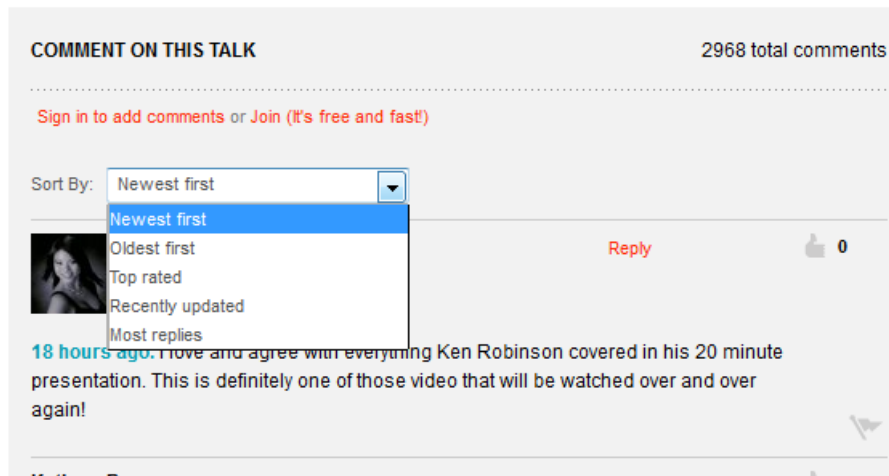
Figure 21, drop down menu

The same order can be found in YouTube comments sections, except that the vlogger's own comments and those that received the highest ratings are automatically placed at the top. Comments on the TED website also have a reply link which allows users to leave a comment that is made visible as a reply in that it appears underneath the comment it replies to and it is indented (cf. Figure 22). TED's reply function is slightly more sophisticated as it allows indenting to further levels, that is, one might leave a reply to a reply, which is then moved even further to the right.

We're not the first ones to build an electric skateboard. The earliest electric skateboards we've seen are from the 1960s, and there are lots of different kinds available today. We're big fans of all of them, and they're all helping people think differently about transportation.

Jed Lubin Reply 👍 +1

Apr 13 2013: What is your "new vision for transportation"…? What are your "new ideas for vehicles"…? All i see in your video is a remote controlled electric skateboard, every feature of which is available on current non-prototype products.

Jeong Kyu Yang Reply 👍 +2

Apr 18 2013: Why are everybody obsessing the "New ideas"? Even iPod was not the fist MP3 in market. The first iPod was not that innovative technically. It was innovative because its design was extraordinary and had easy usability.
In that point of view, Boosted board is little differ from competitors because It weighs little lighter, drives little longer, has a little better remote controller and installs smaller battery package under artistic plate. Aren't they enough to be a portable transportation?

Tim Rogers Reply 👍 0

Apr 21 2013: Hey man! Love the boards. When will these be available and on the market to buy (not pre-order) as I see you have delayed production? I have been following the progress and noticed the design is looking different to your kickstarter pitch. I am guessing this is because production is different from prototype? Last question.....i have been e-boarding for a long time now, from my experience the battery range can be drastically reduced due to the type of riding the user does, you say it does 6 miles, this is not enough. A few more cells really needs to be added as there is nothing worse than running out of juice. People want to carve not just go to the local corner store. These are just my thoughts from an avid e-boarder :)
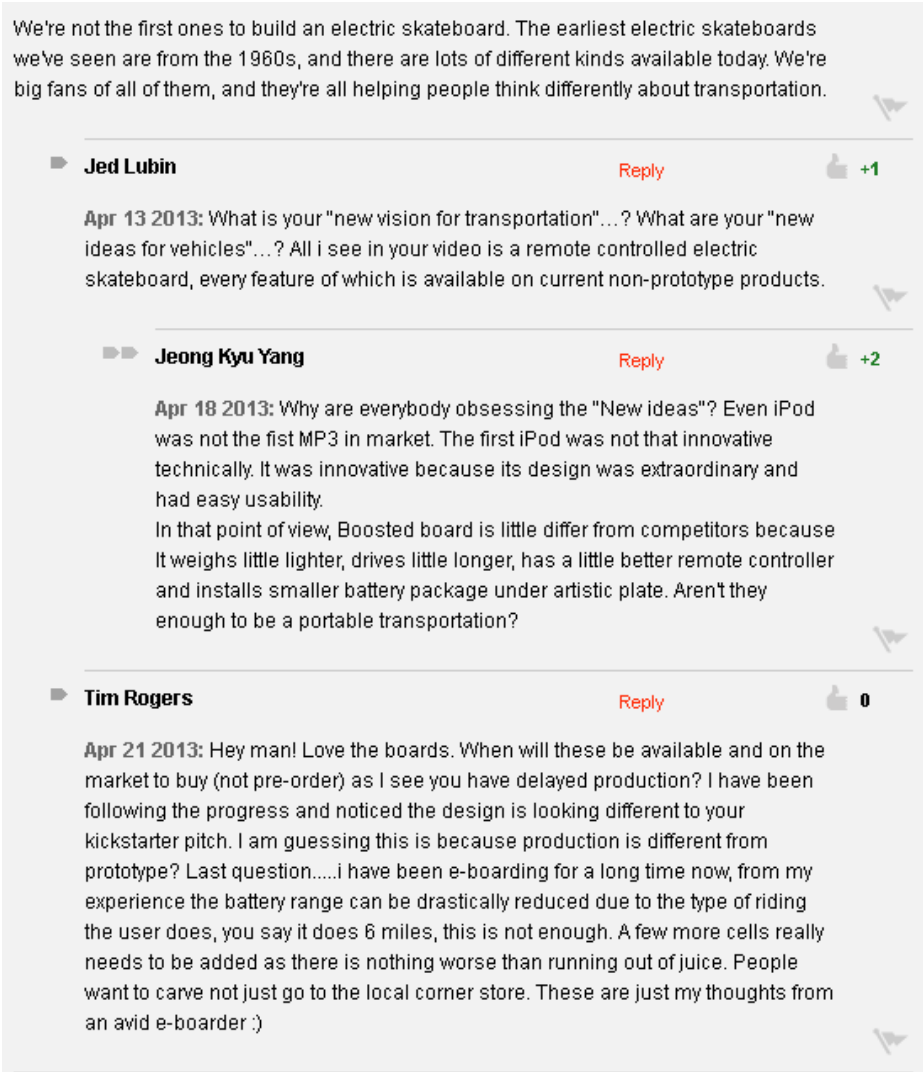
Figure 22, reply comments on two levels

The most striking difference, which has consequences for subsequent interaction, is that the TED website does not organize its contributions as videos that have been uploaded to a user profile, but they are all channeled into one interface. As a consequence, there seems to be a difference in 'ownership' of the commenting space underneath the video. This, I believe, is reflected by YouTube's placing the video uploader's comments at the top of the comments section. The precedence given to uploaders' comments over viewers' comments assigns vloggers a special status in the interaction appearing below the video. TED Talk speakers also have a special status, shown not through a particular placement of their comments, but through additional 'badges' next to their user name. Figure 23 shows such a comment from a speaker in

the comments section underneath his own talk: Mark Bezos's profile is marked as that of a TED attendee, and the yellow "100+" refers to a rating of the contributions of a TED member to the community. As other TED members can earn these same badges (see third comment in Figure 23), there is in principle no difference between viewers' comments and speakers' comments as obviously visible as on YouTube.
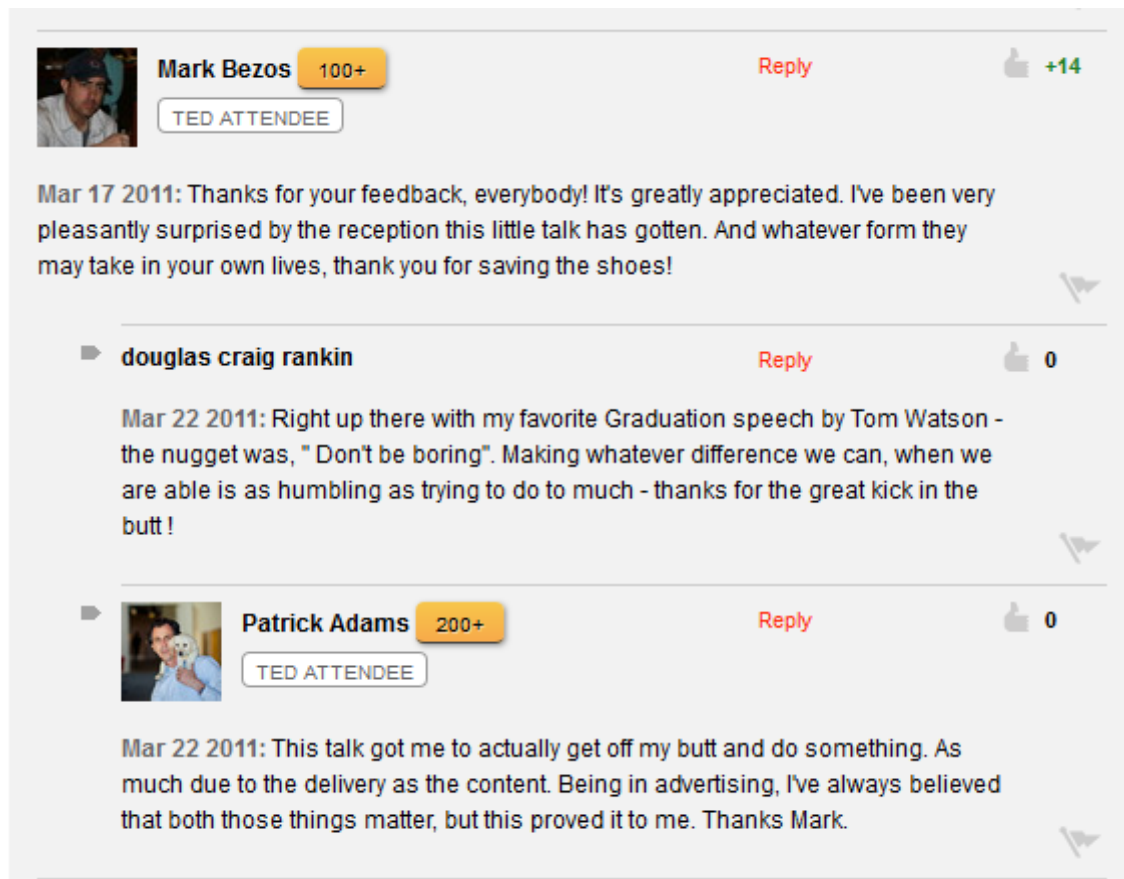

Figure 23, TED attendee 'badges'

The issue of ownership of the commenting space in TED Talks is a matter of negotiation. Figure 23 shows a comment and, judging from the use of the reply function, two replies to this comment. Both replies, however, are not particularly related to the speaker's comment content-wise: they could just as well be original comments themselves, as they do not contain any signals of coherence to the comment, but rather to the talk. So why do commenters choose to place their comment as a reply, even though they appear to be regular comments, especially when other users in this and other talks leave fairly similar comments as original

comments (cf. Figure 24)? Apparently, there is no clear convention regarding the use of the reply function that, if violated, causes other interactants to repair the misuse in some way. However, this use does not appear to be entirely random. What both replies in Figure 23, as well as the comment in Figure 24, have in common is that they address the speaker directly by thanking or praising them. The users leaving replies in Figure 23, as opposed to the one in Figure 24, however, were able to take advantage of the speaker's own participation: by leaving a comment on his own talk, Mark Bezos gave others the chance to use the website's affordance of leaving a reply to extend this space he created. This extension, in turn, transformed a single comment into the speaker's own space within the space of the comments section, where the chance of the speaker's seeing the comments is actually not entirely unrealistic, because the speaker has been to this space before. By contrast, the commenting space that the passage in Figure 24 is taken from does not feature comments or replies by the speaker himself, thus there is no opportunity to extend a comment to such a space. Therefore, a commenter wanting to address this speaker has no way of differentiating.

Figure 24

The use of the reply function to leave a comment addressed to a TED speaker, therefore, represents a renegotiation of the reply function, relieving it from serving only the purpose it was named after. This use of the reply function on YouTube is completely unnecessary as the whole commenting space underneath a vlog is governed by the vlogger (the video uploader can delete comments or not approve them in the first place, so theoretically, he or she sees them all). So while TED speakers can 'own' a sub-space within the commenting section, which is headed by their comment and is followed by comments that are posted as replies, a vlogger 'owns' the whole commenting space, which is headed by their video. The creation of the TED

speaker's space, however, crucially depends on other commenters who extend it through the use of the reply function.

The ambiguity of ownership of this space is also reflected in the following example (Figure 25).
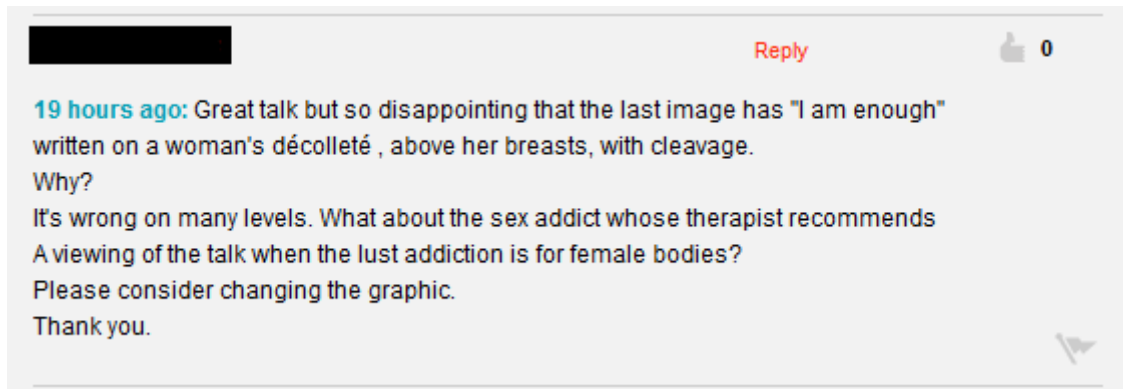

Figure 25, ambiguity of ownership

Here, a commenter complains about an image shown as part of the footage. The commenter does not address anyone in particular, but her request to remove the offensive image is clearly directed at anyone who has any influence over this editorial decision. We do not know whether the commenter has the speaker herself in mind, or anyone else in charge of the website and its contents (although the latter option seems more likely), but the double reading itself allows for the assumption that commenters are aware that not only the speaker but other people involved with TED 'own' this space.

With the lack of responsibility of such a space that is initiated by a video comes a certain passivity: of twenty online TED Talks, only six speakers participate actively in the comments section. Many of the speakers' comments are acknowledgements of the reactions to their talk. By contrast, all twenty speakers interact in various ways with their live audience. Vloggers, simply by uploading a video, interact with their online audience; and many of them participate in the interaction in the comments section, or as pointed out in chapter 7, these comments are recycled in another video.

This recycling as a form of continued interaction is a feature of vlogs that, unsurprisingly, does not occur in TED Talks. For one, the aforementioned passivity of the speakers regarding the written comments online reflects their disinclination to interact in this asynchronous mode. For another, TED speakers predominantly give only one talk (although there are exceptions, even in my small sample of twenty videos). So there simply is no opportunity for most of them to react to a comment to their talk in a subsequent talk. The one exception does refer to reactions to his earlier talk (four years had passed between the two conferences) by way of justifying his second talk. But in no way does he mention online comments to his first talk.

This section has highlighted, once more, that the fact that there is editing and some form of online management of the talks removes the speaker from their online audience. As a consequence, there seems to be less immediate interaction, or if it exists, it is enacted, for example, through the redefinition of the website's affordances. Vlogs, which are created for the online space, are the heart of an environment that fosters mediated interaction.

9 Conclusions

The study of video blogs as a genre with particular features allows for several concluding remarks. This section not only brings together all of the findings from the previous chapters to form a coherent whole, but it goes beyond this: for one, the study of monologue as it occurs in vlogs is supplemented by a similar and comparative investigation of monologue as it occurs in TED Talks. For another, the original study of vlogs and the comparison of its results with those of the study of TED Talks is used to extrapolate from it further options and parameters along which to establish a systematic study of monologue in general. Thus, the structure of this final chapter is made up of three major parts: Conclusion I is a brief review of the findings so far presented; Conclusion II summarizes the comparison of vlog features with the same or similar features in TED Talks; Conclusion III projects a basic theoretical framework for the study of monologue.

9.1 Conclusion I

The present study encompasses six analytic chapters that are mainly based in three approaches established in linguistic theory. Chapters 2 and 3 take up classic topics from Conversation Analytic research, that is, openings and closings. Chapters 4 and 5 cover topics from Interactional Sociolinguistics, that is, repetition and participation frameworks. Chapters 6 and 7 treat topics concerning multimodality, that is, pointing gestures and written interaction in the comments section. Before turning to the comparison of vlogs with another genre, TED Talks, to place the relevance of the research presented so far in the broader area of monologue research, I briefly summarize the main results from these six chapters.

Conversational openings and closings are characterized by their interactive nature as instantiated through fast-paced turn-taking. As monologues are, by definition, characterized by the lack of turn-taking, the beginning of a vlog and its end must be organized according to different parameters. From the basic premise that vloggers

model their monologues at least partially on conversational encounters, chapters 2 and 3 seek to reveal the strategies that speakers use when opening and closing their vlogs, and to discuss their similarity with conversation.

In vlog openings, speakers use verbal and non-verbal strategies. On the one hand, these strategies include linguistic resources such as greetings, terms of address or self-identification, which are borrowed from dialogic genres or conventionalized monologues. On the other hand, they include non-verbal distractive devices, which can trigger side-sequences, and video editing. These phenomena are evidence of audience design, which underlies the production of most monologic media talk. Generally, we can distinguish recurring patterns in a single vlogger's output, and individually produced 'unique' openings.

Regarding monologues, it can be stated that the observation that speakers employ strategies borrowed and adapted from conversation is not surprising. The fixed introductory phrases used in vlog openings are reminiscent of conversational openings from other genres (e.g. TV, radio, answering machine messages, etc.), in that they employ similar elements. Editing allows for strategies that would be very difficult to realize or unconventional in other monologic genres, and multimodal features such as shifts in gaze, facial expression and posture support the vloggers' verbal output.

The lack of turn-taking and consequently the absence of conditional relevance that is a defining feature of adjacency pairs prevent summons-answer or greeting-answer sequences from occurring in vlog openings. Therefore, the joint negotiation of factors such as availability for interaction, identification, social status, and alignment of participants that form the basis of conversational openings does not take place in vlogs. This suggests that openings in vlogs that contain elements similar to those found in conversation have other functions. The inclusion of this type of opening suggests that some vloggers have an imagined audience in mind that they are addressing, which might be helpful in the process of producing a monologue to a camera. It also acts as incentive for the audience to understand the vlog as part of an asymmetric,

asynchronous interaction, inviting the viewers to respond via the communicative channels that YouTube offers.

Like openings, closings are structural elements that have to be 'achieved'. While this achievement entails co-construction and negotiation via turn-taking in (face-to-face) conversation, vloggers are on their own in managing the task they set for themselves. Closings are similar to openings in that many vloggers rely on patterns familiar from conversation or other monologic genres.

Vlog closings contain conventions and phrases from other genres of talk, dialogic and monologic, synchronous and asynchronous, such as conversation, answering machine messages, as well as radio and TV broadcasts. Predominantly, vloggers adapt a conversational closing to monologues: a pre-closing and a terminal exchange are reduced to one or several contributions each instead of one adjacency pair each. In most cases, this is also done with standard lexical items and phrases, which makes closing sections easy to recognize. The lack of negotiation and the resulting shortness of the section are optionally compensated by the insertion of fillers, or further elements that could occupy the closing component slot before the final move. Generally, many vloggers seem to use their experience of answering machine messages as a pattern for their vlog closings.

Those vlog closings that do not adhere to the familiar pattern display a whole spectrum of alternatives that employ alterations of the pattern or leave it out entirely. Editing plays a role in the design of closings. In this second stage of the production of a vlog, crucial decisions are made regarding what remains in the vlog, and where the elements are placed. Editing is used for various effects: speakers can shift footing, change the function of standard elements by assigning them certain slots, shift from spoken to written mode and change the pace of talk. This suggests that vlog closings can be the site of a stage-like performance.

Vloggers commonly adopted terminal components and references to the next encounter as part of their closing section. These are elements typical of conversation as well as all other formats that figure as sources of vlog closing strategies. Retrieving

those familiar elements that occur frequently in interaction in certain slots facilitates online speech production. Likewise, it facilitates comprehension, as these common elements represent expected language use. Thus, the use of familiar elements can be characterized as a form of involvement (in Tannen's 1989 sense).

Parallel to the functions of vlog openings, vlog closings cannot function as the site of a joint negotiation. Instead, they can evoke this joint negotiation: people's conversational experience is an established standard that is sufficiently influential to permeate into other genres. Further, a mutually achieved closing is a highly interactive part of a conversation. If the purpose of a vlog is to connect with the viewers, then the inclusion of this evocation of interactivity can be considered a helpful feature.

Repetition and audience design are phenomena that are closely related to the topic of involvement. While chapters 2 and 3 approached the absence of an audience from a viewpoint that focused on passages that are built around speaker change in conversation, chapters 4 and 5 take an Interactional Sociolinguistic approach that highlights the potential of involvement with an imaginary future audience.

The culture of YouTube values communication between its members (see, e.g., Burgess and Green 2009), with YouTubers frequently asking for feedback of any kind on their videos. Face-to-face communication allows for an immediate exchange, but reactions to a video are necessarily delayed. Consequently, the assignment of participant statuses has to be accomplished differently than in a dialogic, synchronous genre. I have argued that, despite the asynchronicity of the exchange, there is a form of audience involvement present in vlogs resembling that of face-to-face conversation, distinguishing this kind of monologue from the notion of soliloquy. This involvement is based on the assumption that the audience processes the assignment of participant statuses and recognizes the resulting participation structure. Such an argumentation obviously presupposes a cognition-based conceptualization of the notion of involvement, which can be derived from the work of, for example, Chafe (1985) and Tannen (2007 (1989)).

Participant status, or footing, can be traced back to Goffman (1979) and, as an extended model, Bell (1984). These models deconstruct the coarse notions of speaker and hearer into a more differentiated system of participation roles. Vloggers use a variety of devices to assign participant statuses. According to Clark and Carlson (1982), they can be divided into five categories: physical setting, conversational history, gestures, manner of speaking/prosody and content of the talk. These need to be adapted to the online environment of vlogs.

The notion of space, as a primary component of physical setting, emerges as an essential factor in the adaptation of the communicative devices from face-to-face settings to vlogs, as the latter are posted and received in virtual space. This has implications both for the communication that takes place in virtual space and for its origin in separate physical spaces, where the participants are undeniably located. For one thing, people who seek to participate in this communication need access to virtual space, both in terms of internet connection and the skills required to get to this virtual space. For another, once access to virtual space is achieved, the means by which this is carried out, that is, camera plus microphone and computer, determine the physical space along visual and audio parameters due to the limits of the technology used. These circumstances rule out many devices present in face-to-face interaction: As the contextual configuration does not permit visible co-orientation, gestures (e.g. pointing, nodding, posture shifts) and gaze behavior cannot be used by the viewers to negotiate participant status. Physical proximity or distance is not available as a resource to distinguish different vlog viewers.

Multimodal communicative signals are used frequently in vlogs, but to different effects than in face-to-face communication. Changes in volume evoke the re-assignment of participant statuses, such as when whispering excludes those who cannot hear what is being said. The virtual space setting does not allow differentiation in this way, as all viewers are, in principle, equally capable of hearing the vlogger's words as far as the vlogger's influence is concerned. However, the participants are presumably reminded of the effects that whispering has in physical space, so that its use in virtual space may

be a signal to the audience that it is to be understood in terms of its use in an offline environment.

Repetition in spoken language has been shown to be a major device in the creation of involvement (Tannen 2007 (1989)). In vlogs, it can either occur in the form of same-speaker repetition, or in the form of viewer comments cited from earlier videos. While participatory listenership and ratification of listenership, which are based on allo-repetition, cannot occur in vlogs as they depend on synchronous communication, other functions of repetition in conversation can be found in monologues, reflecting in a very basic way that monologic talk has parallels to dialogue. These functions are savoring, evaluation, expanding, stalling, and episode bounding.

The influences of the CMC context are such that communication takes place through an audio-visual channel (video), and as a consequence, editing is possible. The medium video is exploited in combination with repetition, e.g. for a playful or humorous effect. This reflects creativity on the vloggers' part, which is evidence that the vlog format by no means has strict conventions, but fosters experimentation. Editing can play a role even in the first phase of vlog production, as some speakers show explicitly or implicitly an awareness of their future editing options. With this kind of awareness, repetition is assigned a function so far only associated with language acquisition, but not spontaneous adult conversation: repetition can be used to practice talk, especially when it is delivered as part of a performance that is supposed to create a certain effect.

In chapters 6 and 7, the involvement created in vlogs was further investigated from a third perspective which put the focus on elements of multimodal communication. As a mediated form of communication, vlogs not only offer the audio-visual channel (which in itself is a plentiful resource), but the affordances of YouTube surrounding the video screen add various semiotic systems that make vlogs complex sites of interaction. The third piece of analysis is concerned with pointing gestures and written comments by vloggers and viewers.

Four main pointing categories can be identified in the data: (1) (index) finger pointing inside the vlogger's physical space, (2) (index) finger pointing into the two-dimensional virtual space of the website, (3) placing objects into the camera's visual field ('placing for'), (4) camera shifts (as a version of 'placing for'). Pointing gestures usually are small in form due to the spatial restrictions of video footage. As co-orientation of gaze is impossible, placing referents into the visual field of the camera/the viewer is a common practice in vlogs. Often, vloggers employ a combination of 'placing for' and 'directing to' when talking about particular aspects of the referent.

The spatial situation of this genre also brought about the creation of a unique kind of pointing, that is, points that transcend the boundaries between various spaces this communication is situated in. As a result, vloggers point from a three-dimensional space to referents in a two-dimensional space, before those referents even exist. Thus, these gestures create both spatial and temporal deixis.

Vloggers can adopt a viewer's perspective by monitoring their output on screen while shooting the vlog. As a result, some gestures transform three-dimensional space into website, or two-dimensional space: they lack the depth vector. This is compensated for by overlap of gesture and referent on the screen. Although virtual online space of course is not a real, physical space, these kinds of gestures are behavior that indicates people's attempts to inhabit such a space, at least temporarily, as vloggers abandon the meaning these gestures have in their physical surrounding. Such an adaptation to virtual space, which is the space vloggers share with their viewers, is evidence (among many other factors) that vloggers try to interact with their viewers. Possibly, these gestures represent an attempt at social action adapted to the monologue situation.

Actual, traceable interaction between vlogger and viewers on YouTube (rather than offline, for example) necessarily entails the use of several, quite distinct modes: while vloggers initiate contact through the audio-visual channel, viewers and vloggers subsequently interact using written language. The continuation of an exchange across different modes poses problems regarding the achievement of coherence: a comment posted underneath a video must somehow signal what precisely it refers to. The

spatial grammar imposed by our reading practices, that is, reading from top to bottom, is suspended in favor of *reading from video to comment* and in reverse chronological order.

The YouTube comments section design offers functions that facilitate this process of tying to the video and/or other comments, namely the deep link and the reply function. The deep link can be used to signal coherence between a comment and a specific passage in the video, but it is somewhat imprecise as it leaves unclear how long this passage is, and whether the referent of the comment is to be found in the video track or the audio track. The reply function makes visible the coherence between two written comments by re-establishing the spatial order that the English language, for example, follows, that is, from top to bottom, for only those comments that a user marks as part of a particular exchange.

Interactions in YouTube comments sections can extend considerably, not just in the number of 'turns', but also regarding modes. To achieve this, the concept of adjacency needs to be adapted from its original application to two consecutive conversational turns to a more complex model for the description of 'spatial adjacency' in the case of written language, and 'multimodal adjacency' in the case of written language as a reaction to spoken language and vice versa.

9.2 Conclusion II

Conclusion I has given a brief review of the major findings of the six analytic chapters on vlogs. This section draws these results together, mirroring them with their counterparts in TED Talks as developed in chapter 8.

The two main differences in the macro-contextual configuration are that (1) TED Talks are delivered to a physically present audience, where vlogs are delivered into a camera, and (2) speaking and editing of TED Talks are done by different people, where, presumably, vloggers do both. As a consequence, the TED Talk encompasses two sibling genres: one is a live talk on a stage, delivered to an audience, and the other

consists of footage of the first version, which has been edited to fit the online context it is later uploaded to. These differences and the significant consequences regarding the genres that are produced by them are illustrated in Figure 26.

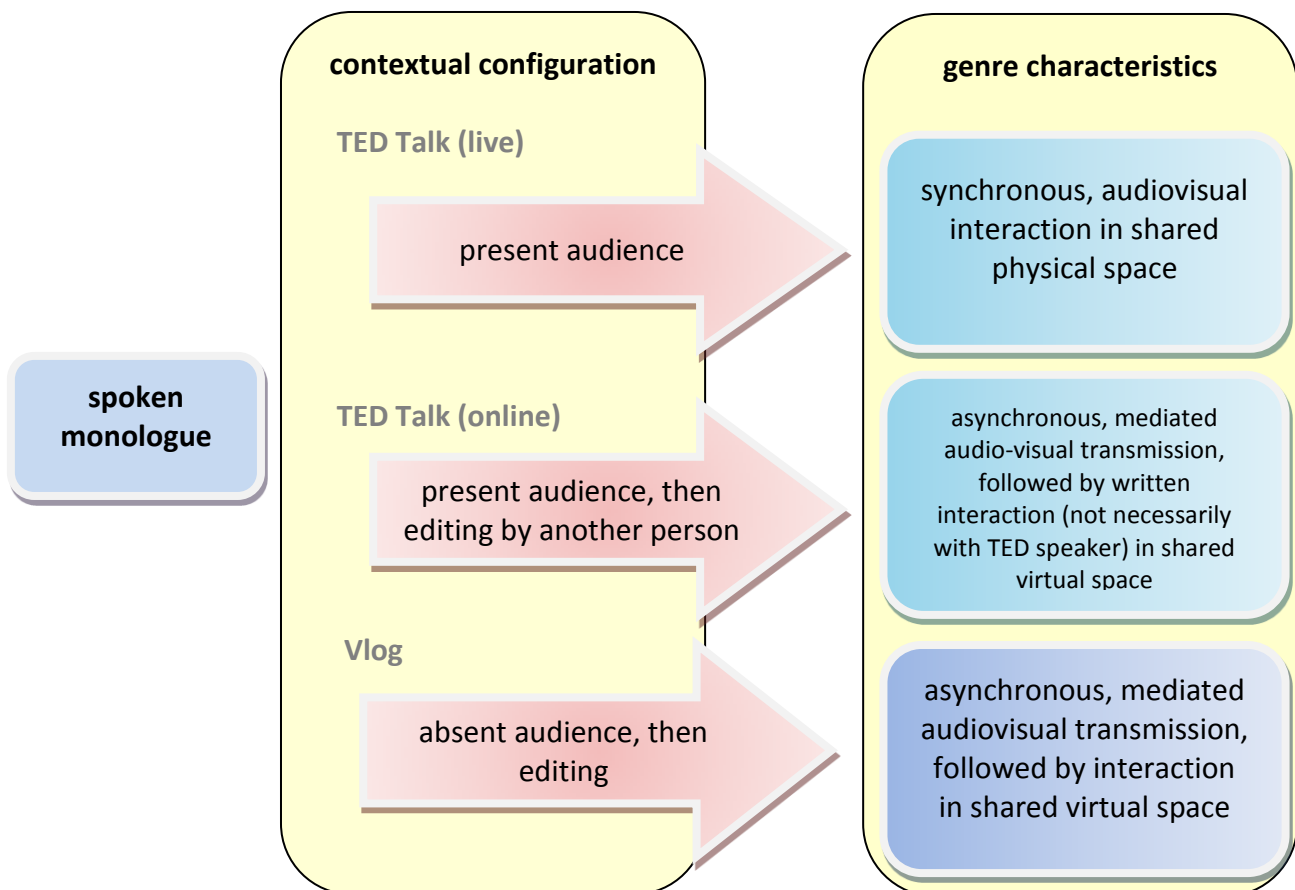| contextual configuration	genre characteristics
TED Talk (live) present audience	synchronous, audiovisual interaction in shared physical space
TED Talk (online) present audience, then editing by another person	asynchronous, mediated audio-visual transmission, followed by written interaction (not necessarily with TED speaker) in shared virtual space
Vlog absent audience, then editing	asynchronous, mediated audiovisual transmission, followed by interaction in shared virtual space

spoken monologue

Figure 26

These general observations shaped the approach with which the comparison of linguistic micro-structures was accomplished: obviously, what is of interest in a pragmatic investigation such as the present one is the influence these contextual factors have on language use in the particular settings. Or in other words, are there variables that remain stable across both contexts? Are there variables that change predictably, that is, according to a pattern? While the comparison in the previous chapter was by no means able to approach laboratory conditions in terms of controlling the settings (as is difficult to do given the many influences on naturally occurring language), it did allow for a number of observations that illuminate, in

principle, the reach of certain genre settings. The comparison was fashioned along the features of talk investigated in chapters 2 through 7, with deviations and selected foci where this promised more interesting results. Table 9.1 summarizes the results of this comparison.

feature	vlog	TED Talk
openings	quasi-conversational pattern with greeting and terms of address; recognizable openings; wide range of openings; some editing	predominantly in medias res opening; all videos have uniform opening credits
closings	recognizable preclosing + closing structure; some variation	predominantly acknowledgement, followed by audience applause
participation framework	multimodal resources (e.g. whispering) and lexical choices determine addressee, by-stander, and referee status of the absent audience members	humor (among other strategies) is employed as a device to trigger reactions from the present audience
involvement through repetition	repetition can be eliminated by editing, or it can be caused by editing, resulting in speech production options unique to vlogging	repetition can be used as a tool for floor management; it is evidence of a pressure to perform without the option of editing
pointing	vloggers adapt their pointing gestures to the surroundings of the website by pointing at	TED speakers stay in the complex, multimodal three-dimensional space of the lecture

	elements outside the video screen, and by inhabiting this two-dimensional virtual space	theater; editorial choices must capture the complexity of the gestures to enable online viewers to make sense of them
written comments	vlogs open a space where vlog viewers can interact with the vlogger and with other viewers	online TED Talks open a space where viewers can interact with other viewers or with those responsible for putting up the video; if a speaker comments, this opens a space to communicate with them directly

Table 9.1

What this shows, among other things, is that the TED Talk, once it is removed from its synchronous version and recontextualized as an asynchronous mediated part of a website, has been changed into a genre other than that of the original talk. Thus we can speak of two TED Talk genres, with the derived, online version being closer in kind to vlogs regarding the medium of communication, and with the original stage talk being closer to vlogs in terms of audience directedness.

9.3 Conclusion III – An agenda for the study of monologue

The two previous sections on video blogs and on the comparison of vlogs and TED Talks represent a fruitful first step in the systematic study of monologue as a spoken form of communication that underlies restrictions imposed by the context it is situated in. However, in light of the existence of many types of monologues and many contextual factors that shape them, it has become increasingly clear throughout the whole study that we still know little about this type of spoken language. This is reflected, as mentioned early on in the introduction to this study, in the relative lack of research on monologues. The present section thus suggests a loose framework along

which a more comprehensive investigation of monologue might be organized in future research endeavors.

Clearly, a large-scale systematic study would benefit from not only large data sets, but also a diverse range of data to enable a comparative approach. A corpus of monologues, therefore, should encompass various types of genres apart from vlogs, such as

- (university) lectures
- (business) presentations
- (political) speeches
- sermons
- debates with assigned speaking slots
- TV/radio presenter talk
- news broadcasts
- voice mail/answering machine messages
- elicited monologues (e.g. in the lab situation)

Regarding the competing definitions of monologue presented in the introduction, it would certainly be desirable to include genres that could challenge the view adopted in the present study (i.e. the definition of monologue as a speech situation where the audience is not expected to interrupt). Although counterintuitive, this implies that further insights might be gained from the inclusion of dialogic genres, for example conversational narrative or joke-telling, where longer turns by single speakers are to be expected. Similarly, the term monologue is used to describe written texts (see, e.g. Biber 1988, Hoffmann 2012), especially academic prose. A better differentiation of the terminology, based on clear criteria, could be achieved by extending the monologue corpus to such written genres. In sum, I propose a data collection as comprehensive as any understanding of the term monologue will allow, with the express purpose of being able to define it more clearly in terms of linguistic or contextual phenomena.

Based on this wide array of data, we can identify parameters that have an impact on language production (and reception) along which to study the strategies applied by language users to adapt to these (and possibly many more) parameters:

- is there an audience
- is there an unknown imaginary audience, a known imaginary audience, a live audience, or all of the above present at the time of speech production
- is the monologue (technologically) mediated
- if so, is it edited
- if so, is the editor the same person as the speaker
- did the speaker self-select to give the monologue, or were they assigned to do so
- is it a planned/scripted monologue, or is it spontaneous
- is there a communicative exchange possible (mediated, audio, audio-visual, etc.)

The comparison presented in the previous section has demonstrated that considerable differences in the speech output are to be expected when there are changes in these contextual variables. I expect that concepts such as audience design and resulting participation frameworks can be illuminated from new angles with the help of an approach that incorporates the variation caused be these factors.

The study of monologue based on a large corpus represents an interesting field for various linguistic disciplines. The present study has adopted a pragmatic approach, or more precisely, a Conversation Analytic, Interactional Sociolinguistic framework, whereby multimodal factors were discussed predominantly from a discourse analytic stance that was informed by close, qualitative analysis. Further studies in this paradigm could investigate topics such as intonation, which had to be left out of the present study, or structural features such as topic transitions or side sequences/insertion sequences.

While I consider this approach quite successful when applied to the investigation of the microstructure of interaction and meaning-making, it is by no means the only way to approach the study of monologue. A tagged and parsed corpus of considerable size would allow for large-scale tendencies of language use based on corpus linguistic methods. This could include the use of grammatical structures in monologues compared to those in dialogic genres. One of the biggest challenges to this approach would be the uniform and reliable transcription of the data, especially regarding multimodal features or editorial practices.

In conclusion, video blogs are a fascinating arena of language use which has given ample inspiration to rethink the notion of interaction. Likewise, this study has revealed that monologues, as a form of human interaction, are shaped by a multitude of contextual factors which have yet to be explored on a large scale for us to make definitive statements about the nature of this type of spoken language.

10 Deutsche Zusammenfassung

Die Pragmatik des Monologs: Interaktion in Video blogs

Die vorliegende Arbeit untersucht mit Mitteln der linguistischen Disziplin Pragmatik die sprachlichen Eigenheiten von Monologen mit Hilfe eines Datenkorpus, welcher aus Video blogs besteht. Die Resultate dieser Untersuchung werden mit ähnlichen Phänomenen aus einem anderen monologischen Kontext verglichen, dem TED Talk. Aus diesem Vergleich wird im Schlusswort ein Ansatz für die systematische Untersuchung von Monologen im Allgemeinen entwickelt, welcher auf mehreren Parametern basiert.

Die Relevanz dieser Arbeit ergibt sich aus mehreren Faktoren. Die untersuchte kommunikative Gattung[21] wird der sogenannten ‚computer-mediated communication' (CMC) zugeordnet, einem ständig wachsenden Bereich kommunikativer Aktivität, die durch das Internet oder Mobilfunknetze mit Hilfe entsprechend ausgestatteter Geräte vermittelt wird. CMC gewinnt zunehmend an Bedeutung als Forschungsfeld für die Linguistik sowie für zahlreiche weitere Disziplinen, da aus ihm Daten gewonnen werden können die Aufschluss über neu entstehende (oder neu verstandene) Konzepte wie zum Beispiel das des ‚virtuellen Raum' liefern. Video blogs sind eine in der Forschung noch stark unterrepräsentierte Form der Kommunikation, so dass die vorliegende Arbeit, als erstes umfassendes Werk zum Thema, Pionierarbeit in der Beschreibung und Analyse der darin vorkommenden Sprache leistet. Insbesondere die der Internetkommunikation eigene Multimodalität, welche hier im weiteren Sinne als ein Konzept verstanden wird, das nicht nur die von Interaktanten ausgehenden audiovisuellen Signale umfasst, sondern ebenso das von den Designvorgaben einer Webseite bestimmte digitale Bild- und Schriftmaterial stellt ein derzeit wichtiges

---

[21] Die germanistische Linguistik unterscheidet zwischen der kommunikativen Gattung und der Kommunikationsform. „Kommunikationsformen bilden den äußeren Rahmen des kommunikativen Geschehens, kommunikative Gattungen sind die in der Kommunikation konstruierten Handlungsmuster, die den Beteiligten eine Orientierung geben" (Dürscheid 2005). Obwohl die genaue Abgrenzung dieser Begriffe nicht vollkommen einheitlich etabliert ist, wird hier der Begriff „kommunikative Gattung" benutzt um das im englischsprachigen Text als „genre" bezeichnete Konzept zu beschreiben.

Forschungsfeld dar, welches durch diese Arbeit weitere Impulse erfährt. Die hier erarbeiteten Ergebnisse leisten wichtige Beiträge zum Verständnis von Organisationmechanismen menschlicher Kommunikation und deren Ausformungen im Internet. Weiterhin wird die Sprechform des Monologs ausführlich betrachtet. Monologe sind linguistisch nur vereinzelt untersucht. Die vorliegende Arbeit stellt die Grundlage für eine systematische Erarbeitung dieses Feldes dar.

Die hier bearbeiteten Fragestellungen und die angewandten Methoden haben sich aus der Beschäftigung mit dem Korpus herausgebildet; es handelt sich bei dieser Untersuchung also um einen grundsätzlich datengetriebenen Ansatz. Dementsprechend ist die Arbeit durch eine Vielzahl von Beispielen und deren Analysen gekennzeichnet. Das Video blog-Korpus besteht aus drei Hauptteilen: einer Sammlung von über 60 video blogs[22] aus den Jahren 2007 bis 2013; eine Sammlung von 46 Video blogs aus dem Frühjahr 2010; den schriftlichen Kommentaren zu 30 Video blogs aus dem zweiten Teil. Darüber hinaus wurde für diese Arbeit ein weiteres, weitaus kleineres Vergleichskorpus angelegt: 20 TED Talks[23] mitsamt ihren auf der Webseite verfügbaren Transkripten lieferten weitere Monologdaten.

Die Analyse der Video blogs erstreckt sich über sechs Kapitel, die in drei Blöcke zu je zwei Themen gebündelt sind. Ihnen zugrunde liegt die Annahme, dass sich Sprecher in Monologsituationen an ihren Erfahrungen in Dialogsituationen orientieren (siehe Haviland 2007, Schegloff 1987). Weiterhin wird hier eine situationsbezogene Definition von Monolog verwendet, die besagt, dass die Unterbrechung des Sprechbeitrages durch eine weitere Person nicht erwartet oder nicht möglich ist (Clark 1996). Der erste Block ist methodologisch an die Konversationsanalyse angelehnt: Kapitel 2 und Kapitel 3 beschäftigen sich demnach mit klassischen Themen aus diesem Bereich, nämlich Anfangs- und Schlusssequenzen. Der zweite Block beschäftigt sich mit Themen aus der Interaktionalen Soziolinguistik: Kapitel 4 und Kapitel 5 wenden die Konzepte des ‚participation framework' und des ‚involvement' auf Video blogs an. Der dritte Block beleuchtet schließlich die Multimodalität inhärent in computergestützter Interaktion,

---

[22] Alle Video blogs, die in dieser Arbeit untersucht wurden, stammen von der Webseite YouTube.
[23] Alle TED Talks wurden der Webseite der TED Organisation (www.ted.com) entnommen.

indem hier Zeigegesten untersucht werden sowie die modalitätsübergreifende Kommunikation in den schriftlichen Kommentaren, die auf der Webseite YouTube stattfinden kann. Ein weiteres Analysekapitel rundet diese drei Blöcke ab; es untersucht die für Video blogs herausgearbeiteten Eigenschaften des gesprochenen Monologs und testet diese auf ihr Bestehen, beziehungsweise ihre Wandelbarkeit, in einer weiteren monologische Kommunikationsgattung, nämlich dem TED Talk. Im Folgenden werden die Ergebnisse dieser sieben Analysekapitel zusammenfassend dargestellt, um darauf aufbauend einen grundlegenden Plan für die weitere systematische sprachwissenschaftliche Erforschung der Sprechform Monolog vorzuschlagen.

Die Konversationsanalyse (CA), wie sie von Harvey Sacks, Emanuel Schegloff und Gail Jefferson begründet und etabliert wurde, betrachtet das ‚turn-taking' (Sprecherwechsel) als fundamentales Organisationsmodell in Konversationssituationen, wobei Sprecher die Präferenz haben, so wenig Pausen und gleichzeitiges Sprechen wie möglich zu produzieren (Sacks et al. 1974). Für die Anfänge von Telefongesprächen gilt, dass diese einem bestimmten Muster folgen. Dieses besteht aus einer Eröffnungssequenz (Schegloff 1968), welche durch das Adjazenzpaar ‚summons-answer' (von Kallmeyer et al. 2001 als „Initiative und Reaktion zum Manifestieren von Gesprächsbereitschaft" umschrieben) geprägt ist. Eine derartige Abfolge setzt zwingend die Gesprächsteilnahme von mindestens zwei Sprechern voraus, da sie einen Sprecherwechsel beinhaltet. Monologische Video blogs weisen laut der hier benutzten Definition diesen Sprecherwechsel nicht auf, so dass eine Anfangssequenz alternativ gestaltet werden muss. Viele Vlogger greifen dabei auf Strategien zurück, die aus der Konversation bzw. anderen monologischen kommunikativen Gattungen entliehen und angepasst werden. Es besteht zum Beispiel eine starke Tendenz bei vielen Vloggern, ihre Zuschauer mit Begrüßung und Anredeformel anzusprechen. Da die in einer Konversation erreichten Funktionen einer derartigen Anfangssequenz, also z.B. das Testen der Bereitschaft für einen Austausch oder die Feststellung der Identität des Gegenübers, in einem Vlog nicht stattfinden können, stellt sich die Frage nach der Funktion interaktiv geprägter Elemente wie der

Begrüßung. Da es sich bei Vlogs um eine kommunikative Gattung handelt, durch die Vlogger laut eigener Aussage den Kontakt und Austausch mit anderen Menschen suchen, liegt die Vermutung nahe, dass eine konversationsähnliche Anfangssequenz diesen Austausch suggerieren soll. Zudem handelt es sich bei diesen Sequenzen um sehr routinierte Abläufe, die dementsprechend die Sprachproduktion erleichtern, so dass sich Vlogger dadurch einen erleichterten Einstieg in den Monolog verschaffen können. Dass es sich dabei aber keinesfalls um eine durchwegs etablierte Konvention der kommunikativen Gattung handelt, zeigt die Tatsache, dass eine weitere starke Tendenz darin besteht, den Vlog vollkommen ohne Gruß und Anredeformel zu gestalten. Bei der Betrachtung mehrerer Vlogs von jeweils einer Person zeigt sich, dass viele Vlogger für sich eine erkennbare Routine entwickeln, die sie am Anfang eines Vlogs benutzen.

Ebenso wie die Anfänge sind auch die Abschlüsse Elemente, die in der Konversation gemeinsam erarbeitet werden. In Vlogs liegt diese Aufgabe allein in der Hand des Vloggers, welcher sich nicht auf die durch den Sprecherwechsel geprägte Bildung von Adjazenzpaaren berufen kann. Wie Schegloff und Sacks (1973) feststellen, werden die Abschlüsse von Telefongesprächen nach einem bestimmten Muster gebildet, welches aus der Abfolge von zwei Adjazenzpaaren besteht: ‚pre-closing – pre-closing' (Abschlußmarkierung) gefolgt von ‚closing – closing' (Abschiedsformel). Unter Vloggern besteht eine starke Tendenz, dieses Muster angepasst zu übernehmen. Die Abfolge pre-closing → closing bleibt bestehen, jedoch sind die Adjazenzpaare auf Einzelredebeiträge reduziert. Dieses Muster lässt sich vielfach in Anrufbeantworternachrichten beobachten, was zu der Annahme führt, dass diese als Vorbild für Vlogger dienen. Gleichermaßen wie in den Vloganfängen wird auch in Vlogabschlüssen nicht die Bereitschaft eines Interaktanten getestet, die Interaktion zu beenden, sondern die Funktion einer solchen Schlussabfolge im Vlog muss anderweitig gesucht werden. Es liegt nahe, dass der interaktive Charakter, wie auch schon im Falle des Vloganfangs, den konversationsähnlichen Austausch suggeriert und die Sprachproduktion erleichtert.

Alternativ zu diesem an Konversationen orientierten Muster gibt es wiederum eine Reihe anderer Realisationen des Schlussteils, wobei das Bearbeiten des Filmmaterials oftmals eine tragende Rolle spielt. So können Vlogs enden, ohne dass der Sprecher eine als Abschluss erkennbare Phase durchlaufen hat, also mit einem abrupten Schnitt. Ebenso gibt es abspannartige Schlussteile wie in Fernsehfilmen und -serien, die den Wechsel von gesprochener Sprache zu geschriebener Sprache ermöglichen. Hierdurch wird klar, dass Vlogger ihre Videos als Teil eines Mediums betrachten, dass durchaus den Konventionen anderer Medien folgen kann.

Wiederholungen und Zuschauerdesign sind sprachliche Phänomene, die eng mit dem Konzept des ‚involvement' (Involviertheit) verwoben sind. Während Kapitel 2 und 3 sich mit dem Fehlen eines Publikums anhand von Passagen beschäftigt haben, die in Konversationen durch den Sprecherwechsel maßgeblich beeinflusst sind, beleuchten Kapitel 4 und 5 mittels Konzepten der Interaktionalen Soziolinguistik die Zuschauereinbindung in Vlogs.

Die Nutzerkultur der Webseite YouTube ermutigt den Austausch zwischen den Nutzern und bewertet diesen positiv (siehe Burgess und Green 2009), so dass YouTuber häufig in ihren Videos um Antworten und Reaktionen jeglicher Art bitten. Im alltäglichen ‚face-to-face'-Gespräch (im direkten Gegenüber) kann ein Austausch unmittelbar erfolgen; doch das Medium Video erlaubt dies nicht, sondern lässt ausschließlich verzögerte Reaktionen zu. Daher muss das Aushandeln bzw. Zuweisen der jeweiligen Teilnehmerrolle im Vlog anders erfolgen als im synchronen, dialogartigen Austausch. Die Involviertheit der Zuschauer eines Vlogs ähnelt dennoch der in der Konversation. Diese Argumentation basiert auf einem kognitiven Verständnis von Involviertheit, wie es zum Beispiel von Chafe (1985) und Tannen (2007 (1989)) abgeleitet werden kann. Hierbei wird vorausgesetzt, dass die Zuschauer das Zuweisen einer Teilnehmerrolle aufnehmen und daraus die resultierende Teilnehmerrollenstruktur erkennen.

Das Konzept der Teilnehmerrolle geht auf Goffman (1979), und in erweiterter Form auf Bell (1984) zurück. Diese Modelle dekonstruieren die grobe Einteilung in Sprecher und Zuhörer, und teilen sie in ein differenzierteres Rollenmodell auf. Vlogger benutzen eine

Reihe von Mitteln, um solche Rollen zuzuteilen. Clark und Carlson (1982) schlagen hier fünf Kategorien vor: Räumlicher Kontext, Konversationsgeschichte, Gestik, Sprechart/Prosodie und Gesprächsinhalt. Diese Kategorien müssen an den online-Kontext von Vlogs angepasst werden.

Das Konzept Raum, als hervortretende Komponente des räumlichen Kontexts, stellt einen wesentlichen Faktor hinsichtlich der Anpassung der kommunikativen Mittel an die Vlog-Umgebung dar, da diese in den virtuellen Raum hochgeladen, und auch dort rezipiert werden. Dies hat Implikationen für die Kommunikation, die in diesem virtuellen Raum stattfindet, sowie für ihre ursprüngliche Form in separaten physischen Räumen, in denen die Teilnehmer sich befinden. Grundsätzlich müssen diejenigen, die an der Kommunikation mit Vlogs interessiert sind, Zugriff auf den virtuellen Raum haben, sowohl hinsichtlich solcher Gegebenheiten wie einem internetfähigen Gerät und einer Verbindung zum Internet, als auch hinsichtlich der Fähigkeiten, solch einen Raum aufzusuchen. Weiterhin bestimmen die technischen Eigenschaften der verwendeten Übertragungsmittel, also Kamera, Mikrophon und Computer, die Dimensionen des physischen Raums entlang audiovisueller Parameter. Diese Umstände verändern die Funktion einiger Mittel der face-to-face-Kommunikation im Vlog: Da die kontextuelle Konfiguration die sichtbare gemeinsame Orientierung unmöglich macht, können Gesten, Nicken, Haltungsänderungen und Blickverhalten nicht zum gemeinsamen Verhandeln der Teilnehmerrolle benutzt werden. Physische Nähe oder Distanz können ebenso nicht als Ressource zum Unterscheiden mehrerer Zuschauer(-gruppen) benutzt werden.

Multimodale Signale werden dennoch häufig in Vlogs eingesetzt. Zum Beispiel evozieren Änderungen der Sprechlautstärke die Neuzuweisung einer Teilnehmerrolle, so wie in einem face-to-face-Gespräch Flüstern diejenigen ausschließt, die nicht mehr in Hörweite sind. Im virtuellen Raum kann diese Differenzierung nicht erfolgen, da prinzipiell, was den Einfluss des Vloggers betrifft, alle Zuschauer den Vlogger gleichermaßen hören können. Die Zuschauer sind jedoch aufgefordert, ein solches

Signal im Vlog so zu verstehen, als wäre es Teil einer Konversation. Dieses Verständnis auf Seiten der Zuschauer macht ihre Involviertheit aus.

Wiederholungen in gesprochener Sprache wurden als wichtiges Mittel identifiziert, mit dem sich Involviertheit erzeugen lässt (Tannen 2007 (1989)). In Vlogs kann diese in Form von Selbstwiederholung auftreten, oder in Form von Zitaten aus (meist geschriebenen) Kommentaren zu früheren Videos. Während ‚participatory listenership‘ (teilnehmendes Zuhören) und das Ratifizieren von Gesprächsbeiträgen, welche auf der ‚allo-repetition‘ (Wiederholung des anderen) beruhen, im Vlog nicht vorkommen können, lassen sich andere Funktionen der Wiederholung im Monolog durchaus finden. Hierin lässt sich erkennen, dass die Sprechform Monolog grundlegende Parallelen zum Dialog hat. Die Funktionen sind das Genießen, die Bewertung, das Erweitern, das Verzögern und das Rahmen von Gesprächsepisoden.

Die Einflüsse des CMC-Kontexts sind hier so beschaffen, dass Kommunikation über einen audiovisuellen Kanal läuft, und dementsprechend besteht die Möglichkeit zur Bild- und Tonbearbeitung. Diese werden im Zusammenhang mit Wiederholung zu spielerischen und humorbezogenen Zwecken genutzt, welches die Kreativität von Vloggern belegt, sowie die Freiheiten des Formats bezüglich der Konventionen der kommunikativen Gattung. Bild- und Tonbearbeitung haben ebenso Auswirkungen auf die erste Phase der Vlogproduktion, dem tatsächlichen Sprechen: hier zeigt sich – explizit wie implizit, dass Vlogger ein Bewusstsein für ihre späteren Bearbeitungsoptionen haben. Durch dieses Bewusstsein kommt der Wiederholung eine Funktion zu, die bisher hauptsächlich mit Spracherwerb assoziiert wird, aber nicht mit spontaner Konversation: Wiederholung kann für einen Übungsdurchgang genutzt werden, der später ausgeschnitten werden soll, um als Endprodukt eine überzeugende Vorführung zu haben.

Die in Vlogs erzeugte Involviertheit wurde abschließend aus einer dritten Perspektive untersucht, welche sich hauptsächlich auf Elemente multimodaler Kommunikation konzentriert. Als kommunikative Gattung, die durch Medien vermittelt wird, basieren Vlogs nicht nur auf dem audiovisuellen Kanal (welche an sich eine reichhaltige

Ressource darstellt), vielmehr werden durch die Einbettung in eine Webseite mit spezifischen Gegebenheiten dem Video mehrere semiotische Systeme hinzugefügt, die den Vlog zum Schauplatz komplexer Interaktion machen. Der dritte Teil der Vlog-Analyse befasst sich mit Zeigegesten und geschriebenen Kommentaren von Vloggern und ihren Zuschauern.

In den Daten lassen sich vier Hauptkategorien von Zeigegesten identifizieren: (1) Zeigefinger/Daumen-Zeigegesten innerhalb des physischen Raums des Vloggers; (2) Zeigegesten in den virtuellen Raum der Webseite; (3) Plazieren von Objekten in das Sichtfeld der Kamera ('placing for'); (4) Kamerabewegungen. Zeigegesten sind generell klein in ihrer Ausführung aufgrund der räumlichen Restriktionen der Bildaufnahmen. Da die gemeinsame, koordinierte Ausrichtung des Blicks nicht möglich ist, werden in Vlogs häufig Objekte vor die Kamera, und damit in das Sichtfeld des Zuschauers, plaziert. Dies wird häufig mit einer zusätzlichen Zeigegeste kombiniert.

Die räumliche Besonderheit des Vlogs hat die Entwicklung einzigartiger Zeigegesten befördert, nämlich Gesten, die die Grenzen der Räume, in denen Vlogs produziert und gesehen werden, überschreiten. Vlogger benutzen Gesten, um von ihrem dreidimensionalen Raum aus auf die noch nicht existente Umgebung der Webseite zu zeigen, in der das Video später zu sehen sein wird. Somit erzeugen sie räumliche wie temporale Deixis.

Vlogger können die Perspektive der Zuschauer annehmen, indem sie ihr Videomaterial bereits beim Filmen auf einem Bildschirm mitverfolgen. Daraus folgt, dass manche Gesten den dreidimensionalen Raum in zweidimensionalen Webseitenraum umwandeln: es fehlt die Tiefendimension. Dies wird kompensiert durch das Überlappen von Geste und Objekt auf dem Bildschirm. Für den Moment, in dem diese Gesten ihre Bedeutung im dreidimensionalen Raum verlieren, zeugen sie von dem Versuch, den zweidimensionalen Raum zu 'bewohnen'. Dieser Versuch, den virtuellen Raum der Webseite, den Vlogger mit ihren Zuschauern teilen, einzunehmen, deutet auf die starke Motivation hin, mit ihnen in Kontakt zu treten, da er als 'social action' (soziale Handlung) betrachtet werden kann.

Tatsächlicher, nachvollziehbarer Austausch zwischen Vloggern und Zuschauern auf YouTube erfordert die Nutzung mehrerer, recht unterschiedlicher Modalitäten. Vlogger initiieren den Austausch über den audiovisuellen Kanal, worauf anschließend Kommentare und Antworten in geschriebener Sprache erfolgen. Die Fortführung eines Austausches über mehrere Modalitäten hinweg verursacht Probleme hinsichtlich der Herstellung von Kohärenz: Ein Kommentar, der unter ein Video gesetzt wird, muss in irgendeiner Form signalisieren, worauf genau er sich bezieht. Die räumliche Grammatik, die durch unsere Schreib- und Lesegewohnheiten vorgegeben ist, also von links nach rechts und von oben nach unten, wird hier aufgehoben für die *Video zu Kommentar Lesart*, und die chronologisch inverse Anordnung.

Die Kommentarsektion auf YouTube enthält Funktionen, die die Erzeugung von Kohärenz durch ‚tying‘ (Anbinden) unterstützen, nämlich der ‚deep link‘ und die Antwortfunktion. Der deep link kann benutzt werden, um die Anbindung an eine bestimmte Stelle im Video zu signalisieren, jedoch ist diese Wirkung unpräzise, da unklar bleibt, wie lange die Stelle ist, und ob sich der Referent des Kommentars im Bild- oder Tonmaterial der Passage findet. Die Antwortfunktion macht die Kohärenz zwischen zwei schriftlichen Beiträgen sichtbar, indem sie die gewohnte Raumgrammatik wiederherstellt, also einen chronologisch vorausgegangenen Kommentar direkt über seine Antwort plaziert.

Interaktionen auf YouTube können beträchtlich expandieren, nicht nur hinsichtlich der Anzahl ihrer Beiträge, sondern auch bezüglich der angewandten Modalitäten. Um diesen Austauschsituationen analytisch gerecht zu werden, muss das zentrale Konzept der Adjazenz angepasst und erweitert werden, um nicht mehr nur auf zwei aufeinanderfolgende Redebeitrage in einer Konversation anwendbar zu sein. Für geschriebene Sprache müsste das Konzept um die Komponente der räumlichen Adjazenz erweitert werden, und für Interaktionen über mehrere Modalitäten hinweg um die Komponente der multimodalen Adjazenz.

Der Vergleich mit einer weiteren monologischen Gattung, dem TED Talk, entlang der Phänomene und Konzepte, mittels derer Vlogs beleuchtet wurden, zeigt eine Reihe

von Unterschieden und Variationen auf. Diese sind auf zwei Hauptunterschiede in den kontextuellen Gegebenheiten der beiden Gattungen zurückzuführen, nämlich die Gegenwart eines Publikums und das Aufteilen der Produktionsprozesse des Sprechens und der Bild- und Tonbearbeitung auf mehrere Personen. Daher muss der TED Talk auch in zwei Gattungen unterteilt werden, zum einen die live-Version, die ein Sprecher auf einer Bühne vorträgt, und zum anderen die bearbeitete Videoversion, die online gestellt wird.

Für die Anfangssequenzen von TED Talks bedeutet dies, dass alle denselben Vorspann haben, unabhängig von Sprecher und Thema. Im Gegensatz zu Vlogs werden die meisten TED Talks ohne Begrüßung und Anredeformel begonnen, sondern ‚in medias res'. Die TED Schlusssequenz ist insofern unterschiedlich zu Vlogs, als sie den Übergang von Monolog zu einer bestimmten zu erwartenden Zuschauerreaktion markiert, nämlich Applaus. Dieses wird fast durchgehend durch eine relativ unvermittelt eingesetzte Dankesformel erreicht, die für das Publikum als Zeichen für die Übergabe des Beitragsrechts dient.

Im Vergleich zu Vlogs bieten TED Talks die Möglichkeit, direkte Publikumsreaktionen als Daten mit auszuwerten, deshalb beleuchtet die Analyse hier die Involviertheit der Zuschauer anhand der tatsächlich im Raum stattfindenden Interaktion. Insbesondere ist der Einsatz von Humor ein wichtiges und häufig gebrauchtes Mittel zur Steuerung bestimmter Reaktionen. Die Analyse arbeitet heraus, dass, obgleich im Monolog die Sprecherrolle ausschließlich beim Vortragenden liegt, Beiträge des Publikums in Form von Gelächter grundsätzlich erwünscht und durch Feinabstimmung als Element der Mikroorganisation der Interaktion mit eingearbeitet werden. In ähnlich organisierender Funktion kann auch die Wiederholung eingesetzt werden, nämlich um eine Situation, in der die Teilnehmerrollen ausgehandelt werden (wie in Passagen, in denen das Publikum durch Gelächter teilnimmt), zu strukturieren. Hier kann eine kippende Teilnehmerstruktur durch die Wiederholung einer vorangegangenen Äußerung wieder in der Ausgangslage gefestigt werden, indem der Vortragende die Sprecherrolle durch die damit verbundene Autorität bestätigt. Weiterhin können

Wiederholungen als Stilmittel auf die intensive Vorbereitung hinweisen. Diese ist, im Vergleich zum Vlog, der Tatsache geschuldet, dass der Vortrag vor einem live-Publikum stattfindet, und die Ton- und Bildbearbeitung also wenn überhaupt nur in der online Version das Ergebnis verändern könnte.

Der Einsatz von Gesten in TED Talks weist nicht die Besonderheiten auf, die in Vlogs festgestellt werden können. Da TED Sprecher sich in der Hauptsache an das anwesende Publikum richten, nicht aber an die spätere online Zuschauerschaft, bleiben die Gesten im dreidimensionalen Raum des Vortragssaales. Dennoch bieten die multimodalen Ressourcen, zum Beispiel eine Leinwand im Hintergrund der Bühne, Spielraum für komplexe Gesten, deren Interpretation durch das Publikum einigen Aufwand bedeutet. Die Herausforderung für den zweiten Schritt der Produktion, also die Bild- und Tonbearbeitung des Videomaterials, besteht darin, diese komplexen audiovisuellen Kommunikationssignale derart darzustellen, dass ein Konsument des online Videos ebensogut in der Lage ist, die Bedeutung der Signale zu interpretieren, wie ein Zuschauer im Raum, der durch die Änderung der Blickrichtung beispielsweise selbst entscheiden kann, welchen Signalen er Aufmerksamkeit schenken will, um das Geschehen als kohärente Interaktion interpretieren zu können. Die geschriebene Interaktion, die in der online Version des TED Talks stattfindet, unterscheidet sich von der in Vlogs durch die ‚Besitzerdynamik' des virtuellen Raums, in dem sie stattfindet. Das Vlog-Video eröffnet einen derartigen Raum mit dem Beitrag des Vloggers, so dass dieser automatisch Teilnehmer ist. Der TED Talk wird nicht direkt vom Sprecher selbst hochgeladen, so dass dieser nicht zwingend Teilnehmer in der Interaktion des virtuellen Raums ist. Sobald ein Sprecher aber selbst einen geschriebenen Kommentar hinterlässt, kann dieser durch die Antwortfunktion, welche nachfolgende, als Antwort eingestufte Kommentare nach rechts einrückt, benutzt werden, um einen dem Sprecher gehörenden Raum zu etablieren. Dieser Raum ähnelt dann der gesamten Kommentarsektion eines Vlogs, die die Teilnahme des Vloggers per se voraussetzt.

Die Arbeit schließt mit einem Ausblick auf die Untersuchung von Monologen im Allgemeinen. Diese könnte durch ein großangelegtes Korpus systematisch und

umfassend erfolgen, indem monologische Gattungen, dialogische Gattungen, sowie geschriebene Sprache anhand mehrerer Parameter untersucht werden. Dies würde erheblich zu unserem Verständnis des kontextuellen Einflusses auf Sprachproduktion im Monolog, sowie zu einer Klärung und eindeutigeren Benutzung der einschlägigen Terminologie beitragen.

11 Appendix

Transcription conventions

she's out.	Period shows falling tone in the preceding element, suggesting finality.
oh yeah?	Question mark shows rising tone in the preceding element; cf. yes-no question intonation.
so, now,	Comma indicates a level, continuing intonation; suggesting nonfinality.
bu- but	A single dash indicates a cutoff (often with a glottal stop); including truncated intonation units.
DAMN	Capitals show heavy stress or indicate that speech is louder than surrounding discourse.
°dearest°	Utterances spoken more softly than the surrounding discourse are framed by degree signs.
says "oh"	Double quotes mark speech set off by a shift in the speaker's voice.
(2.0)	Numbers in parentheses indicate timed pauses.

If the duration of the pauses is not crucial and not timed:

..	a truncated ellipsis is used to indicate pauses of one-half second or less.
...	An ellipsis is used to indicate a pause of more than a halfsecond.
ha:rd	The colon indicates the prolonging of the prior sound or syllable.
<no way>	Angle brackets pointing outward denote words or phrases that are spoken more slowly than the surrounding discourse.
>watch out<	Angle brackets pointing inward indicate words or phrases spoken more quickly than surrounding discourse.
[and so-]	
[WHY] her?	Square brackets on successive lines mark beginning and end of

	overlapping talk; multiple overlap is marked by aligning the brackets.
and=	
=then	Equal signs on successive lines show latching between turns of different speakers; they can also indicate that the turn of one speaker continues after e.g. backchannels of interlocutors.
H	Clearly audible breath sounds are indicated with a capital H.
.h	Inhalations are denoted with a period, followed by a small h. Longer inhalations are depicted with multiple hs as in .hhhh
h	Exhalations are denoted with a small h (without a preceding period).
	A longer exhalation is denoted by multiple hs.
.t	Alveolar suction click
( )	In the case that utterances cannot be transcribed with certainty empty parentheses are employed.
(hard work)	If there is a likely interpretation, the questionable words appear within the parentheses.
/ /	Slashes are used for phonetic transcriptions.
((laugh))	Aspects of the utterance, such as whispers, coughing, and laughter, are indicated with double parentheses.
{points at board}	Nonverbal behavior, such as movements and looks, are indicated with braces.
CUT, BLEND	The editorial features cutting and blending are indicated in capital letters.
Numbering:	Number each intonation unit is numbered consecutively (e.g. from 1 to n).

12 References

References

Aijmer, Karin. 2007. Idiomaticity in a cultural and activity type perspective: Conventionalisation of routine phrases. *Topics in English linguistics, vol. 54: Phraseology and culture in English*, ed. by Paul Skandera, 323-349. New York: de Gruyter.

Alajoutsijärvi, Seppo; and Kaarina Määttä. 2011. What was said?: A discourse analysis of a famous Finnish radio journalist's virtuous monologues. *Review of European Studies* 3 (2). 22-32.

Ariel, Mira. 2010. *Defining pragmatics*. Cambridge: Cambridge University Press.

Askehave, Inger; and Anne Ellerup Nielsen. 2005. Digital genres: a challenge to traditional genre theory. *Information Technology & People* 18 (2). 120 – 141.

Bakhtin, Mikhail. 1984. *Problems of Dostoevsky's poetics*, trans. by Caryl Emerson. Minneapolis: University of Minnesota Press.

Bangerter, Adrian; and Daniel M. Oppenheimer. 2006. Accuracy in detecting referents of pointing gestures unaccompanied by language. *Gesture* 6 (1). 85-102.

Baron, Naomi S. et al. 2005. Tethered or mobile?: Use of away messages in instant messaging by American college students. *Mobile communications: Re-negotiation of the social sphere*, ed. by Rich Ling; and Per E. Pederson, 293-311. London: Springer.

Bauman, Richard. 2006. Speech Genres in Cultural Practice. *Encyclopedia of Language & Linguistics*, vol. 11, ed. by Keith Brown, 745 - 758. Oxford: Elsevier.

Bavelas, Janet B.; Chovil, Nicole; Lavrie, Douglas A.; and Allan Wade. 1992. Interactive gestures. *Discourse Processes* 15. 469-489.

----; Chovil, Nicole; Coates, Linda; and Lori Roe. 1995. Gestures for specialized dialogue. *Personality and Social Psychology Bulletin* 21. 394-405.

----; Kenwood, Christine; Johnson, Trudy; and Bruce Phillips. 2002. An experimental study of when and how speakers use gestures to communicate. *Gesture* 2 (1). 1-17.

----; Gerwing Jennifer; Sutton, Chantelle; and Danielle Prevost. 2008. Gesturing on the
telephone: independent effects of dialogue and visibility. *Journal of Memory
and Language* 58, 495–520.

Beißwenger, Michael. 2008. Multimodale Analyse von Chat-Kommunikation. *Arbeit mit
Transkripten in der Praxis: Forschung, Lehre und Fortbildung*, ed. by Karin
Birkner; and Anja Stukenbrock. Mannheim: Verlag für Gesprächsforschung.

Besnier, Niko. 1994. Involvement in linguistic practice: An ethnographic appraisal.
Journal of Pragmatics 22. 279-299.

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge
University Press.

Biber, Douglas; Johansson, Stig; Leech, Geoffrey; Conrad, Susan; and Edward Finegan.
1999. *Longman grammar of spoken and written English*. Essex: Pearson.

Boas, Franz. 1940. *Race, Language, and Culture*. New York: Free Press.

Boyd, Michael S. 2010. Participation, democratization and discrimination in new
media: Political discourse and YouTube. *T2PP Workshop, 9-10 April 2010, Vrije
Universiteit Amsterdam.*
http://www2.let.vu.nl/oz/cltl/t2pp/docs/ws2010/papers/P7-Boyd.pdf.
(22.10.2012)

Bublitz, Wolfram and Christian Hoffmann. 2011. 'Three men using our toilet all day
without flushing – This may be one of the worst sentences I've ever read':
Quoting in CMC. *Anglistentag Saarbrücken 2010 Proceedings*, ed. by Joachim
Frenk and Lena Steveker, 433-448. Trier: WVT.

Burgess, Jean; and Joshua Green. 2009. *YouTube: Online video and participatory
culture*. Cambridge: Polity Press.

Cayari, Christopher. 2011. The YouTube effect: How YouTube has provided new ways
to consume, create, and share music. *International Journal of Education & the
Arts* 12 (6). 1-29.

Cha, Meeyoung et al. 2007. I tube, you tube, everybody tubes: Analyzing the world's
largest user generated content video system. *Proceedings of the 7th Internet*

Measurement Conference 2007, ed. by Constantine Dovrolis; and Matthew Roughan, 1-14. New York: ACM.

Chafe, Wallace. 1982. Integration and involvement in speaking, writing, and oral literature. *Advances in discourse processes, vol 9: Spoken and written language: Exploring orality and literacy*, ed. by Deborah Tannen, 35-54. Norwood, NJ: Ablex.

----. 1985. Linguistic differences produced by differences between speaking and writing. *Literacy, language, and learning: The nature of consequences of reading and writing*, ed. by David R. Olson; Nancy Torrance; and Angela Hildyard, 105-123. Cambridge: Cambridge University Press.

----. 1994. *Discourse, consciousness and time*. Chicago: University of Chicago Press.

Chau, Clement. 2010. YouTube as a participatory culture. *New Directions for Youth Development* 128. 65-74.

Cheng, Xu; Cameron Dale; and Jiangchuan Liu. 2007. Understanding the characteristics of internet short video sharing: YouTube as a case study. Technical report, Cornell University. *arXiv:0707.3670.* (25.10.2012)

Clark, Herbert H. 1996. *Using language*. Cambridge: Cambridge University Press.

----; 2003. Pointing and Placing. Cf. *Pointing: Where language, culture, and cognition meet,* ed. by Sotaro Kita, 243-268. Mahwah: Erlbaum.

----; and Catherine R. Marshall. 1981. Definite reference and mutual knowledge. *Elements of discourse understanding,* ed. by Aravind K. Joshi; Bonnie L. Webber; and Ivan A. Sag, 10-63. Cambridge: Cambridge University Press.

----; and Thomas B. Carlson. 1982. Hearers and speech acts. *Language* 58 (2). 332-373.

----; and Thomas B. Carlson. 1992. Hearers and speech acts. *Arenas of Language Use,* ed. by Herbert Clark, 205-247. Chicago: University of Chicago Press.

Cook-Gumperz, Jenny; and John Gumperz. 1994. The politics of conversation: Conversational inference in discussion. *What's going on here?: Complementary studies*, ed. by Allen D. Grimshaw, 373-395. Cambridge: Cambridge University Press.

Coulthard, Malcolm; and Martin Montgomery. 1981. The structure of monologue. *Studies in discourse analysis,* ed. by Malcolm Coulthard; and Martin Montgomery, 30-39. London: Routledge.

Couper-Kuhlen, Elisabeth. 1998. Coherent voicing: On prosody in conversational reported speech. *InLiSt: Interaction and Linguistic Structures* 1. [*http://www.inlist.uni-bayreuth.de/issues/1/index.htm*] (25.10.2012)

----. 1999. Coherent voicing: On prosody in conversational reported speech. *Coherence in spoken and written discourse: How to create it and how to describe it*, ed. by Wolfram Bublitz; and Uta Lenk, 11-32. Amsterdam: John Benjamins.

----; and Margaret Selting. 1996. Towards an interactional perspective on prosody and a prosodic perspective on interaction. *Studies in interactional sociolinguistics, vol 12: Prosody in conversation: Interactional studies*, ed. by Elizabeth Couper-Kuhlen; and Margaret Selting, 11-56. Cambridge: Cambridge University Press.

Crabb, Peter. 1996. Answering machines take the "answering" out of telephone interactions. *Journal of Social Behavior and Personality* 11 (2). 387-397.

Crawford Camiciottoli, Belinda. 2008. Interaction in academic lectures vs. written text materials: The case of questions. *Journal of Pragmatics* 40. 1216-1231.

Crystal, David. 1985. Things to remember when transcribing speech. *Child Language Teaching and Therapy* 1 (2). 235-239.

Dressler, Richard A.; and Roger J. Kreuz. 2000. Transcribing oral discourse: A survey and a model system. *Discourse Processes* 29 (1). 25-36.

Dubin, Fraida. 1987. Answering machines. *English Today* 10. 28-30.

Duman, Steve; and Miriam Locher. 2008. 'So let's talk. Let's chat. Let's start a dialog:' An analysis of the conversation metaphor employed in Clinton's and Obama's YouTube campaign clips. *Multilingua* 27. 193-230.

Duranti, Alessandro. 1992. Language in context and language as context: The Samoan respect vocabulary. Cf. *Rethinking context: Language as an interactive phenomenon*, ed. by Alessandro Duranti; and Charles Goodwin, 77-99. Cambridge: Cambridge University Press.

----; and Charles Goodwin (eds.). 1992. *Rethinking context: Language as an interactive phenomenon*. Cambridge: Cambridge University Press.

Durkheim, Émile. 1894. Les règles de la méthode sociologique. *Revue Philosophique de la France et de l'Étranger* 37. 465-498.

Dürscheid, Christa. 2005. *Medien, Kommunikationsformen, kommunikative Gattungen*. Linguistik online 22/1: 3–16. http://www.linguistik-online.de/22_05/duerscheid.pdf (19.05.2013)

Efron, David. 1941 (1972). *Gesture, race, and culture*. The Hague: Mouton.

Enfield, Nick J.; Sotaro Kita; and Jan Peter De Ruiter. 2007. Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics* 39 (10). 1722-1741.

Ess, Charles; and the AoIR ethics working committee. 2002. *Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee*, approved by AoIR, November 27, 2002. *www.aoir.org/reports/ethics.pdf*. (28.02.2013)

Fallis, Don. 2008. Towards an epistemology of Wikipedia. *Journal of the American Society for Information Science and Technology* 59 (10). 1662-1674.

Feldmann, Valerie; and Axel Zerdick. 2004. E-merging media: The future of communication. *E-merging media: Communication and the media economy of the future*, ed. by Axel Zerdick et al., 19-32. Berlin: Springer.

Fetzer, Anita; and Etsuko Oishi. 2011. *Context and contexts*. Amsterdam: John Benjamins.

Fortanet, Immaculada. 2004. The use of 'we' in university lectures: Reference and function. *English for Specific Purposes* 23. 45-66.

Frank, Jane. 1989. On conversational involvement by mail: The use of questions in direct sales letters. *Text* 9 (2). 231-259.

Frobenius, Maximiliane. 2011. Beginning a monologue: The opening sequence of video blogs. *Journal of Pragmatics* 43 (3). 814-827.

----. 2013. Pointing gestures in video blogs. *Text&Talk* 33(1). 1-23.

----. 2014. Audience design in monologues: How vloggers involve their viewers. *Journal*

of Pragmatics, DOI: 10.1016/j.pragma.2014.02.008. available online April 12, 2014.

----; and Richard Harper. forthcoming. Tying in comment sections: The production of meaning and sense on Facebook. accepted for publication in *Semiotica*.

Garfinkel, Harold; and Harvey Sacks. 1970. On formal structures of practical action. *Theoretical sociology*, ed. by John C. McKinney; and Edward A. Tiryakian, 338-366. New York: Appleton-Century-Crofts.

Garrod, Simon; and Martin Pickering. 2004. Why is conversation so easy? *Trends in Cognitive Sciences* 8 (1). 8-11.

Gauntlett, David. 2011. *Making is connecting*. Cambridge: Polity.

Gehl, Robert. 2009. YouTube as archive: Who will curate this digital Wunderkammer? *International Journal of Cultural Studies* 12 (1). 43-60.

Gerhardt, Cornelia. 2008. *Talk by television viewers watching live football matches: Coherence through interactionality, intertextuality, and multimodality*. Saarbrücken: Saarland University dissertation.

Giltrow, Janet. 2013. Genre and computer-mediated communication. Cf. *Handbooks of Pragmatics, vol 9: Pragmatics of computer-mediated communication*, ed. by Susan Herring; Dieter Stein; and Tuija Virtanen, 717-738. Berlin: Mouton de Gruyter.

----; and Dieter Stein. 2009. *Genres in the Internet*. Amsterdam: John Benjamins.

Glick, Douglas. 2007. Some performative techniques of stand-up comedy: An exercise in the textuality of temporalization. *Language & Communication* 27 (3). 291-306.

Goffman, Erving. 1974 (1986). *Frame analysis*. New York: Harper & Row.

----. 1981. *Forms of talk*. Oxford: Blackwell.

----. 1981. *Forms of talk*. Philadelphia: University of Pennsylvania Press.

Gold, Ruby. 1991. Answering machine talk. *Discourse Processes* 14 (2). 243-260.

Goodwin, Charles. 1986. Audience diversity, participation and interpretation. *Text* 6 (3). 283-316.

----. 2000a. Action and embodiment within situated human interaction. *Journal of Pragmatics* 32. 1489-1522.

----. 2000b. Practices of seeing visual analysis: An ethnomethodological approach. *Handbook of Visual Analysis*, ed. by Theo van Leeuwen; and Carey Jewitt, 157-182. London: Sage.

----. 2003. Pointing as situated practice. *Pointing: Where language, culture, and cognition meet,* ed. by Sotaro Kita, 217-241. Mahwah: Erlbaum.

----; and Alessandro Duranti. 1992. Rethinking context: An introduction. *Rethinking Context: Language as an Interactive Phenomenon,* ed. by Alessandro Duranti; and Charles Goodwin, 1-42. Cambridge: Cambridge University Press.

Gracy, Karen F. 2007. Moving image preservation and cultural capital. *Library Trends* 56 (1). 183-197.

Gumperz, John J. 1982. *Studies in interactional sociolinguistics, vol 1*: *Discourse strategies*. Cambridge: Cambridge University Press.

----. 2003. Response essay. *Language and interaction: Discussions with John J. Gumperz*, ed. by Susan L. Eerdmans; Carlo L. Prevignano; and Paul J. Thibault, 105-126. Amsterdam: John Benjamins.

Hanks, William F. 2005. Explorations in the deictic field. *Current Anthropology* 46 (2). 191-220.

Haridakis, Paul; and Gary Hanson. 2009. Social interaction and co-viewing with YouTube: Blending mass communication reception and social connection. *Journal of Broadcasting & Electronic Media* 53 (2). 317-335.

Harley, Dave; and Geraldine Fitzpatrick. 2009. Creating a conversational context through video blogging: A case study of Geriatric1927. *Journal of Computers in Human Behavior* (special issue *Enabling elderly users to create and share self authored multimedia content*) 25 (3). 679-689.

Hasan, Ruqaiya. 1978. Text in the systemic-functional model. *Current Trends in Text Linguistics*, ed. by Wolfgang Dressler, 228–246. Berlin: de Gruyter.

Hasegawa, Yoko. 2010. *Soliloquy in Japanese and English*. Amsterdam: John Benjamins.

Have, Paul ten. no date (1990). Methodological issues in conversation analysis. *Bulletin de Méthodologie Sociologique* 27. 23-51.
[*http://www.pscw.uva.nl/emca/mica.htm*] (12.3.2013).

Haviland, John B. 2004. Gesture*. A companion to linguistic anthropology,* ed. by Alessandro Duranti, 197-221. Malden, MA: Blackwell.

----. 2007. Master speakers, master gesturers: A string quartet master class. *Gesture and the dynamic dimension of language: Essays in honor of David McNeill*, ed. by Susan D. Duncan; Elena T. Levy; and Justine Cassell, 147-172. Amsterdam/Philadelphia: John Benjamins.

Herring, Susan C. 1999. Interactional coherence in CMC. *Journal of Computer-Mediated Communication* 4 (4). [*http://jcmc.indiana.edu/vol4/issue4/herring.html*] (03.01.2013)

----. 2007. A faceted classification scheme for computer-mediated discourse. *Language@Internet 4*, article 1.
[*http://www.languageatinternet.de/articles/2007/761*] (11.03.2013)

----. forthcoming. Discourse in web 2.0: Familiar, reconfigured, and emergent. *Georgetown University Round Table on Languages and Linguistics 2011: Discourse 2.0: Language and new media,* ed. by Deborah Tannen; and Anna Marie Trester, 1-26. Washington, DC: Georgetown University Press.

----; Scheidt, Lois; Bonus, Sabrina; and Elijah Wright. 2004. Bridging the gap: A genre analysis of weblogs. *Proceedings of the thirty-seventh Hawaii International Conference on System Sciences (HICSS-37) 2004.* Los Alamitos, CA: IEEE Press.

----; Scheidt, Lois; Wright, Elijah; and Sabrina Bonus. 2005. Weblogs as a bridging genre. *Information Technology & People* 18 (2). 142-171.

----; Dieter Stein; and Tuija Virtanen (eds.). 2013. *Handbooks of Pragmatics, vol 9: Pragmatics of computer-mediated communication*. Berlin: Mouton de Gruyter.

----; Dieter Stein; and Tuija Virtanen. 2013. Introduction to the pragmatics of computer-mediated communication. Cf. *Handbooks of Pragmatics, vol 9: Pragmatics of computer-mediated communication*, ed. by Susan Herring; Dieter Stein; and Tuija Virtanen, 3-31. Berlin: Mouton de Gruyter.

Hess, Aaron. 2009. Resistance up in smoke: Analyzing the limitations of deliberation on YouTube. *Critical Studies in Media Communication* 26 (5). 411-434.

Hincks, Rebecca. 2004. Standard deviation of F_0 in student monologue. *Proceedings of the 17th Swedish Phonetics Conference Fonetik 2004*, 132-135. Department of Linguistics, Stockholm University.

Hoffmann, Christian. 2012. *Cohesive Profiling. Meaning and interaction in personal weblogs*. Amsterdam: John Benjamins.

Honeycutt, Courtenay; and Susan Herring. 2009. Beyond microblogging: Conversation and collaboration via Twitter. *Proceedings of the forty-second Hawaii International Conference on System Sciences (HICSS-42) 2009*. Los Alamitos, CA: IEEE Press.

Horton, Donald; and R. Richard Wohl. 1956. Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry* 19 (3). 215-229.

Hutchby, Ian; and Robin Wooffitt. 1998. *Conversation analysis: Principles, practices, and applications*. Cambridge: Polity.

Hutchins, Edwin; and Leysia Palen. 1997. Constructing meaning from space, gesture, and speech. *Discourse, tools and reasoning: Essays on situated cognition,* ed. by Lauren Resnick, 23-40. Berlin: Springer.

Hymes, Dell. 1974. *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania.

Jean Christian, Aymar. 2009. Real vlogs: The rules and meanings of online personal videos. *First Monday* 14 (11). [*http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/ article/view/2699/2353*] (29.12.2012)

Jeffries, Laura. 2011. The revolution will be soooo cute: YouTube "hauls" and the voice of young female consumers. *Studies in Popular Culture* 33 (2). 59-75.

Jenkins, Henry. 2006. Youtube and the vaudeville aesthetic. *http://www.henryjenkins.org/2006/11/youtube_and_the_vaudeville_aes.html.* (29.10.2012)

Jisa, Harriet. 1987. Sentence connectors in French children's monologue performance. *Journal of Pragmatics* 11. 607-621.

Johnstone, Barbara. 1987. Perspectives on repetition. *Text* 7 (3). 205-311.

Jones, Graham M.; and Bambi B. Schieffelin. 2009. Talking text and talking back: ''My BFF Jill'' from boob tube to YouTube. *Journal of Computer-Mediated Communication* 14 (4). 1050-1079.

Jones, Rodney. 2004. The problem of context in computer mediated communication. *Discourse and technology: Multimodal discourse analysis*, ed. by Philip LeVine; and Ron Scollon, 20-33. Washington, DC: Georgetown University Press.

----. 2008. The role of text in televideo cybersex. *Text and Talk* 28 (4). 453-473.

----; and Christoph Hafner. 2012. *Understanding digital literacies*. New York: Routledge.

Jules-Rosette, Benetta. 1985. Harold Garfinkel: La contribution de l'ethnomethodologie à la recherche sociologique. *Sociétés: Revue des sciences humaines et sociales* 5. 35-38.

Jung, Euen Hyuk (Sarah). 2006. Misunderstanding of academic monologues by nonnative speakers of English. *Journal of Pragmatics* 38 (11). 1928-1942.

Kallmeyer, Werner; Schütte, Wilfried; and Daniela Heidtmann. 2001. GAIS – Gesprächsanalytisches Informationssystem Grundlagentheoretisches Glossar. http://www.ids-mannheim.de/prag/lehre/HS_SS2001/Alt/GlossBeginnEnde.html (19.05.2013)

Kazandjian, Seta; Joan Borod; and Adam Brickman. 2007. Facial expression during emotional monologues in unilateral stroke: An analysis of monologue segments. *Applied Neuropsychology* 14 (4). 235-246.

Keating, Elizabeth; and Gene Mirus. 2003. American Sign Language in virtual space: Interactions between deaf users of computer-mediated video communication and the impact of technology on language practices. *Language in Society* 32 (5). 693-714.

----; Terra Edwards; and Gene Mirus. 2008. Cybersign and new proximities: Impacts of new communication technologies on space and language. *Journal of Pragmatics* 40. 1067-1081.

Kendon, Adam; and Laura Versante. 2003. Pointing by hand in "Neapolitan." Cf. *Pointing: Where language, culture, and cognition meet,* ed. by Sotaro Kita, 109-138. Mahwah: Erlbaum.

Kita, Sotaro. 2003. *Pointing: Where language, culture, and cognition meet*. Mahwah: Erlbaum.

----. 2009. Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes* 24 (2). 145-167.

Knoblauch, Hubert. 2008. The performance of knowledge: Pointing and knowledge in powerpoint presentations. *Cultural Sociology* 2 (1). 75-97.

Koch, Peter; and Wulf Oesterreicher. 1985. Sprache der Nähe – Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 39. 15-43.

Kotthoff, Helga. 1994. Zur Rolle der Konversationsanalyse in der interkulturellen Kommunikationsforschung: Gesprächsbeendigungen im Schnittfeld von Mikro und Makro. *Zeitschrift für Literaturwissenschaft und Linguistik* 93. 75-96.

Kozinets, Robert V. 2010. *Netnography: Doing ethnographic research online*. London: Sage.

Krauss, Robert M.; Yihsiu Chen; and Rebecca F. Gottesman. 2000. Lexical gestures and lexical access: A process model. *Language and gesture,* ed. by David McNeill, 261-283. Cambridge: Cambridge University Press.

Kress, Gunther; and Theo van Leeuwen. 2001. *Multimodal discourse: The modes and media of contemporary communication*. London: Arnold.

Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

----; and David Fanshel. 1977. *Therapeutic discourse: Psychotherapy as conversation*. New York: Academic Press.

Lange, Patricia. 2007a. Searching for the 'You' in 'YouTube:' An analysis of online response ability. *Being heard: Proceedings of the Ethnographic Praxis in Industry Conference 2007*, *National Association for the Practice of Anthropology,* 36-49. Berkeley, CA: University of California Press.

----. 2007b. Publicly private and privately public: Social networking on YouTube. *Journal of Computer-Mediated Communication* 13 (1). 361-380.

----. 2007c. The vulnerable video blogger: Promoting social change through intimacy. *The Scholar and Feminist Online* 5 (2). [*http://sfonline.barnard.edu/blogs/lange_01.htm*] (29.10.2012)

----. 2007d. *Commenting on comments: Investigating responses to antagonism on YouTube.* Paper presented at the Society for Applied Anthropology Conference, Tampa, Florida.

----. 2008. (Mis)Conceptions about YouTube. *Video vortex reader: Responses to YouTube*, ed. by Geert Lovink; and Sabine Niederer, 87-100. Amsterdam: Institute of Network Cultures.

Laver, John. 1975. Communicative functions of phatic communion. *Organization of behavior in face-to-face interaction*, ed. by Adam Kendon; Richard M. Harris; and Mary Ritchie Key, 215-238. The Hague: Mouton.

----. 1981. Linguistic routines and politeness in greeting and parting. *Rasmus Rask studies in pragmatic linguistics, vol 2: Conversational routine: Exploration in standardized communication situations and prepatterned speech*, ed. by Florian Coulmas, 289-304. The Hague: Mouton.

Leech, Geoffrey. 1983. *Principles of Pragmatics*. London: Longman.

Liddicoat, Anthony. 1994. Discourse routines in answering machine communication in Australia. *Discourse Processes* 17 (2). 283-309.

Lin, Chia-Yen. 2010. '. . . that's actually sort of you know trying to get consultants in . . .': Functions and multifunctionality of modifiers in academic lectures. *Journal of Pragmatics* 42. 1173-1183.

Lindgren, Simon. 2012. "It took me about half an hour, but I did it!": Media circuits and affinity spaces around how-to videos on YouTube. *European Journal of Communication* 27 (2). 152-170.

Lister, Martin; Dovey, Jon; Giddings, Seth; Grant, Iain; and Kieran Kelly. 2009. *New media: A critical introduction*, 2nd edition. London: Routledge.

Malinowski, Bronislav. 1948 (1926). *Magic, science and religion*. New York: Doubleday.

Markham, Annette; and Elizabeth Buchanan. *2012. Ethical decision-making and Internet research 2.0: Recommendations from the aoir ethics working committee.* [*www.aoir.org/reports/ethics2.pdf*] (28.02.2013)

Marlow, Cameron; Naaman, Mor; boyd, danah; and Marc Davis. 2006. HT06, tagging paper, taxonomy, Flickr, academic article, to read. *HYPERTEXT '06: Proceedings of the 17th conference on Hypertext and Hypermedia 2006, Odense, Denmark*, 31-40. New York: ACM.

Matthews, Nicole. 2007. Confessions to a new public: Video nation shorts. *Media, Culture & Society* 29 (3). 435-448.

McKee, Alan. 2011. YouTube versus the National Film and Sound Archive: Which is the more useful resource for historians of Australian television? *Television & New Media* 12 (2). 154-173.

McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: Chicago University Press.

----. 2003. Pointing and morality in Chicago. *Pointing: Where language, culture, and cognition meet,* ed. by Sotaro Kita, 293-306. Mahwah: Erlbaum.

----. 2005. *Gesture and thought*. Chicago: Chicago University Press.

----; Cassell, Justine; and Elena Levy. 1993. Abstract deixis. *Semiotica* 95: 1/2, 5–19.

Mead, George. 1934. *Mind, self, and society: From the standpoint of a social behaviorist*. Chicago: University of Chicago Press.

Meichenbaum, Donald. 1977. *Cognitive-behavior modification: An integrative approach*. New York: Plenum Press.

Mindt, Ilka. 2008. Appropriateness in discourse: The adjectives *surprised* and *surprising* in monologue and dialogue. *Journal of Pragmatics* 40 (9). 1503-1520.

Montgomery, Martin. 1981. The structure of monologue. *Studies in discourse analysis,* ed. by Malcolm Coulthard; and Martin Montgomery, 33-39. London: Routledge.

Moor, Peter; Ard Heuvelman; and Ria Verleur. 2010. Flaming on YouTube. *Journal of Computers in Human Behavior: Online interactivity: Role of technology in behavior change* 26 (6). 1536-1546.

Morley, David. 2006. Unanswered questions in audience research. *The Communication Review* 9 (2). 101-121.

murmur project. http://murmur.info/. [no author, no publishing year] (19.05.2013)

Norrick, Neal R. 1987. Functions of repetition in conversation. *Text* 7 (3). 245-264.

----. 2002. Poetics and conversation. *Connotations: A Journal for Critical Debate* 10. 243-267.

----. 2007. Conversational storytelling. *The Cambridge companion to narrative*, ed. by David Herman, 127-141. Cambridge: Cambridge University Press.

----; and Claudia Bubel. 2009. Direct address as a resource for humor. *Humor in interaction*, ed. by Neal R. Norrick; and Delia Chiaro, 29-47. Amsterdam: John Benjamins.

Norris, Sigrid. 2004. Multimodal discourse analysis: A conceptual framework. *Discourse and technology: Multimodal discourse analysis*, ed. by Philip Levine; and Ron Scollon, 101-115. Washington, DC: Georgetown University Press.

Ochs, Elinor. 1979. Transcription as Theory. *Developmental Pragmatics*, ed. by Elinor Ochs; and Bambi Schieffelin, 43-72. New York: Academic Press.

----. 1993. Constructing social identity: A language socialization perspective. *Research on Language and Social Interaction* 26 (3). 287-306.

O'Donnell, Susan; Gibson, Kerri; Milliken, Mary; and Janice Singer. 2008. Reacting to YouTube videos: Exploring differences among user groups. *Proceedings of the International Communication Association Annual Conference May 22-26, 2008 (ICA 2008)*, ed. by . Montreal, Quebec. *http://iit-iti.nrc-cnrc.gc.ca/iit-publications-iti/docs/NRC-50361.pdf.* (17.02.2013)

O'Reilly, Tim. 2005. What is Web 2.0: Design patterns and business models for the next generation of software.

http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/
what-is-web-20.html. (29.10.2012)

Othman, Zarina. 2010. The use of okay, right and yeah in academic lectures by native speaker lecturers: Their 'anticipated' and 'real' meanings. *Discourse Studies* 12 (5). 665-681.

Page, Ruth. 2010. Introduction. *Routledge Studies in Multimodality, vol 1: New Perspectives on Narrative and Multimodality*, ed. by Ruth Page, 1-13. London: Routledge.

Paltridge, Brian. 1994. Genre analysis and the identification of textual boundaries. *Applied Linguistics* 15 (3). 288-299.

Pascual, Esther. 2006. Questions in legal monologues: Fictive interaction as argumentative strategy in a murder trial. *Text & Talk* 26 (3). 383-402.

Peirce, Charles. 1960. *Collected Papers of Charles Sanders Peirce, vol. 4.6.* Cambridge: Harvard University Press.

Pfister, Manfred. 1977. *Das Drama*. München: Fink.

Piaget, Jean. 1959. *The language and thought of the child*, 3rd edition, trans. by Marjorie and Ruth Gabain. London: Routledge & Kegan Paul. [French original: Le langage et la pensée chez l'enfant. Paris.]

Puschmann, Cornelius. 2010. *The corporate blog as an emerging genre of computer-mediated communication: Features, constraints, discourse situation*. Göttingen: Universitätsverlag Göttingen.

Rawls, Anne W. 2004. Conversation Analysis. *Encyclopedia of Social Theory*, ed. by George Ritzer, 146-150. Thousand Oaks, CA: Sage. SAGE Reference Online. 5 Mar. 2012.[http://www.sagepub.com/ritzerintro/study/materials/reference/77708_5.1r.pdf] (26.02.2013)

Rendle-Short, Johanna. 2005. Managing the transitions between talk and silence in the academic monologue. *Research on Language and Social Interaction* 38 (2). 179-218.

Richards, Jack C.; John Platt; and Heidi Platt. 1992. *Longman dictionary of language teaching and applied linguistics*. Harlow: Longman.

Rintel, E. Sean; Joan Mulholland; and Jeffery Pittam. 2001. First things first: Internet relay chat openings. *Journal of Computer-Mediated Communication* 6 (3). [*http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2001.tb00125.x/full*] (09.11.2012)

Rosen, Jay. 1987. An equal right to inequality: The sociology of the answering machine. *ETC: A Review of General Semantics* 44 (3). 288-290.

Sacks, Harvey. 1987. On the preferences for agreement and contiguity in sequences in conversation. *Talk and social organisation*, ed. by Graham Button; and John R.E. Lee, 54-69. Philadelphia: Multilingual Matters.

----. 1995. *Lectures on conversation*. Oxford: Blackwell.

----; Emanuel Schegloff; and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50 (4). 696-735.

Schegloff, Emanuel A. 1967. The first five seconds: The order of conversational openings. Unpublished Ph. D. Dissertation. University of California, Berkeley.

----. 1968. Sequencing in conversational openings. *American Anthropologist* 70 (6). 1075-1095.

----. 1986. The routine as achievement. *Human Studies* 9 (2-3). 111-151.

----. 1987. Between macro and micro: Contexts and other connections. *The micro-macro link*, ed. by Jeffrey C. *Alexander* et al., 207-234. Berkeley, CA: University of California Press.

----. 1992. In another context. *Studies in the social and cultural foundations of language, vol 11: Rethinking context: Language as an interactive phenomenon*, ed. by Alessandro Duranti; and Charles Goodwin, 191-228. Cambridge: Cambridge University Press.

----; and Harvey Sacks. 1973. Opening up closings. *Semiotica* 8 (4). 289-327.

Scheidt, Lois Ann. 2009. Diary weblogs as genre. Indiana University dissertation. *http://professional-lurker.com/linked/2009/qp/Diary_Weblogs_as_Genre2009.pdf.* (29.06.2010)

Siersdorfer, Stefan; Chelaru, Sergiu; and Wolfgang Nejdl. 2010. How useful are your comments?: Analyzing and predicting YouTube comments and comment

ratings. *Proceedings of the 19th international conference on World Wide Web (WWW 2010) 2010*, 891-900. New York, NY: ACM.

Schiffrin, Deborah. 1987. *Discourse markers*. Cambridge: Cambridge University Press.

----. 1994. *Approaches to discourse*. Oxford: Blackwell.

----. 2001. Discourse markers: Language, meaning, and context. *The handbook of discourse analysis*, ed. by Deborah Schiffrin; Deborah Tannen; and Heidi E. Hamilton, 54-75. Oxford: Blackwell.

Schwarz, Jeannine. 2010. *Linguistic aspects of verbal humor in stand-up comedy*. Saarbrücken: Saarland University dissertation.

Sinclair, John; and Michael Coulthard. 1975. *Toward an analysis of discourse: The English used by teachers and pupils.* Oxford: Oxford University Press.

Smith, Sara W.; Noda, Hiromi Pat; Andrews, Steven, and Andreas H. Jucker. 2005. Setting the stage: How speakers prepare listeners for the introduction of referents in dialogues and monologues. *Journal of Pragmatics* 37. 1865-1895.

Snelson, Chareen; and Ross A. Perkins. 2009. From silent film to YouTube: Tracing the historical roots of motion picture technologies in education. *Journal of Visual Literacy* 28 (1). 1-27.

Sperber, Dan; and Deirdre Wilson. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.

Strodt-Lopez, Barbara. 1991. Tying it all in: Asides in university lectures. *Applied Linguistics* 12 (2). 117-140.

Swales, John M. 1990. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Tannen, Deborah. 1987. Repetition in conversation as spontaneous formulaicity. *Text* 7 (3). 215-243.

----. 1989. *Talking voices: Repetition, dialogue, and imagery in conversational discourse*. New York/Cambridge: Cambridge University Press.

----. 2004. Talking the dog: Framing pets as interactional resources in family discourse. *Research on Language and Social Interaction* 37 (4). 399-420.

----. 2007 (1989). *Talking voices: Repetition, dialogue, and imagery in conversational discourse*. Cambridge: Cambridge University Press

TED. http://www.ted.com/. [no author, no publishing year] (19.05.2013)

TED – About TED. http://www.ted.com/pages/204. [no author, no publishing year] (19.05.2013)

TED – Speaking at TED. http://www.ted.com/pages/speakingatted. [no author, no publishing year] (19.05.2013)

Thompson, Susan. 1994. Aspects of cohesion in monologue. *Applied Linguistics* 15 (1). 58-75.

----. 2003. Text-structuring metadiscourse, intonation and the signalling of organisation in academic lectures. *Journal of English for Academic Purposes* 2 (1). 5-20.

Thurlow, Crispin; and Kristine Mroczek (eds.). 2011. *Oxford studies in sociolinguistics: Digital discourse: Language in the new media*. New York: Oxford University Press.

Tolson, Andrew. 2005. *Media talk: Spoken discourse on TV and radio.* Edinburgh: Edinburgh University Press.

----. 2010. A new authenticity?: Communicative practices on YouTube. *Critical Discourse Studies* 7 (4). 277-289.

Van Dijk, Teun. 2008. *Discourse and context: A sociocognitive approach*. Cambridge: Cambridge University Press.

Van Leeuwen, Theo. 2009. Parametric systems: The case of voice quality. *The Routledge handbook of multimodality*, ed. by Carey Jewitt, 68-77. London: Routledge.

Verschueren, Jef. 1995. The pragmatic perspective. *Handbook of Pragmatics*, ed. by Jef Verschueren; Jan-Ola Östman; and Jan Blommaert, 1-20. Amsterdam: John Benjamins.

Vygotsky, Lev. 1986. Thought and language, 2nd edition, trans. by Eugenia Hanfmann; and Gertrude Vakar, ed. by Alex Kozulin. Cambridge: Massachusetts Institute of Technology Press. [Russian original: Myshlenie i rech.]

Waldvogel, Joan. 2007. Greetings and closings in workplace email. *Journal of Computer-Mediated Communication* 12 (2). [*http://jcmc.indiana.edu/vol12/issue2/waldvogel.html*] (07.11.2012)

Walther, Joseph B.; DeAndrea, David; Kim, Jinsuk; and James C. Anthony. 2010. The influence of online comments on perceptions of antimarijuana public service announcements on YouTube. *Human Communication Research* 36 (4). 469-492.

Watson, John. 1925. *Behaviorism*. New York: Norton.

Watson, Rod. 2009. *Analysing practical and professional texts: A naturalistic approach*. Farnham, U.K: Ashgate.

Weir, Ruth. 1962. *Language in the crib*. The Hague: Mouton.

Wells, Pam; and Peter Bull. 2007. From politics to comedy: A comparative analysis of affiliative audience responses. *Journal of Language and Social Psychology* 26 (4). 321-342.

Werry, Christopher C. 1996. Linguistic and interactional features of internet relay chat. *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*, ed. by Susan Herring, 47-63. Amsterdam: John Benjamins.

YouTube. *www.youtube.com.* [no author, no publishing year]

YouTube. Terms of Service. [*http://www.youtube.com/t/terms?gl=US*] (24.03.2013) [no author, no publishing year]

Yus, Francisco. 2011. *Cyberpragmatics: Internet-mediated communication in context*. Amsterdam: John Benjamins.

Zink, Michael; Yu Gu, Kyoungwon Suh; and Jim Kurose. 2009. Characteristics of YouTube network traffic at a campus network: Measurements, models, and implications. Content Distribution Infrastructures for Community Networks. *Computer Networks* 53 (4). 501-514.

Zwicky, Arnold M. 1974. Hey, Whatsyourname. *Papers from the tenth regional meeting: Chicago Linguistic Society April 19-21, 1974*, ed. by Michael W. La Galy; Robert A. Fox; and Anthony Bruck, 787-801. Chicago: Chicago Linguistic Society.