

Potentials and Limitations of Discourse-Centred Online Ethnography

Jannis Androutsopoulos

King's College London

urn:nbn:de:0009-7-16100

Abstract

This article outlines aims and procedures of *discourse-centred online ethnography*, a combination of systematic observation of online activities and interviews with online actors, which was developed as a complement to the linguistic analysis of log data. Drawing on research on computer-mediated discourse on German-based websites of hip-hop and diaspora groups, I discuss how ethnographic data fit in with a sociolinguistic approach to computer-mediated communication. I suggest that ethnographic insights may supplement and support the linguistic analysis of log data in three ways: reconstructing fields of computer-mediated discourse, reconstructing participants' literacy practices, and analysing their sociolinguistic awareness.

Introduction

Ethnography on the Internet is a multifaceted trend. Terms such as "virtual ethnography" (Hine, 2000), "network ethnography" (Howard, 2002), "netnography" (Kozinets, 2002), "cyberethnography" (Domínguez et al., 2007) and "webnography" (Puri, 2007) indicate researchers' attempts at transferring principles and techniques of ethnography to settings of computer-mediated communication (CMC). Such transfer is discussed in this article from the point of view of language-focused CMC research.

Over the last 10 years or so, this research has broadened its scope from specific linguistic features to the wider "social and discourse complexities of the socio-technical system" of CMC (Cherny, 1999, p. 298). In what I identify as the "first wave" of language-focused CMC research (Androutsopoulos, 2006), the focus was on features and strategies that are (assumed to be) specific to new media; the effects of communications technologies on language were given priority over other contextual factors. The data were often randomly collected and detached from their discursive and social contexts, and generalisations were organised around media-related distinctions such as language of emails, newsgroups, etc. (Ferrara, Brunner, & Whittemore, 1991; Crystal, 2001). A second wave of language-focused CMC studies is informed by pragmatics, sociolinguistics, and discourse studies, and

emphasises situated language use and linguistic diversity (Androutsopoulos, 2006); however, the exclusive study of log data¹ still prevails. Even though data collection involving direct contact with Internet users through means such as surveys, interviews, and participant observation has been often advocated and occasionally carried out in combination with log data (e.g., Baym, 2000; Cherny, 1999; Herring, 1996), such combinations have played a somewhat peripheral role in language-focused CMC studies thus far.

My point of departure in this article is the assumption that research based exclusively on log data is not ideally positioned to examine participants' discourse practices and perspectives or to relate these practices and perspectives to observable patterns of language use. The notion of practices and perspectives includes questions about people's motivations for the use of particular linguistic resources online and the meanings they attach to those resources; people's awareness and evaluation of linguistic diversity online; their knowledge about the origin and circulation of linguistic innovations in CMC; and the relationship between participants' and researchers' interpretations. I suggest that addressing these and similar issues requires going beyond what is observable on the screen (see also Beißwenger and Marcoccia, Atifi, & Gauducheau, this issue), and I propose *discourse-centred online ethnography* as a step in this direction.

Discourse-Centred Online Ethnography (DCOE)

The combination of methods I tentatively call discourse-centred online ethnography (DCOE)² combines the systematic observation of selected sites of online discourse with direct contact with its social actors. It thus encompasses, and extends beyond, systematic observation, which is part of Herring's computer-mediated discourse analysis (2004) framework and other language-focused CMC work. DCOE uses ethnographic insights as a backdrop to the selection, analysis, and interpretation of log data, in order to illuminate relations between digital texts and their production and reception practices.

Before discussing procedures of DCOE in the next section, I first contextualize my approach in relation to previous work on ethnography, especially in sociolinguistics. Ethnography is understood in this article primarily as a method, not an epistemology (see Agar 1995)—i.e., as a way of doing Internet research. The specific background to my approach is the ethnography of communication (Hymes, 1996; Saviile-Troike, 2003) and subsequent uses of ethnography in socially-oriented linguistics (see Eckert, 2000 and Rampton, 2006 for two different cases in point). Such research

aims at studying patterns of communication and social relationships accomplished through language in a community or group. It seeks to understand the social meaning of different ways of using language by taking into account participants' awareness and interpretation of their practices, and by relating language to the social categories and activities of a community (rather than to abstract macro-sociological classifications). Even though many sociolinguistic studies eventually focus on details of linguistic variability, ethnography is an essential background to such a focus and the main resource for understanding the meaning of variability from the members' perspective. In this sense, an effect of combining ethnography and linguistics is "opening linguistics up" (Rampton, 2006, pp. 394-395) to the richness of the social context of language use. Deppermann (2000) has argued for a similar opening with regard to conversation analysis by proposing an "ethnographic conversation analysis" in which ethnographic knowledge is "inserted" into conversation analysis, broadening the scope of interpretation beyond what can be immediately accounted for from transcripts alone.

The value of doing ethnography on the Internet is not only as a research tool but also as a conceptual and methodological bridge to other research traditions. In this respect, it seems safe to say that it is ethnography (rather than linguistics) that connects language-focused research to other streams of CMC studies. This is perhaps most obvious with regard to an adjacent discipline such as (new) literacy studies and its branches, such as multiliteracy and digital literacy research (e.g., Hawisher & Selfe, 2000), in which ethnography is the canonical method. In this context, the ethnography of writing, a rather neglected side aspect of the ethnography of communication, has been rediscovered in the study of Internet literacies. Danet (2001) summarises the main questions of the ethnography of writing as follows:

Researchers ask: Who uses writing for what purposes? What genres and subgenres of texts are recognized, and how do they develop? What media are considered appropriate for which kinds of messages, and what are the norms governing usage in the various genres? (p. 11)

Likewise, other branches of CMC research, based on anthropology, sociology, media studies, and social psychology, draw heavily on ethnography to study the local and situated character of Internet practices, to reconstruct the emergence of virtual communities, and to chart the unfolding of online activities in relation to offline events. As Döring (2003) puts it:

The reason ethnographic observation is so important to internet research is that it helps to describe larger social formations such as specific mailing lists, news groups, chat channels, and MUDs holistically, in their own structures and processes, from the participants' perspective. (pp. 174-175; my translation)

Thus even though the analytic questions explored below are specifically geared towards language use, the potential of online ethnography extends much further than that.

Rather than being a homogeneous project, online ethnography comes—as is perhaps typical for emergent fields of research—in different versions, which may be divided into two main types, depending on how they strike the balance between research online and offline (Greschke, 2007).³ The first type focuses on the Internet in everyday life, asking how new communications technologies are integrated into the life and culture of a community. It proceeds as blended ethnography, i.e., a blend of online and offline ethnography, with offline activities receiving equal or even more attention than online ones. An example is Miller and Slater's (2000) work on Internet use in Trinidad and among diaspora Trinidadians, which used interviews, door-to-door surveys, and observation of Internet use in social spaces. The second approach is concerned with everyday life *on* the Internet, theorising the Internet as a site where culture and community are formed. There is a tradition of ethnographies of online communities in chat channels or role-playing environments that make much use of participant observation but involve little or no other contact with members (Döring, 2003). An attempt to bridge the gap between Internet as culture and as cultural artefact is Hine's (2000) Internet ethnography. Hine's research takes as its point of departure an offline event (a court case) and follows the online activities related to that event, combining content analysis of websites and interviews with website producers.

Doing DCOE

This second type of online ethnography, and Hine's approach more specifically, provides the methodological background for my discourse-centred online ethnography, which was developed in two German-based projects (Androutsopoulos, 2003, 2007a, 2007b). The first (2000-2004) studied web environments dedicated to music youth cultures, especially hip-hop, looking in particular at the formation of sociolinguistic style and the construction of identities in CMC. The second project (2004-2005) studied

websites by and for ethnic minority groups in Germany, focusing on multilingual practices on these websites. Both projects investigate three dimensions of linguistic variability—spoken/written style, standard/non-standard varieties, and German/English contact—in a range of communication formats (home pages, edited genres, discussion boards) involving multimodal combinations. Most examples in the remainder of this article come from the hip-hop project, which involved more ethnographic observation and interviews. Drawing on this research, I discuss in this section a number of practice-based guidelines; these are summarised in Table 1 below.

The first pillar of DCOE, systematic observation, is of course implicit in much previous work on language in new media, but its procedures and its relationship to log data are seldom discussed in detail (cf. Baym, 2000; Cherny, 1999). My take on observation over time is rooted in the (tacit) assumption that the continuous monitoring of given sites of discourse affords insights into discourse practices and patterns of language use on these sites.⁴ It seems important to emphasise that ethnographic observation on the Internet does not target "static" textual artefacts that are simply being "gazed at." Rather, its targets are sets of relationships on two levels, i.e., within a particular site of discourse as well as across a set of such sites .

The first level comprises the sum of discourse units that make up a particular space of CMD, for instance a single website or discussion board. Systematic observation aims at charting the complex architecture of that space and understanding the various relations among its components. An example is the comparison of German-based migrant websites with respect to their multilingual practices (Androutsopoulos, 2007b). The second set of relationships is what I call a "field of computer-mediated discourse," i.e., a set of interconnected websites that represent a lifestyle or a social scene on the web (Androutsopoulos, 2003, 2007a). The notion of field here relates both to the ethnographer's sense of the space in which fieldwork is carried out and to Bourdieu's (1991) theory of social organisation and symbolic capital, to which I return below. German-speaking hip-hop is such a field in my research, and observation of it aims at uncovering its extension, boundaries, internal distinctions, and characteristic discourse practices. Such fields might seem vast and "limitless" at first sight (as suggested by McLelland 2002, p. 394, for the example of Japanese gay websites and discussion boards), but I argue that they can be charted and delimited (Androutsopoulos, 2007a).

Against this backdrop, systematic observation is concerned with the dynamics of communication and semiotic production within web environments. In

analogy to Danet's (2001) ethnography of writing, my first guideline is to ask: What activities are unfolding in these environments, what is their pace or rate of change, who are their main actors, and how do they interact or interrelate? Questions of this kind, tailored to the observation of virtual interaction in discussion boards and chat channels, are complemented by questions that focus on edited web content and its semiotic features: What are the semiotic (including linguistic) resources recurrently deployed in this field, what characteristic clusters do they form, and how do different environments, participants, and genres differ in their use of these resources? It seems accurate to say that the targets of ethnographic observation are both communicative activities *and* the semiotic artefacts produced through such activities, i.e., both "products" and processes taking place around these products.

Practice-derived guidelines for systematic observation:

- 1) Examine relationships and processes rather than isolated artefacts
 - 2) Move from core to periphery of a field
 - 3) Repeat observation
 - 4) Maintain openness
 - 5) Use all available technology
 - 6) Use observation insights as guidance for further sampling
-

Practice-derived guidelines for contact with Internet actors:

- 1) Contacts should be limited, non-random, and include various participation formats
 - 2) Pay attention to the initial contact
 - 3) Formulate and customise interview guidelines
 - 4) Confront participants with (their own) material
 - 5) Seek repeated and prolonged contacts
 - 6) Make use of alternative techniques whenever possible
-

Table 1. Practice-derived guidelines for DCOE

The second guideline for ethnographic observation is to move from the core to the periphery of the field under investigation. Researchers should first identify key nodes in a network of websites and then browse through to peripheral nodes, using for this purpose any orientation resources available in the field, such as link directories. The third guideline is to visit websites and discussion areas of interest repeatedly, in order to develop a "feel" for their discourses, emblems, and language styles. Sites of interest should be monitored over longer periods of time with regard to their genres, topics,

information sources, participation patterns, updating frequency, etc. Fourth, research should attend to the openness and fluidity of online discourse. Rather than isolating a particular forum section from the outset, it is better to browse around the whole forum regularly, identifying core participants, reading members' self-descriptions, etc. This point is aptly captured in the metaphor of "guerilla ethnography" (Yang, 2003, p. 471), which emphasises resisting too early a closure of the observation scope and legitimises the researchers' temptation to get "carried away" by, and immersed into, their material. As in traditional face-to-face ethnography, researchers should "take time to observe" their field of interest (Rampton, 2006, p. 397; see also Agar, 1995). Fifth, in this process the researcher should employ all technological resources that are available to participants to make sense of others' displayed identities and participation patterns. When observing discussion boards, for example, the researcher may use popularity statistics to select threads; numbers of posts as indicators of core participants; lists of posts by members (where available) to determine individual linguistic preferences; the search function of a forum to locate keywords; etc.

The data thus collected can be used as a starting point for linguistic analysis, e.g., forum statistics may be used to distinguish between regulars and newcomers in a forum, and this distinction may then be used in an analysis of sociolinguistic variation or conversational style. In this way, independent sociolinguistic variables are based on the discourse of the environments studied. Finally, observational data may also be used to provide guidance for further sampling. Systematic observation forms a backdrop against which to select text samples for fine-grained linguistic analyses or participants for interview contacts. Such sampling may be random/systematic or non-random/purposive: For instance, establishing by observation the main genres on a website can set the backdrop for compiling a random sample of these genres to be analysed with respect to spoken/written language (Androutsopoulos, 2007a). To take another example, identifying controversial discussion issues in a forum can be the basis for selecting discussion threads in order to examine the identity constructions these controversies prompt. If one is interested in multilingualism, identifying discussion topics that favour the use of a minority language can be the basis for selecting threads and posts for detailed analysis of code-switching (Androutsopoulos, 2007b; Siebenhaar, this issue).

The second dimension of DCOE, i.e. the contact with Internet actors, follows from the first, in that it draws on and is guided by observation and log-based analysis of CMD. My first practice-based guideline here is to work with non-

randomly selected informants; in my research, these were limited in number due to practical reasons. Approximately 25 interviews were carried out in the hip-hop project (with almost all participants being males in their teens or mid to late twenties), whereas the diaspora project included only four interviews with professional webmasters (all of them young adults in their twenties and thirties). Their selection was based on prior observation and textual analysis, taking into consideration both the "richness" of individual cases—e.g., how well they exemplify a participation format—and practical issues such as regional location. The selection of interviewees should offer insights into a range of perspectives within a field. It is therefore crucial to contact interviewees who exemplify different *participation formats*, e.g., amateur and professional ones, as identified by observation.

Preparation for an interview involves initiating and negotiating the contact, as well as formulating and customizing interview guidelines. The initial contact needs to be paid particular attention to, as it frames the interviewees' understanding of the research relationship and the researcher's identity. In my case it was established with an informal email, which clearly positioned me as a researcher, i.e., an outsider, but at the same time indicated my familiarity with the field. Especially in fields characterised by scepticism towards outsiders, researchers need to strike a balance between informality (but not flattery) and professional distance (but not stiffness). This email also guaranteed anonymity and offered an opt-out option.

I conducted semi-structured interviews, most of them face-to-face, although telephone or email had to be used in some cases. A generic interview guideline was produced and carefully customised based on analysis of the interviewee's homepage, posts, and other discursive engagement in the field. The interviews covered a broad range of topics, including the interviewee's aims and intended audience, their production practices, semiotic resources, personal opinion on hip-hop on the Web, and perceived relation between online and offline cultural practices. What proved to be quite fruitful for my purposes was confronting members with excerpts from their own or other websites and using the subsequent discussion to elicit their awareness and evaluation of language styles in these excerpts. The benefits of repeated and prolonged contacts with participants cannot be over-emphasised. I met a couple of amateur homepage producers twice, and my contact with one professional webmaster started with a phone interview, continued in an "academy-meets-community"-type conference where the webmaster was invited to participate in a panel discussion, and was carried over in a hip-hop

festival that the webmaster covered for his online magazine; different types of data were collected each time.

Beyond interviews, it is useful to make use of alternative techniques of direct contact with participants whenever this seems necessary or feasible. In one instance, a simple questionnaire on dialect writing on the Internet was administered to members of a chat channel. In a couple of cases a public chat discussion was followed by an individual chat with the character of an informal interview (my researcher identity was already disclosed). Even though this is obviously not consistent across sites, it is consistent with the notion of "guerilla ethnography," i.e., seizing the opportunity to use whatever methods are possible under the circumstances of each particular context.

I mentioned in passing how the initial contact email addressed issues of research ethics by disclosing myself as a researcher, guaranteeing anonymity, and offering an opt-out option, thus preparing the groundwork for participants' informed consent. While these measures conform to social-scientific standards with regard to research with human participants, ethical issues surrounding Internet observation seem less straightforward, not least due to the novel situation of an "invisible observer" who is not forced to disclose his or her presence and activity, so that participants may not be aware at all of being observed. There are different responses to this issue in the research literature,⁵ and its legal regulation varies across countries.⁶ A standard measure in my own research was to anonymise all personal information, including nicknames, which are a resource of individuation and, as evidence suggests, quite recognizable within the relevant online community. With respect to names of web environments, while I regularly anonymised amateur homepages, I did not do so for some commercial portals and large online magazines.⁷ With regard to log data, I consider obtaining permission a standard requirement with regard to data from restricted-access networks or private exchanges (e.g., mailing lists, text messaging), but treat logs from publicly-accessible boards and websites as public domain data. Some webmasters were contacted, notably with a view to conducting interviews, but seeking everyone's permission just did not seem feasible. Overall, it seems that if a minimum standard of privacy protection (such as anonymisation) is maintained, flexible case-by-case decisions may be tailored to the specific conditions of each project. Different research aims and web environments require different degrees of privacy protection, and a generic solution does not seem to do justice to the tremendous variety of digital communication and research.

The two projects the above discussion is based on did not exhaust the available ethnographic techniques. For example, researcher-initiated forum discussions, surveys, attending offline meetings or observing production and reception practices were not used, albeit due more to practical constraints than to theoretical decisions. Nonetheless, the data collected by systematic observation and interviews formed an adequate basis for ethnographic triangulation. Constant moving back and forth among observation notes, interview data, and web (textual) data offers insights that could not be gained by a purely log-based (or a purely observational) procedure; this is what the remainder of this article attempts to illustrate.

Assessing DCOE

In the remainder of this article, I draw on findings and examples from my research on hip-hop on the German-speaking web to illustrate how DCOE can contribute to the study of sociolinguistic style and variability in CMC. In a nutshell, the benefits of ethnography for my research were threefold: reconstructing fields of computer-mediated discourse, literacy practices, and participants' "lay sociolinguistics."

Reconstructing Fields of Computer-Mediated Discourse

Following previous CMC studies (Bräuchler, 2005; Döring, 2003), I assumed that the structure of a field of computer-mediated discourse could be conceived of in terms of a core-periphery scheme, a website's core position being indicated by its popularity and awareness among users. More than 800 sites were listed in German hip-hop directories during the time of my research, but the core of the field consisted of no more than 12-15 sites, which self-categorised as portals, online magazines, or discussion forums (for further discussion, see Androutsopoulos, 2003, 2007a). "Around" them, as it were, there operated smaller online magazines, dedicated discussion boards and chat channels, and a large number of homepages, both commercial (i.e., run by artists or music labels) and personal (run by fans, activists, and amateur artists). Thus the field encompasses a range of diverse formats that may be subdivided with regard to their genres, authors, or topics.

Understanding the components and the extension of a field, in turn, serves to contextualize linguistic variability. As the notion of field (Bourdieu, 1991) implies, a field of computer-mediated discourse is not a random collection of websites but a space of positions with differential access to resources (e.g., number of visitors or advertisement revenue) and different laws of price

formation. This captures quite well the contrast between sites of edited content (homepages, magazines) and community interaction discourse (boards, chats). Even though they are often adjacent to each other in virtual space, edited and community discourse differ significantly in terms of authorship, participation formats, institutionalisation and professionalisation. While online magazines are under pressure to conform to the conditions of the media market, discussion boards are characterised by a reduced authorial responsibility and considerable topical and stylistic freedom. This results in strikingly different language norms and stylistic practices, which were clearly borne out by micro-linguistic analyses and interviews alike (Androutsopoulos, 2007a). Thus, combined etic and emic evidence helps to establish a main distinction in the field, around which other findings may be organised.

Reconstructing Literacy Practices

While understanding the outline of a field is largely based on systematic observation, understanding member practices depends crucially on direct contact with them. It is especially through interviews that I grasped typical patterns of engagement in the online hip-hop culture. These include: the occasional poster to a board or chat; the regular member; the author of a personal homepage; those who assume various institutional functions such as forum moderators or freelance editors; and the managers of large, commercial websites, i.e., those who turn their passion into a profession. The positions on such a cline of increasing web engagement are not mutually exclusive but rather are in an implicational relationship. They tend to correlate with social variables such as age, as well as with different writing styles and repertoires. Members assuming professional roles (editors, moderators, webmasters) display a larger stylistic repertoire and style-switch as they move across different participation formats; at the same time, there is evidence of participants being excluded from certain functions (such as freelancers) because they are deemed to lack the necessary writing skills in standard German. Moreover, interviews reveal some of the (multimodal) strategies members develop in order to get on with their engagement in the field, such as naming themselves and their sites, creating logos, composing and decorating their sites, and promoting them. Small and partial as they may be, such insights into literacy practices intersect with and inform sociolinguistic analysis, inasmuch as they offer background knowledge and emic views of phenomena that would otherwise be accessed based on the researcher's understanding only.

Reconstructing Participants' "Lay Sociolinguistics"

A third benefit of ethnographic insights was their role in reconstructing participants' "lay sociolinguistics," i.e., their awareness of linguistic variability and its social meanings (Niedzielski & Preston, 2000). Observation and in particular interviews were essential to this end. Confronting hip-hop activists with samples of hip-hop writing style, or asking them, e.g., "What do you see as typical features of hip-hop writing?" helped to elicit their own categories and distinctions (e.g., "That's how I write in a forum, but not on my homepage"). These sometimes coincide with and confirm the analyst's findings, but at other times they offer new, unexpected insights that can be taken up in linguistic analysis.

Consider the case of "Alex," a mid-twenties manager of a core portal with a large forum. When asked about typical features of hip-hop style, Alex pointed out the following (all interviews were conducted in German and are given here in English translation):

Example 1. "Alex" (by email)

The ending –er is spelt as –a. All endings with –s are spelt as –z. Some hardcore rap freaks even respell every S as Z or double ZZ within words. This is judged as "underground" affiliation.

Alex formulates here two sets of "folk variable rules" with corresponding social meanings. The first set, i.e., the substitution of <a> for final <er>, and <z> for <s>, captures two widespread hip-hop spellings, which originated in African American Vernacular English (AAVE) discourse traditions and spread to German via global hip-hop.⁸ Their social meaning is implicit in my question, i.e., typical of hip-hop writing style. The second rule is to substitute <z> or <zz> for <s> *within* words. In terms of linguistic structure, this is an extension of the first rule, as it generalises the substitution of <z> for <s> to a wider environment, and the doubling, <zz>, lacks phonetic motivation. This is interpreted by Alex as an index of an even stronger commitment to the hip-hop community ("an underground affiliation"). The order of these rules is not random, but mirrors the evolution of vernacular spellings in hip-hop discourse (from phonologically motivated to purely visual ones), and their interpretation mirrors the sociolinguistic insight that in adolescence, extreme variants are produced by subculturally-engaged speakers (Eckert, 2000). However, Alex's analysis also confirms that "conscious folk perception is often categorical—speakers of certain "types" either use or do not use certain stereotypical variants" (Niedzielski & Preston, 2000, p. 148). His substitution rules do not

account for individual variability or for linguistic constraints on the variables. As a matter of fact, analysis of data from Alex's website suggests that most <z> variation is restricted to word-final instances in originally English words with cultural significance to hip-hop (such as *beatz*), while only a few observable cases of categorical substitution occur, and <zz> is quite rare and idiosyncratic.

As a whole, members' awareness about style features such as <z> was quite variable. While Alex's views are remarkably elaborated, others' awareness was much more limited and shaped by specific, often local, sources. Consider the case of "Anita," a female teenager who used <z> and other stylised "Black English" features (e.g., *2* for 'to', *da*, *friendz*) on her homepage:

Example 2. "Anita" and "Tim" (face-to-face interview)

Anita: Wu Tang. I've listened to them since I was eleven. They always have these statements in their booklets, where you can see how they write... that sort of slang!

Jannis: What's the difference between da und tha?

Anita: That's what Wu Tang use.

(Later on about *zed*)

Tim: Not everyone uses it... it delimits a bit from other people

Anita references a famous U.S. rap group (Wu Tang) and a local (German) fan forum as sources for her Black English features. But even though she did not seem to be overtly aware of the phonological distinctions these spellings represent, Anita and her friend Tim were quite outspoken about the aesthetic and social values of <z> to them. It fulfils the classic double function of slang or group-specific language, i.e., affiliation with a particular group and demarcation "from other people" (by which they meant young people who listen to mainstream pop music). Such a foregrounding of aesthetic distinctiveness at the expense of linguistic distinctions, and of local influence at the expense of "global" alliances, occurred repeatedly in my interviews. I therefore suggest that participants' understanding of marked linguistic features may be considerably more restricted than we might assume, and their social indexicalities do not necessarily coincide with our assumptions. In this case, rather than relating <z> to a global hip-hop alliance, they relate it to local practices and quite specific sources of influence.

Another unexpected trait that emerged from the interviews relates to participants' overt attitudes toward "hip-hop slang"—a member category for a writing style characterised by the frequency of spoken-like features and hip-hop related English (Androutsopoulos, 2007a). Consider the following example from a guestbook entry (the gloss features original English items in italics):

Example 3. Guest book entry

Yo Leude! Was sup? Phatte Seite. Grüße gehen raus an jeden einzelnen Head, der Hip Hop liebt und auch supported! smoke weed und schieß auf peace. I am out tha 'Scarface'

'Yo people! Whas sup? Phat page. Greetings to all heads who love hip-hop and support it too. smoke weed and don't give a damn about peace. I am out tha 'Scarface'

This is fairly typical for hip-hop guest books, with regard to the speech acts involved (greeting the site producers, praising their site, greeting like-minded readers, calling for actions or stances, concluding self-naming) and to linguistic choices: There is an abundance of English lexis, discourse markers, and chunks, including "Black English" stereotypes in lexis and spelling (see Androutsopoulos, 2009 for discussion). Consider now what an activist in his mid-twenties replied when asked whether he uses hip-hop slang on his homepage:

Example 4. "Max" (by email)

I avoid using current slang expressions, which I often encounter on some other webpages. For this reason I assume that *Hauptschüler* [original language use, J.A.] are less comfortable on my page ;->

Here the interviewee mockingly implies that hip-hop slang is suited for *Hauptschüler*, i.e., pupils from the lowest-status tier of secondary schooling in the German education system. By so doing, he links slang expressions to a rather stigmatised social position from which he distances himself. Thus rather than unanimously accepting hip-hop slang, e.g., as "symbolic resistance" to mainstream language in the unregulated space of online hip-hop, participants' attitudes seem to vary according to age and web experience. Older, more experienced interviewees such as Max treated hip-hop slang as indexical of social class. What we see here and elsewhere in the interview data is that lay sociolinguistic attitudes in the hip-hop field are neither homogeneous nor independent of the stratification of the field.

Rather, participants with certain positions within the field dissociate themselves from the field's most characteristic writing style, reproducing in this process mainstream stereotypes of the relation of vernacular language to education and social class.

My last point returns to one of the main research questions in Hine's (2000) Internet ethnography, i.e., the relation between online and offline processes. In the context of my research, this translates to both cultural and linguistic relations; I focus here on the latter. Log data made it clear that writing in community sections (boards, chats) was full of spoken style markers. But at the same time, some of my interviewees dissociate hip-hop slang from offline language use. Discussing text samples with two homepage producers in their late teens, I confronted them with example 4, which happens to be an entry from their guestbook. This prompted the following comments (the symbol (.) indicates a small pause, and :: a lengthening of the preceding vowel):

Example 5. "Wolfgang" (face-to-face interview)

Wolfgang: Based on this statement you notice that (.) that he is using again those ah clichés, well this 'yo', 'leute' spelt with a 'd', 'wassup', was geht, a mixture of German and English, 'phatte seite' with a 'ph', double 't', 'heads' of course once again, 'supported', 'smoke weed', and ah –

Jannis: Well is it all clichés then?

Wolfgang: Well these are those typical words that you use in hip-hop (.) now if you – let's put it this way, I don't go to my friend and say, "hey alter, wassu::p, und äh supportest du mich," I talk normally to people, well I never met one like this one who would come over to me and (.) "wassu::p, ich supporte dich, smoke nich so viel weed" and so on—well I never met such a guy.

Wolfgang does not deny that the language style exemplified by this guestbook entry is typical for hip-hop online. In fact, linguistic analysis of these youngsters' own writing indicates that their guestbook entries on others' websites are not dissimilar to example 3. However, they wrote differently in the edited parts of their own homepage. But Wolfgang here dissociates such a writing style from speech in the local hip-hop community, and judging from the stylised direct speech in his second turn, the main difference in what he calls "talking normally to people" is that it is not overloaded with English hip-hop slang as these stylizations are. Again, ethnographic knowledge suggests that "normal" speech in Wolfgang's local

community was marked by regional and ethnic variation rather than just English lexis and idioms. Here, as in other cases in my data, participants' accounts invite us to see CMC as a discourse space with its own potentials of style and styling, which draw on offline speech styles without being identical to them.⁹ At the same time, ethnographic insights warn against taking the most "deviant" (exotic, hybrid) style in the field as its default pattern.

Conclusions

Much language-focused CMC research to date is based primarily on log data, and methodological reflection on its combination with other types of data is lagging behind. In this study I argue that an ethnographic framing of the linguistic analysis of log data may bear significant benefits for the contextualization of the data, the interpretation of findings, and in shaping what research questions are asked. In particular, systematic observation and interviews with Internet actors can be used to overview the online field or community under study; to elicit participants' awareness of linguistic heterogeneity; and to alert researchers to emic categories and views, which may act as correctives to their assumptions and interpretations. In this article, evidence was presented of the power of ethnography to uncover the unexpected (Agar, 1995); for instance, the importance of local, German-based practices and influences on the way young hip-hoppers use and perceive variation in English spelling (see example 2); the fact that hip-hop slang is not unanimously accepted, but rather treated as an index of young age and low-status education by some hip-hoppers (see example 4); or the fact that the spoken/written relationship is deemed considerably more complex than just a transfer of spoken features to written texts or a not-further-specified "hybrid" amalgamation of both (see example 5). In all these cases ethnography reveals, as Agar (1995) puts it, that "what appears on the surface means something other than what one initially thinks" (p. 583)—and, I would add, something other than one might think based on log data alone.

However, reflecting on my own methodological choices leads me to emphasise the strength of examining different types of data together: While interviews may offer insights that are not (or only marginally) accessible through systematic observation, observation may disclose aspects of structure that are difficult to elicit in an interview. At the same time, linguistic analysis of log data may contextualize emic views, indicating where participants' distinctions are generalised or biased (see example 1). Thus a combination of "objective" and "subjective" data may offer considerable advantages over a log-data-only research design with respect to

contextualising data, accessing emic perspectives, and enhancing interpretation (i.e., closing interpretation leaks, deepening interpretations, protecting from false interpretations; see Deppermann, 2000).

An important caveat is that the work reported here does not claim to represent a full-fledged ethnography, i.e., an in-depth, long-term study of a specific "virtual community" (as in Baym, 2000). Rather, it adopts an ethnographic perspective and uses elements of ethnographic method in various settings. While I acknowledge that my use of ethnography could be enhanced, I also endorse Hine's suggestions that virtual ethnography remains necessarily partial and is not "quite the real thing" in methodologically purist terms, but rather is an "adaptive ethnography which sets out to suit itself to the conditions in which it finds itself" (Hine, 2000, p. 65). At the same time, the understanding and use of ethnography discussed in this article are consonant with the supplementary role of ethnography in sociolinguistics. To the extent that we as researchers are mainly concerned with the use of language in digital social life (and not with social life as such), maintaining a strong focus on discourse is a legitimate decision; a partial ethnographic engagement may therefore be sufficient for the aims of language-focused research.

Notes

1. "Log data" and "logs" are used in the sense of Herring (2004, p. 339) to refer to "characters, words, utterances, messages, exchanges, threads, archives, etc."
2. The label itself elaborates on "discourse-centred ethnography," a term coined by Haines (1999) in a study of computer-mediated discourse.
3. See a recent theme issue of *Forum Qualitative Social Research* (Domínguez et al., 2007), several articles in *Journal of Computer-Mediated Communication*, and discussion in Baym (2000), Bräuchler (2005), and Cherny (1999).
4. In this process, some researchers opt for participant observation (e.g., Baym, 2000, Cherny, 1999), while others refrain from active participation (e.g., Paolillo, 1999). Interestingly, a clash of traditions might be lurking here, with the ethnographic tradition supporting active participation and the (socio) linguistic tradition rather favouring detachment. On the tension between linguistics and ethnography, see also Rampton (2006).
5. See, e.g., Boehlefeld (1996), Döring (2003), Ess (2007), Herring (1996), papers in Buchanan (Ed.), (2004), and the *International Journal of Internet Research Ethics* (<http://www.uwm.edu/Dept/SOIS/cipr/ijire/>).
6. At the time and place of this fieldwork, i.e., 2000-2002 in Germany, I was not legally required to obtain approval by a research ethics committee or to have my

informants sign a consent form. It was therefore possible to establish contacts with Internet users on a basis of mutual trust, which felt all the more suited for the youth-cultural field in which I was doing research.

7. However, even anonymisation cannot offer bullet-proof protection of privacy as long as verbatim excerpts are used which may be traceable to their context of origin through a purposeful web search.
8. The <a> spelling reflects the vocalization of word-final postvocalic /r/, while <z> is a phonetic spelling, originally restricted to the noun plural marker in postvocalic position (e.g., *boyz*) and then generalised to all instances of plural <s>.
9. Tsiplakou (2009, forthcoming), who also supplements linguistic analysis of email data with evidence from interviews and questionnaires, also finds that her informants' English/Greek code mixing was considerably higher in their emails than in face-to-face interaction.

References

- Agar, M. (1995). Ethnography. In J. Verschueren, J.-O. Östman, & J. Blommaert (Eds.) *Handbook of pragmatics* (pp. 583-590). Amsterdam/Philadelphia: Benjamins.
- Androutsopoulos, J. (2003). Musikszenen im Netz: Felder, Nutzer, Codes. In H. Mertens & J. Zinnecker (Eds.), *Jahrbuch Jugendforschung 3* (pp. 57-82). Opladen: Leske + Budrich.
- Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics*, 10(4), 419-438.
- Androutsopoulos, J. (2007a). Style online: Doing hip-hop on the German-speaking web. In P. Auer (Ed.), *Style and social identities: Alternative approaches to linguistic heterogeneity* (pp. 279-317). Berlin & New York: de Gruyter.
- Androutsopoulos, J. (2007b). Language choice and code-switching in German-based diasporic web forums. In B. Danet & S. C. Herring, (Eds.), *The multilingual Internet. Language, culture, and communication online* (pp. 340-61). New York: Oxford University Press.
- Androutsopoulos, J. (2009). Language and the three spheres of hip hop. In H. Samy Alim, A. Ibrahim, & A. Pennycook (Eds.), *Global linguistic flows. Hip hop cultures, youth identities, and the politics of language* (pp. 43-62). New York/Oxon: Routledge.
- Baym, N. K. (2000). *Tune in, log on: Soaps, fandom, and on-line community*. Thousand Oaks, CA: Sage.
- Boehlefeld, S. P. (1996). Doing the right thing: Ethical cyberspace research. *The Information Society*, 12(2), 141-152.
- Bourdieu, P. (1991). *Language and symbolic power*. Oxford, UK: Polity Press.

- Bräuchler, B. (2005). *Cyberidentities at war. Der Molukkenkonflikt im Internet*. Bielefeld: Transcript.
- Buchanan, E. (Ed.) (2004). *Readings in virtual research ethics: Issues and controversies*. Hershey, PA: Information Science Publishers.
- Cherny, L. (1999). *Conversation and community: Chat in a virtual world*. Stanford, CA: CSLI Publications.
- Crystal, D. (2001). *Language and the Internet*. Cambridge, UK: Cambridge University Press.
- Danet, B. (2001). *Cyberpl@y. Communicating online*. Oxford & New York: Berg.
- Deppermann, A. (2000). Ethnographische Gesprächsanalyse. *Gesprächsforschung*, 1, 96-124. Retrieved August 20, 2008 from <http://www.gespraechsforschung-ozs.de/heft2000/heft2000.htm>
- Domínguez, D., Beaulieu, A., Estalella, A., Gómez, E., Read, R., & Schnettler, B. (Eds.). (2007). *Virtual ethnography*. Thematic issue, *Forum: Qualitative Social Research*, 8(3). Retrieved August 20, 2008 from <http://www.qualitative-research.net/fqs/fqs-e/inhalt3-07-e.htm>
- Döring, N. (2003). *Sozialpsychologie des Internet*. 2nd ed. Göttingen: Hogrefe.
- Eckert, P. (2000). *Linguistic variation as social practice*. Oxford, UK: Blackwell.
- Ess, C. (2007). Internet research ethics. In A. N. Joinson, K. Y. A. McKenna, T. Postmes, & U.-D. Reips (Eds.), *The Oxford handbook of Internet psychology* (pp. 487-502). Oxford, UK: Oxford University Press.
- Ferrara, K., Brunner, H., & Whittemore, G. (1991). Interactive written discourse as an emergent register. *Written Communication*, 8(1), 8-34.
- Greschke, H. M. (2007). Bin ich drin? – Methodologische Reflektionen zur ethnografischen Forschung in einem plurilokalen, computervermittelten Feld. *Forum: Qualitative Social Research*, 8(3). Retrieved August 20, 2008 from <http://www.qualitative-research.net/fqs-texte/3-07/07-3-32-d.htm>
- Haines J. (1999). 'Oi! skins:' Trans-Atlantic gay skinhead discourse on the Internet. *Intercultural Communication*, 1. Retrieved August 20, 2008 from <http://www.immi.se/intercultural/nr1/haines.htm>
- Hawisher, G. E., & Selfe, C. L. (Eds.) (2000). *Global literacies and the World Wide Web*. London and New York: Routledge.
- Herring, S. C. (1996). Linguistic and critical research on computer-mediated communication: Some ethical and scholarly considerations. *The Information Society*, 12(2), 153-168.
- Herring, S. C. (2004). Computer-mediated discourse analysis: An approach to researching online communities. In S. A. Barab, R. Kling, & J. H. Gray (Eds.), *Designing for virtual communities in the service of learning* (pp. 338-376). Cambridge & New York: Cambridge University Press.
- Hine, C. (2000). *Virtual ethnography*. London: Sage.

- Howard, P. N. (2002). Network ethnography and the hypermedia organization: New media, new organizations, new methods. *New Media & Society*, 4(4), 550-574.
- Hymes, D. (1996). *Ethnography, linguistics, narrative inequality: Toward an understanding of voice*. London: Taylor & Francis.
- Kozinets, R. V. (2002). The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research*, 39, 61-72.
- McLelland, M. J. (2002). Virtual ethnography: Using the Internet to study gay culture in Japan. *Sexualities*, 5(4), 387-406.
- Miller, D., & Slater, D. (2000). *The Internet: An ethnographic approach*. Oxford & New York: Berg.
- Niedzielski, N. A., & Preston, D. R. (2000). *Folk linguistics*. Berlin: Mouton de Gruyter.
- Paolillo, J. C. (1999). The virtual speech community: Social network and language variation on IRC. *Journal of Computer-Mediated Communication*, 4(4). Retrieved August 20, 2008 from <http://jcmc.indiana.edu/vol4/issue4/paolillo.html>
- Puri, A. (2007). The Web of insights—The art and practice of webnography. *International Journal of Market Research*, 49(3), 387-408.
- Rampton, B. (2006). *Language in late modernity. Interaction in an urban school*. Cambridge: Cambridge University Press.
- Saville-Troike, M. (2003). *The ethnography of communication: An introduction*. 3rd edition. Malden, MA: Blackwell.
- Tsiplakou S. (2009, forthcoming). Acting bilingual: Code-switching as performative construction of online identities. *Pragmatics*. Special issue on *Language, discourse and identities. Snapshots from Greek contexts*. Ed. by A. Georgakopoulou & V. Lytra.
- Yang, G. (2003). The Internet and the rise of a transnational Chinese cultural sphere. *Media, Culture & Society*, 25, 469-490.

Biographical Note

Jannis Androutsopoulos [jannis.androutsopoulos@kcl.ac.uk] is a reader in sociolinguistics and media discourse in the Department of Education & Professional Studies, King's College London. His research interests include computer-mediated communication, language in popular culture, youth and youth cultures, sociolinguistic aspects of orthography, and linguistic diversity in media discourse.