Introduction: Data and Methods in Computer-Mediated Discourse Analysis

Jannis Androutsopoulos
King's College London

Michael Beißwenger
TU Dortmund University

urn:nbn:de: 0009-7-16090

Background

Originating in a section by the same name at the 10th International Pragmatics Conference held July 2007 in Gothenburg, Sweden, this collection of articles aims at addressing what we consider a less-attended to area of computer-mediated discourse studies: the need for methodological reflection on data collection and analysis.

Computer-mediated discourse (CMD) encompasses all kinds of interpersonal communication carried out on the Internet, e.g., by email, instant messaging, web discussion boards, and chat channels (Herring, 2001, 2004). In the last decade, CMD has attracted a great deal of research attention from linguistic—especially pragmatic, discourse-analytic, and sociolinguistic—perspectives. However, methodological reflection is lagging behind compared to other areas of discourse studies. To begin with, while data collection on the Internet seems trivial at first sight, researchers conducting CMD studies are confronted with a variety of non-trivial questions. These may relate to the size and representativeness of data samples, data processing techniques, the delimitation of genres, and the kind and amount of contextual information that is necessary, as well as to ethical issues such as anonymity and privacy protection. Much research in the area has been based on small, ad-hoc data sets; there is a lack of standard guidelines for CMD corpus design and a lack of publicly-available CMD corpora (Beißwenger & Storrer, 2008).

In terms of methodology, language-focused research on CMD has drawn on methods and key concepts from a variety of research traditions in linguistics (including pragmatics, conversation analysis, sociolinguistics, genre analysis, and the ethnography of communication), which have been applied fruitfully to study how individuals use linguistic resources to establish contacts, manage interactions, and construct identities within computer networks. What is

largely lacking, however, is critical reflection on the problems and challenges that arise when these research traditions are applied to the new settings and environments of CMD. For example, does a one-to-one transfer of research frameworks lead to contextually rich understandings of language use and interactional processes in CMD, or does it rather conceal some of its essential new aspects? Research findings suggest that CMD has important implications for understanding key concepts in discourse studies, such as interactional coherence, participant frameworks, intertextuality, language-identity relationships, and the notion of community. Adapting or reconceptualizing existing concepts and methods seems a necessary step in the further development of CMD studies, and new research frameworks are already emerging, such as Herring's approach to the study of online communities (Herring, 2004).

Research questions that arise against this backdrop relate, for example, to the following issues:

- The compilation and design of CMD corpora, including the acquisition, pre-processing, and annotation of data and meta-data, tools for corpus storage and maintenance, and ethical issues;
- Benefits and challenges of the combination of various data sets (e.g., log files, online observation, user interviews) for specific research questions;
- Ways of doing “online” or “virtual” ethnography as a contextually rich window into the study of online activities and online communities;
- Social and technical conditions of CMD that need to be taken into account when adapting established concepts and frameworks to the analysis of online discourse; and
- Combining qualitative and quantitative methods in CMD research.

Contents of the Special Issue

The contributions to this special issue address various aspects of this research agenda; we highlight here some of their common threads and trajectories. In a nutshell, these relate to 1) the use and combination of innovative types of data; 2) the application and extension of Herring's (2004) CMDA framework; 3) the critical examination of established methodologies; and 4) the combination of quantitative and qualitative approaches.

In terms of data, several articles in this issue emphasize the need to go "beyond the screen," i.e., to extend the present research focus on log file data. *Marcoccia, Atifi and Gauducheau*, as well as *Beißwenger,* contribute to the study of multimodality in the production of CMD, a strand of research still in its infancy, by arguing that the study of video recordings of participants' gaze and body movements and screen recordings of the production of text entries can provide new insights into how new media users design and experience their participation in online interaction. Likewise, *Androutsopoulos* discusses the combination of log files and ethnographic data such as systematic observation and interviews, and suggests that ethnographic insights may work as a corrective to the limitations of log-based interpretations. *Siebenhaar* offers an example of the use of large data sets (9 million chat messages with 41.7 million words from 13 Swiss IRC channels over seven years) as a basis for both quantitative and qualitative analyses. *Holmer* outlines a tool and methodology for the "discourse structure analysis" of chat data, which combines coding the relations between chat messages with the software-aided visualization of threads and retrieval of statistical metrics from coded CMD data. One goal of this approach is to provide a tool for qualitative CMD analyses by allowing researchers to represent chat discourse structure in a more vivid way than the usual linear list representation of log files.

A second meeting point of several papers is their use of Herring's (2004) computer-mediated discourse analysis (CMDA) framework. Being perhaps the most explicit and fully-articulated framework in the field today, it invites other researchers to put it to work, explore its applicability, but also to extend it and take some of its aspects further. Thus *Nishimura* elaborates on Herring's theorizing of politeness by drawing on Japanese data and politeness theory; *Stommel* argues for the addition of interaction analysis and the concept of communities of practice (Wenger, 1998) to the study of online communities; and *Androutsopoulos* suggests extending the framework's grounding on logs of verbal interaction to include ethnographic evidence.

A third point of concern is to what extent a transfer of methodological paradigms developed for "traditional" verbal interaction may be suitable for the study of CMD. *Androutsopoulos* discusses the transfer of principles and techniques of ethnography to online environments and new media users, and *Stommel* draws on Conversation Analysis (CA) for the study of identities and relationships in a discussion forum. In contrast, *Beißwenger* critically questions the applicability of conversation-analytic categories such as turn-taking to the analysis of chat conversation. Thus CA offers a case in point for

the debate on the transferability of an existing methodological paradigm to online discourse analysis. These articles suggest that the motivation for, and the usefulness of, such a transfer will vary depending on the discourse sites examined and the questions asked by different online researchers.

A further point of convergence is the relationship between quantitative and qualitative approaches, which CMD is well equipped to bridge and strengthen (Georgakopoulou, 2006). This direction is most visible here in the papers by *Nishimura* and *Siebenhaar. Nishimura's* comparative analysis of two Japanese bulletin boards combines a quantitative analysis of politeness features with a qualitative analysis of speech acts (thanks, insults), which are taken as indicators of *online community* based on Herring's framework. *Siebenhaar's* paper discusses implications of quantitative research for qualitative analysis and suggests that quantitative analyses may provide a backdrop and guidelines for qualitative research. Due to the variety of online discourse spaces and the immense fluctuation of participation patterns within them, *Siebenhaar* argues, the exact slice of data chosen for CMD analysis will crucially shape the findings. A quantitative analysis step may help qualitative researchers in selecting their sample and avoiding false conclusions.

A similar point is made by *Beißwenger* and *Marcoccia, Atifi, and Gauducheau* when they point out that the structure of chat communication will be different depending on whether the researcher's focus is on screen discourse only or also involves users' physical (corporeal) actions or gaze movements. In these articles, the tension between text-centred and multimodal (or "situated") analysis is resolved in favour of the latter, thereby echoing the argument by *Androutsopoulos* with respect to the contrast between log-based and ethnographically enhanced research.

In terms of language coverage, the articles in this special issue analyze French, German, and Japanese data. In terms of communication modes (or formats), e-chat gets the lion's share of attention (*Beißwenger; Holmer; Marcoccia et al.; Siebenhaar*), followed by forums/bulletin boards (*Nishimura; Stommel*), while one article is based on user interviews rather than online discourse data (*Androutsopoulos*).

Future Directions

Discussion of data and methods in CMDA has just begun, and language-focused researchers will need to align their methodologies with the continuing evolution and social spread of digital technologies. This collection is an initial

contribution toward that goal. A useful next step would be to examine methodological differences across synchronous and asynchronous communication modes, and to consider whether different social uses of the same technology can be examined with the same methodology. Examples of research questions that could be addressed in future work include the following:

- Comparing virtual communities: One might consider the cases discussed in this issue (Stommel: a forum on eating disorders; Nishimura: discussion boards on film and language) alongside other types of online community, such as a social chat environment with a strong core of regular users or a MMORPG, i.e., a 'massively multiplayer online role-playing game', with frequently changing users. What are the commonalities and differences among different types of online communities, and to what extent do different community types call for distinct analytic categories?
- Multi-platform communication: People often use multiple digital communication tools in order to pursue their communicative tasks, e.g., to exchange opinions, seek advice, chat with friends, or play a game online. What types of data and fieldwork methods are needed for studying such complex, but increasingly common, forms of CMD?
- Different social uses of CMC technologies: Comparative studies are needed to examine the use of a CMC technology in different social situations: For instance, participation in an informal social chat implies quite different social relations than does participation in a chat implemented for academic learning and teaching, in which interpersonal relations are institutionally predefined. What are the impacts of such differences on the interactional structure and the linguistic resources used in each case? How might tools such as Holmer's *Chatline* be implemented in such research?
- Research on "Web 2.0": Platforms for social networking (e.g., Facebook, MySpace), content sharing (e.g., Flickr, YouTube) and collaborative authoring (e.g., Wikipedia, Wictionary) have received little attention from language-oriented scholars so far, with the exception of blogs (see, e.g., Herring, Scheidt, Bonus, & Wright, 2004; Huffaker & Calvert, 2005). Much remains to be done here, and the implications of these new computer-mediated discourse environments for data collection and methodology need to be assessed.

Acknowledgements

We're grateful to all the contributors to this issue, who went with us through a series of revisions with accuracy and speed; to the anonymous reviewers of all the articles for their helpful feedback; and especially to Susan Herring for the invaluable support and critique with which she accompanied the making of this issue all the way. We are, of course, solely responsible for any shortcomings that remain

References

Beißwenger, M., & Storrer, A. (2008). Corpora of computer-mediated communication. In A. Lüdeling & M. Kytö (Eds), *Corpus linguistics. An international handbook. Volume 1.* Berlin/New York: de Gruyter (Handbooks of Linguistics and Communication Sciences 29.1).

Georgakopoulou, A. (2006). Postscript: Computer-mediated communication in sociolinguistics. *Journal of Sociolinguistics, 10*(4), 548-557.

Herring, S. C. (2001). Computer-mediated discourse. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *Handbook of discourse analysis* (pp. 612-634). Oxford: Blackwell.

Herring, S. C. (2004). Computer-mediated discourse analysis: An approach to researching online behavior. In S. A. Barab, R. Kling, & J. Gray (Eds.), *Designing for virtual communities in the service of learning* (pp. 338-376). Cambridge/New York: Cambridge University Press.

Herring, S., Scheidt, L. A., Bonus, B., & Wright, E. (2004). Bridging the gap: A genre analysis of weblogs. *Proceedings of the Thirty-Seventh Hawai'i International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society Press. Retrieved September 2, 2008 from http://www.blogninja.com/DDGDD04.doc

Huffaker, D. A., & Calvert, S. L. (2005). Gender, identity, and language use in teenage blogs. *Journal of Computer-Mediated Communication, 10*(2), article 1. Retrieved September 2, 2008 from http://jcmc.indiana.edu/vol10/issue2/huffaker.html

Wenger, E. (1998). *Communities of practice; learning, meaning and identity*. Cambridge: Cambridge University Press.

Biographical Notes

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.

Michael Beißwenger [michael.beisswenger@uni-dortmund.de] is a research assistant and lecturer in German linguistics in the Department of Cultural Sciences at TU Dortmund University. His research interests include computer-mediated communication, lexicology, orthography, and corpus linguistics.