Language and communication in computer-mediated contexts: A rich and challenging research field

Miriam A. Locher

1. Introduction

This special issue offers an invitation to think about the first decade of the 21st century. Acknowledging that millennia boundaries are always arbitrary such that any development is likely to have started earlier and to still be ongoing, I decided to write about the advent of computer-mediated communication (CMC) as one of the linguistic areas that are topical and prominent in today's world. In recent overview articles on 'electronic discourse' (Locher 2014; Locher and Mondada 2014) and on challenges in research methodology for scholars working on computer-mediated discourse (Bolander and Locher 2014), a number of observations were made that I want to revisit in light of the aim of this special issue. Given the fact that communication through computer-mediated means is a fairly recent phenomenon and that the vast majority of research on this use of language has been published since the turn of the millennium, we can indeed state that this topic is relevant. In line with Crystal, "[t]he Internet is the largest area of language development we have seen in our lifetimes. Only two things are certain: it is not going to go away, and it is going to get larger" (2011, 149). In this chapter, I will briefly sketch the object of study, show some of the research questions and developments that have occurred and discuss some of the ways in which linguists have responded to language online.
2. Computer-mediated communication as a rapidly developing form of communication

Those of us who still remember their Commodore 64, one of the first affordable home computers in the 1980s, will appreciate to what extent access to computer-mediated communication has generally increased in the last three decades.¹ Not only has the computer clearly conquered the office with text creation and calculation programs, but the advent of the Internet with its possibilities of providing information and services (from online shopping to entertainment) and the fairly recent addition of social network sites has become a phenomenon that is hard to dismiss in our daily lives. The continual development of new technology, such as fully online accessible smartphones and tablets, gives people continual and instant access to different types of information. At the same time, people not only consume language but also produce it. It is this latter fact that I want to focus on in this chapter, i.e. that we are confronted with a corpus of written and multi-modal data that linguists can explore for their research.

To describe language on the web, scholars have used a number of alternative labels, such as 'digital discourse,' 'electronic discourse,' 'e-communication,' 'digitally mediated communication,' 'computer-mediated communication,' 'computer-mediated discourse' or 'keyboard-to-screen communication' (for an overview, see Crystal 2011; Jucker and Dürscheid 2012). Crystal ([2001] 2006) published one of the first monographs that dealt with language in computer-mediated contexts and proposed Netspeak as an umbrella term to describe the different linguistic practices observed. However, in recent times, he and others have dropped this concept to do justice to the fact that the use of language on the web is so diverse: "It is difficult to find linguistic generalizations that apply comfortably to Internet language as a whole" (Crystal 2011, 10).

There are a number of reasons for this diversity. On the one hand, technology has constantly developed new programs to make new ways of communication possible. While the early use of the Internet was largely geared to providing information, the communicative aspect was nevertheless important from the beginning in the form of email, instant messenger, chat and language-based games (MUDs; MOOs; see Crystal [2001] 2006). The 2000s have seen a strengthening of participatory options, so that we find an increase in blogging, responding to blogs and newspaper articles, sharing videos and pictures and commenting on these in platforms such as YouTube or flickr, joining social network sites such as MySpace or Facebook, engaging in virtual worlds such as Second Life, or sharing news via Twitter, etc. This latter development has been termed Web 2.0. In Zappavigna's words:

The social web, or Web 2.0, are popularized terms used to signal a shift toward the internet as an interpersonal resource rather than solely an information network. In other words, the social web is about using the internet to enact relationships rather than simply share information, although the two functions are clearly interconnected. (2012, 2)

¹ However, access to CMC affordances (cell phones, Internet, computers, etc.) is not equally distributed (cf., e.g., the literature on the 'digital divide' in America, as reported at http://www.pewinternet.org/topics/digital-divide/ or globally as discussed in Guillén and Suárez 2005).
Web 2.0 is also described as participatory, dynamic, no longer restricted to specialists but open to everyone and no longer tied to a computer but also accessible via other electronic devices (see Zappavigna 2012, 2; Locher 2014, 558). In addition to the increase in multi-modal communication, there is also an increase in platforms where formerly separate communication forms are provided within one and the same interface (Thurlow and Mroczek 2011; Yus 2011; Herring 2013; Jucker and Dürscheid 2012; Locher 2014). Facebook, for example, allows written interaction in the form of sharing status updates, commenting on them and engaging in private messaging and chat communication, while also providing the option of uploading and sharing pictures and videos (for an illustration see Locher 2014). In other words, the sheer number of possibilities by means of which a person can engage in online communication has dramatically increased.

On the other hand, it is not enough to simply point to the fact that technology has provided new options for communication. This view would dramatically downplay the role of the users, who are constantly adopting and adapting the new technology for their own communicative purposes. As we can see with the use of email, it was a combination of technical developments but also a need generated bottom-up that turned this form of communication from a purely content-oriented means to quickly and efficiently exchange factual information to a dynamic and everyday tool of communication that is used for formal and informal, personal and business interaction (see Locher 2014). Comparing early computer-mediated language practices with today's use of the same medium might therefore often be difficult because both technological as well as functional parameters will have changed. Crystal (2011, 10) refers to this "speed of change" as one of three research challenges for scholars aiming to study language use online (the other two being the "rapidly growing language corpus" and the "diversity of language encountered on the Internet").

Looking back over the last few decades of computer-mediated communication, we can therefore state that people have appropriated the technological means for their own purposes and have creatively helped to give direction to further technological developments. The first decade of the 21st century, especially, has witnessed the large increase of participatory platforms so that linguists are confronted more than ever with data waiting to be studied. In what follows, I will elaborate eclectically on a number of research issues that have transpired in recent times.

### 3. Research directions: digital humanities and linguistics

One of the recent 21st-century buzzwords is 'digital humanities,' which, according to the UCLA website of a program in digital humanities, is defined as follows:

> Digital Humanities interprets the cultural and social impact of new media and information technologies – the fundamental components of the new information age – as well as creates and applies these technologies to answer cultural, social, historical, and philological questions, both those traditionally conceived and those only enabled by new technologies. (UCLA Center for Digital Humanities)
In other words, this field is fuzzy by definition and combines both the study of the use and development of digital technology, as well as the exploration of cultural practices that are enabled by computer-mediated means in the first place. Scholars who are studying computer-mediated practices in their own right (communication studies, discourse analysis, media studies, etc.) exist alongside those who are using the recently developed digital tools for archiving (e.g. digitalizing material that does not necessarily originate from the Internet) and data mining these sources to approach their research questions (e.g. literary and cultural studies; history; the humanities in general).

3.1 Research areas

With respect to the study of computer-mediated communication practices, linguists have approached a number of issues as identified in recent overview work by, for example, Androutsopoulos (2006; 2008), Crystal ([2001] 2006; 2011), Herring (2007; 2013), Herring, Stein and Virtanen (2013), Locher (2014), Locher and Mondada (2014), and Thurlow and Mroczek (2011). Work on computer-mediated communication can be split into a number of general research areas:2

(a) developing a theoretical framework for CMC contexts by providing tools for the description of data and interaction formats;
(b) exploring particular modes of CMC, such as chat, blogs, e-mail, etc.;
(c) focusing on the management of interactional organisation (e.g. turn-taking, etc.) in CMC;
(d) studying activities (e.g. disagreeing, apologising, requesting, etc.) within CMC contexts;
(e) focusing on the complex emergence of situated interpersonal language use (e.g. the study of identity and group membership, etc.);
(f) exploring the interweaving of online and offline interaction;
(g) studying general language change and maintenance by taking CMC into account;
(h) addressing the challenges of literacy and computer literacy.

Focus (a) is about critically reflecting on research findings on CMC and the wish to synthesize the findings into a theoretical perspective. For example, the early discussion on the written form of oral features such as the description of false starts, self-corrections, disruption, truncation, etc. as observed in chatrooms ultimately aimed at fitting the observed phenomena into an already established theoretical idea of how written and spoken language use works. Scholars such as Herring (2007) then proposed to look at CMC in its own right and proposed social as well as technical parameters to

---

2 Due to the brevity of this contribution, the individual points cannot be backed up with extensive references. Please consult Locher (2014) and the indicated overview work for references on some of the issues addressed.
describe computer-mediated practices.\textsuperscript{3} Giving importance to both sets of parameters avoids giving predominance to the technical affordances (an early tendency in research that was criticized as technological determinism; for a discussion, see, e.g., Androutsopoulos 2006; Baym 1995; Herring, Stein and Virtanen 2013). Scholars working on CMC are thus open to developing new methodologies that are adapted to online contexts while clearly drawing on existing insights and theories (e.g. Herring’s social parameters are based on Hymes’ SPEAKING mnemonic).

Focus (b) draws attention to a continually growing body of literature on particular CMC modes, such as chat, blogs, e-mail, instant messaging, listserv, websites, wikis, tweets, Facebook status updates, interactive online games, etc. This body of work is growing both because of increased interest in different CMC practices by scholars and because new modes are continually being invented or old ones changed. To use blogs as a case in point: they have been around since the 1990s, but monographs on blogging in linguistics only started to appear in the 2000s (e.g. Smyk-Bhattacharjee 2009; Hoffmann 2012; Bolander 2013). As Crystal points out, the software enabling blogging and commenting on blog entries may have changed again in the meantime, so that the "speed of change" (Crystal 2011, 10) might make it impossible to re-produce similar studies or to retrieve the original data. For example, the spread of the “reply” option in blogs, which allows commentators to respond to other people’s comments, is a relatively recent addition, which changes the participatory framework of blogs by making them more akin to forms of CMC in which polylogues are common.\textsuperscript{4} As already outlined above, social network platforms also often combine previously separate modes of interaction so that new synergies in communication strategies might emerge. While stressing the novel aspect of many interaction formats is certainly deserved, it is nevertheless worthwhile pointing out that many text practices on the internet have their predecessors in literary and non-literary print and face-to-face genres, which by definition have fuzzy boundaries and are characterized by intertextuality (see Busse 2014a; 2014b). For example, online health advice columns clearly have their roots in the printed newspaper and magazine versions (see Locher 2006).

Focus (c) is concerned with the interactional organisation of computer-mediated practices, such as with how interactants create coherence, manage turn-taking, use code-switching, etc. The interest lies in establishing how interactants engage with each other in the different modes and how they deal with the technological affordances that might be perceived as both restricting and empowering. For example, chat allows quasi-

\textsuperscript{3} Herring (2007) proposes a “faceted classification scheme,” which is open-ended. The described situational/social and medium/technological factors both shape CMC practices, so that giving the medium/technological factors precedence is explicitly avoided. The medium/technological factors comprise aspects of synchronicity, message transmission, the persistence of transcript, the size of message buffer, the channels of communication, the possibility for anonymous messaging, private messaging, filtering, quoting, and the message format. The set of situational/social factors consist of the participation structure, the participant characteristics, the purpose, topic or theme, tone, activity of the interaction, the norms developed and invoked, and the code.

\textsuperscript{4} Thank you to Brook Bolander for providing this example.
synchronous written communication that might pose challenges with respect to lag, turn disruption and the problem of establishing coherence in multiparty conversations. On the other hand, next to the fact that users have developed turn-holding strategies (such as posting incomplete sentences so that the addressee knows that more is to come) etc., chat programs often provide options for using emoticons or sharing pictures and videos, which create new (multi-modal) communication possibilities (see, e.g., Beisswenger 2000; 2001).

Focus (d) builds on the findings about interactional organization and expands our interest to the ways in which particular activities are performed online, e.g. disagreeing, apologizing, requesting, advice-giving, etc. In other words, the use of language from a content and functional perspective gains center stage. Research on health practices online can serve as an example since the Internet provides a vast resource of health information for lay people who turn to professional and lay sites for their research. Next to purely looking for information, we also find many instances of advice and support seeking/providing. Studies have explored online health advice for young adults and teenagers provided by professional health experts (e.g. Locher 2006; 2010; Harvey 2013; Harvey et al. 2013) or peer-to-peer online fora (Harrison and Barlow 2009; Kouper 2010).

Research focus (e) builds on research interests that are well established in face-to-face interaction but enlarges the scope to include computer-mediated communication. Here we can name studies on the emergence of situated interpersonal language use, i.e. the study of identity construction and group membership, style and (im)politeness, etc. In other words, people appropriate computer-mediated means to not just search and pass on information but to engage in community building, facework and identity construction (e.g. Baym 1995; 1998; Herring 2004; Bedijs et al. 2014). As a case in point, we can once again mention research on health communication. In the project Language and Heath Online, funded by the Swiss National Science Foundation, we explore how relational work (i.e. “the work invested by individuals in the construction, maintenance, reproduction and transformation of interpersonal relationships among those engaged in social practice;” Locher and Watts 2008, 96) manifests itself in email counselling where the client and counsellor strive to create a therapeutic alliance and how relational work informs acts of persuasion in professional and lay sites on smoking cessation. Other areas of interest in this field are the construction of identity in social media, such as in status updates in Facebook, microblogging in tweets and online narratives (see, e.g., Bolander and Locher 2010; 2015; Locher and Bolander 2014; 2015; Seargeant and Tagg 2014; Zappavigna 2012; 2014; Page 2012).

Focus (f) refers to insights by scholars who propose that CMC practices should not be studied in isolation but should include an exploration of their connections to other online and offline practices (see, e.g., Androutsopoulos 2008; Miller and Slater 2000). As we cannot separate offline from online interaction in a non-problematic way and since CMC

---

5 More information on this project can be found here: https://language-health-online.unibas.ch/?Home (22 February 2015).

is as real to its users as face-to-face interaction (Baym 1998), there are studies that explore how people multi-task and how they draw on different communication affordances during their activities (e.g. Lee 2011; Jones et al. 2011; Georgakopoulou 2013). In other words, how practices are organized and how acts are performed is not only of interest within computer-mediated communication modes but also across online and offline contexts. As a case in point, Mondada (2012) looks at online gamers and films them in their physical environment to study how people simultaneously draw on CMC and face-to-face modes for their communicative purposes.

Research area (g) is concerned with the influence of CMC interaction on language change and language maintenance in general. In other words, we can ask to what extent and in what ways CMC affordances might influence language development beyond CMC contexts. First, the topic of whether CMC practices might influence offline language practices was tabled early on and has also often found reflections in the discourse of the general public, which, for example, worried about deteriorating language standards (see, e.g. Baron 1984; 2000). Second, turning to the particular case of migration and language change, we can observe that CMC has dramatically changed the situation of migrants who live in the diaspora. Simplifying the picture quite consciously: in the not so distant past, people who moved away from their home had to rely on letters sent by postal services via carriages, trains, boats and airplanes, and were often without any news from their family and friends for many months. Moving back and forth between the new location and the homeland was often neither practically possible nor affordable. While there may have been people who spoke the same mother tongue as the recent migrants, the diaspora might have been rather small. Today, cheaper costs for international phone calls, internet cafés, cheaper access to CMC devices and programs such as skype that include video broadcasting will make it easier for many migrants to stay in touch with their home community and to keep up with changes that occur in the home language during their absence. Cheaper airfares (to/from some areas of the world) might also allow for more home visits. In other words, there is both more mobility and more opportunity for maintaining links with the homeland and its languages. Migrants draw on the Internet as a rich resource to keep using their native tongues and to stay in touch, for example, by reading online newspapers, reading and writing blogs or participating in fora, etc. (see Jacquemet 2013 on digital transnational spaces). It remains to be seen whether these developments will have any long-term influence on the maintenance of first languages in the diaspora. Migrants and locals thus use the Internet for their own purposes in many different languages. This multilingual Internet has indeed started to receive the attention it deserves (e.g., Androutsopoulos 2010; 2013a; Danet and Herring 2007; Mair 2013). For example, studies explore how previously non-written or rarely written forms of dialect and

6 While the sociolinguistics literature gives precedence to face-to-face oral communication for the locus of language change and transmission, this exclusiveness is up for discussion nowadays (see, e.g., Gao 2008 on Internet influence; Sayers 2014 on TV and film). Jacquemet argues that “digital communication is now a home base for the development of language for a broad section of the general population” (2014, 478).

7 Major changes in migration patterns are often subsumed under the buzzword “super-diversity” (Vertovec 2007; cf., also Arnaut et al. 2012; Blommaert and Rampton 2011; Blommaert et al. 2012).
patois are used online (e.g. Siebenhaar 2003 on some dialects of Swiss German) and "how the new media impact on the spread of standard and nonstandard varieties of European ex-colonial languages in conditions of economic, political, cultural and media globalization" (Mair 2013: 17).

Focus (h) tries to capture the important point that much of CMC still appears in the written form despite recent multi-modal developments. This means that CMC users have to be literate in the sense of being able to read and write and, in addition, they have to develop particular computer savviness. While many members of the younger generation grow up familiar with CMC technology (cf. the discussion around the concept of the 'digital native;' Prensky 2001), older people have to make an effort to learn how to use it. As a case in point, nowadays many businesses in Switzerland pressure people into making use of the new technological affordances. For example, many banks punish people with processing fees when they do not make use of online banking programs.\(^8\) This ultimately penalizes those who are not computer savvy or who do not want to take part in this process for privacy and security reasons. Since writing and reading literacy as such are not simply a given (cf. work on adult illiteracy by, e.g., Eme 2011), people who struggle with literacy and/or computer literacy are clearly at a disadvantage. Linking focus (h) with the questions of migration raised above, linguists can explore the language use of many second-generation migrants who might have oral skills in the first language of their parents but lack the corresponding literacy skills as they have never undergone formal schooling in their home language. Active participation in, for example, written online fora in their parents' language might therefore out them as non-proficient or non-standard language users and might stigmatize them. On the other hand, as pointed out above, the Internet also provides migrants and locals alike with the opportunities for writing in creoles, dialects and patois for which no standard written forms exist so that computer-mediated forms of writing can also be empowering and contribute to the maintenance of language diversity.

3.2 Methodological challenges

Scholars working on the types of research subsumed under (a) through (h) also face challenges with respect to methodology. Of particular note are (i) to (k), which are not unique to computer-mediated communication but which have to be addressed when engaging in research with computer-mediated data (see also Bolander and Locher 2014):

(i) using language on the internet to conduct localized case studies and/or for big data mining;
(j) combining methodologies and embracing the challenge of multi-modal communication in online and offline settings;

---

\(^8\) The online banking programs of many banks in turn require the clients to have a cell phone so that secure access codes can be sent which need to be entered into the online login form.
(k) addressing the ethical challenges of CMC research.

Focus (i) shows the scope in which scholars explore computer-mediated language. There are many studies which look at small CMC data sets or custom-made corpora for linguistic purposes (such as the study of turn-taking, code-switching, identity construction, community creation, etc.). Within ethnographic studies, the researchers are often involved in the CMC community itself while at other times they merely act as observers (see Androutsopoulos 2013b; Markham 1998). In addition, there are also projects that move away from studying communication between specific (groups of) interlocutors and instead work with large databases where the individual contributors are no longer known or traceable (e.g. Mair 2013 on compiling data from online fora in West Africa and the Caribbean to explore English on the web). Linguistics has of course a long-standing tradition of working with digitalized and marked-up corpora such as the British National Corpus and its predecessors, which compile English texts of a particular period and make them available for exploration with a linguistic lens. This knowledge of working with large amounts of texts tagging language and annotating it now comes in handy in the new projects that use the language stored on the Internet as a source for research (see, e.g., Hundt et al 2007; Beisswenger and Storrer 2008). The latter quantitative orientation also profits from research in other disciplines within the digital humanities, such as distant reading proposed by Moretti (2013) in literary studies, and is constantly developing new technological means of visualizing and exploring data (see, e.g., Bubenhofer 2014; Jannidis and Lauer 2014).9

As mentioned above, computer-mediated discourse is increasingly multimodal so that scholars have to rise to the challenge of (j) taking different modes of communication into account that might be made available within the same communication platforms (see Herring 2013). In addition, interactants often multi-task and switch back and forth between forms of communication (see, e.g., Lee 2011; Georgakopoulou 2013; Jones et al. 2011). Yet again we find evidence that keeping CMC and face-to-face communication conceptually apart as if they had nothing to do with each other is not fruitful (see research focus f above). Scholars have started to address the resulting methodological challenges by creatively mixing methodologies (for an overview, see Bolander and Locher 2014; Locher and Mondada 2014).

Finally, focus (k) refers to the obvious fact that research ethics cannot be suspended because we are dealing with data from CMC sources. Since the ethical situation of obtaining data from the internet is often unclear and further complicated because of copyright and authorship considerations, scholars have to be especially careful when dealing with CMC data so as not to infringe the rights of the people who provide the language that we study. There are no universal rules that can be followed, but a

---

9 The focus of this chapter is on the language produced and created for communication within a computer-mediated environment. In contrast, many scholars working within the digital humanities in different disciplines often explore data which originally either existed in print or oral form and which was subsequently prepared to appear in digitalized form (such as collections of literary works, digitalized historical archives, etc.).
Community of Internet scholars is continually working on providing guidelines that scholars can draw on when trying to make decisions about issues including consent, anonymity, harm, rights and obligations, and refer to when explaining their decisions to include or exclude particular data (e.g. Markham et al. 2012; Wawra 2014).

4. Conclusions

When looking at the issues raised in this chapter, it becomes evident that computer-mediated communication is not a phenomenon that occurs in isolation. Instead, while legitimately proposing concepts to do justice to the technical affordances that shape CMC to some extent, much recent scholarship highlights that it is ultimately the same individual human beings who are communicating face-to-face and in traditional print forms who also make use of computer-mediated communication tools for their communicative needs and aims. For linguists interested in synchronous language use and present-day linguistic developments, forms of computer-mediated communication provide a rich source of inspiration. This advent of new and exciting language use available for study goes hand-in-hand with technological developments for researching language, so that nowadays there is also an excellent choice of programs which facilitate quantitative and qualitative research (e.g., AntConc, Wordsmith or Atlas.Ti, Ethnograph, ELAN, HyperRESEARCH, Kwalitan, MAXQDA, and NVivo). The recent communication developments in the late 20th and early 21st century discussed in this chapter will keep the linguistics community on their toes for many more years to come, and it will be exciting to witness how the interconnectedness between CMC practices and face-to-face interaction develops and pans out for interactants, and what this implies for research.

Acknowledgments

Many thanks to Brook Bolander for her helpful and critical reading of a first draft of this chapter.

Works cited


Hundt, Marianne, Nesselhauf, Nadja, and Biewer, Carolin, eds.. *Corpus Linguistics and the Web*. Amsterdam: Rodopi, 2007.


