

Herring, S. C. (In press, 2008). Language and the Internet. In: W. Donsbach (Ed.), *International Encyclopedia of Communication*. Blackwell Publishers.

Language and the Internet

Language and (or 'on') the Internet refers to human language (or language intended to be human-like, such as the linguistic output of artificial intelligence agents) produced and displayed through computer-mediated communication (CMC) systems that are mostly text-based and mostly reciprocally interactive, such as email, listserv lists, newsgroups, chat, instant messaging, text messaging via mobile phones (SMS), blogs, and wikis. The term Internet language is somewhat of a misnomer, in that some of this communication takes place on intranets and some is mediated by mobile technologies, rather than by the global networked infrastructure known as the Internet per se (--> Internet, Technology of). "Internet" is used here in an extended sense to include these related communication technologies. Variant terms for Internet language include computer-mediated language, computer-mediated discourse, online discourse, and electronic discourse (--> Technologically-Mediated Discourse). All of these are intended to distinguish language- and discourse-related phenomena as a focus of interest from the broader phenomenon of computer-mediated communication, of which they form a part.

Intellectual and Social Context

Language in the form of typed text is one of the most pervasive and visible manifestations of Internet use. Internet language has attracted the interest of scholars, educators, and the general public, and has been at the center of controversies in each domain. Linguists have argued about how it should be classified; ethnographers and ethnomethodologists have grappled with how to apply the methods of their disciplines to it; and sociologists and communication scholars have debated the status of linguistically-performed online identities (--> Identities and Discourse) and of virtual communities whose group identity is constructed and maintained through online discourse (--> Virtual Communities). Educators and language purists have expressed concern about the non-standard, informal nature of the language found in chat rooms and text messaging, and have wondered whether and how to teach the new mediated language varieties. And despite warnings in the mass media about deception in online dating sites (--> Deceptive Message Production) and the risks of revealing too much information about oneself in other online social contexts, people frequent such sites in large numbers, managing their self-presentation through language and images.

The concerns of scholars are mainly descriptive—they seek to describe, classify, and interpret language online as it is actually used—while those of educators, language purists, and the mass media are largely prescriptive—they seek to provide direction for how online language should (or should not) be used. As with traditional written and spoken language, tensions arise between these two approaches, especially as concerns the Internet and language change. From a prescriptivist perspective, the non-standardness of much Internet language and its alleged effects on non-mediated language indicate a dangerous trend toward a decline in the overall quality of the language (which is usually English in these debates, although prescriptive concerns have also been raised about other languages, especially Greek). There is a tendency to attribute this trend to the medium itself, consistent with a belief in technological determinism (see Baron, 2000). Descriptivists, in contrast, are interested in documenting trends and understanding the

factors that influence them, whether those factors are technical, social, or a combination of both, and they do not assume that non-standard language is degenerate. On the contrary, they often characterize such language as playful, creative, and performing useful social functions (Cherny, 1999).

Moreover, linguists, in particular, are cautious about assuming that variation in online language will necessarily lead to long-term language change. They distinguish between practices primarily restricted to online environments—including acronyms such as 'ttyl' (talk to you later) and "l33tspeak" (from 'leet,' a shortening of 'elite'), which employs letters, numbers, and non-alphanumeric keyboard characters to represent words—and practices that have extended to offline language. In the latter case, they distinguish further between relatively superficial lexical changes— a shift in meaning of words such as 'spam' and 'lurk'; the addition of new vocabulary such as 'email' and 'emoticon;' new word formative elements such as the prefixes 'e-' and 'cyber-'—and deeper, syntactic change, of which there is little evidence as yet (Stein, 2006).

Major Areas of Research

In addition to, and cross cutting, the issues identified above, research on language on the Internet may be grouped into the five major areas discussed below.

Classification research aims to characterize and label computer-mediated language. Much early research focused on the relationship of computer-mediated language to the modalities of speech and writing. Like the latter, it is produced and read as text, but like the former, Internet language tends to be informal and context-dependent, especially in synchronous modes, and message exchanges can feel like conversation, leading some scholars to characterize Internet language as "written speech." Others consider it to be a distinct, third modality (see "Netspeak" below). Another level of classification differentiates among modes or genres of computer-mediated discourse, such as Instant Messaging, webboards, and blogs, noting that different language practices are characteristic of each (--> Genre). A more fine-grained approach classifies computer-mediated discourse samples according to a set of features, such as synchronicity, participant structure, and topic, that cut across modes (Herring, 2007).

Historically and continuing to the present time, the most popular area of language and Internet research has been the *structural features* of Internet language, especially typography, orthography, and neologisms (new word formations). Crystal (2001) has popularized the term "Netspeak" to refer to the use of abbreviations, emoticons (combinations of keyboard symbols that represent, for example, a smiling face), and playful typography that is claimed to characterize Internet language as a unique language variety. Most structural analyses of Internet language catalogue lists of features or speculate about the reasons for their existence; as Androutsopoulos (2006) notes, there has been little systematic study as yet of language variation involving such features.

Internet language also manifests a rich variety of *discourse patterns* (Herring, 2001). These include pragmatic phenomena such as politeness (and rudeness, including 'flaming' --> Verbal Aggressiveness), violations of relevance, and the performance of various speech acts (--> Linguistic Pragmatics); interactional phenomena such as turn-taking, repairs, and topic establishment, maintenance, and drift (--> Conversation

Analysis); and register phenomena such as gender styles, regional dialects, and in-group language practices characteristic of particular online communities (--> Language Varieties). These phenomena necessitate the analysis of language in context.

Internet language has also proven to be a useful *lens through which to study human behavior* more generally. Most online activity is instantiated through typed text, whether through information exchange, political debate, online learning, making friends, or flirting (--> e.g., Personal Communication by CMC). The discourse of each of these activities has characteristic properties and can be mined for patterns (--> Discourse). Internet language in this sense has attracted the interest of non-language scholars whose goals are not to describe language for its own sake but rather to use it to gain a purchase on otherwise elusive but theoretically-rich concepts such as collaboration, community, democracy, identity, influence, performance, power, reputation, and trust (Herring, 2004).

Finally, *languages and language ecologies* have increasingly attracted attention as the Internet expands its global scope. The Internet has been claimed to accelerate the ongoing spread of English and perhaps other large regional languages such as Chinese and Spanish, although commentators disagree as to whether this is at the expense of smaller languages. While the numerical domination of English-language users has decreased considerably over the past decade, the use of English as a lingua franca appears to be growing as speakers of different languages come into contact via the Internet and use English as a common language. At the same time, differences across languages and cultural contexts are increasingly being documented that range from structural features such as emoticons and script-conditioned typography to gender patterns in online interaction (--> Gender and Discourse) (Danet & Herring, 2007).

Changes over Time in Internet Language and its Treatment

The order of presentation of the five areas above corresponds approximately to the order in which each emerged as a strand of language and Internet research, since the time of the first studies in the mid-1980s, although all of these areas continue to attract attention. A general trend can be noted, from a fascination with the technologically-conditioned features of Internet language in the early years, to a growing awareness that social and contextual factors shape online language use, much as they do offline language, and that culture and geographical setting may condition more variation than originally envisioned (Georgakopoulou, 2003).

Simultaneously, from the two basic modes of email (asynchronous) and chat (synchronous) that were popular in the early 1990s (--> Electronic Mail), the types and capabilities of CMC systems that support linguistic communication have expanded to include web-based modes such as blogs, wikis, and social network sites; semi-synchronous modes such as instant messaging; graphical modes such as avatar-based virtual worlds; and Internet telephony (Voice over IP). Inevitably, Internet language research lags somewhat behind these developments. Audio communication, in particular, has been little researched to date.

Methodological Issues and Challenges Associated with Language and the Internet

Internet language offers a number of advantages for research, including an abundance of naturally-occurring (i.e., non-experimental) data that, unlike speech, does not require transcription and that can be readily analyzed using computational means (--> Logfile Analysis). The ability of researchers to "lurk" in online environments without their presence being visible or salient to participants is another advantage, especially for studies of social interaction, albeit one that raises ethical issues (--> Research Ethics, Internet Research).

From the linguist's point of view, text-based Internet language is lacking in sound, and therefore questions of phonetics and phonology, which are central to linguistics, can not directly be addressed (--> Linguistics). Variationist sociolinguists are challenged to apply their methods, because it is often difficult or impossible to ascertain the demographics of text producers, which typically constitute the independent variables in studies of linguistic variation. Researchers interested in online multilingualism (--> Bi- and Multilingualism) or in describing online discourse in other languages may encounter the confounding effect of English influence on those languages, especially as regards borrowed vocabulary and popular abbreviations.

An ongoing challenge associated with language and the Internet is the rate at which new technologies continue to be introduced. Since the affordances of media can affect language use through those media, it is necessary to consider each new CMC mode first in its own terms, a situation that has distracted attention somewhat from the development of theory about online language. At the same time, the ready availability of new modes provides a rich opportunity to study the emergence of language practices, norms, and social behaviors as expressed through discourse, and to theorize about emergent language phenomena.

Future Directions in Research, Theory, and Methodology

There is a need to move beyond description to theorize CMC effects on language. Theories should also be tested empirically on large corpora of contextually-classified (i.e., tagged) computer-mediated language samples that can be compared systematically across modes, contexts, and languages. Related to this is a need for Internet language preservation efforts, particularly in the case of synchronous data that are not automatically logged, and for longitudinal studies to investigate mediated language change. It is likely that computer-mediated language will be perceived as just plain language by future generations who have grown up with it; it will thus need to be studied to address broader questions of language change.

Future language research will almost certainly devote increased attention to spoken and visually-enhanced modes of networked and mobile communication. At the same time, the ascendancy of multimedia CMC over text-based CMC has been predicted for nearly a decade, yet text remains the most popular format. Whatever else the future holds, it seems certain that people will continue to use new media to communicate and that they will do so using human languages, and thus that the study of language and digital media will be relevant for years to come.

SEE ALSO: Bi- and Multilingualism; Conversation Analysis; Deceptive Message Production; Discourse; Electronic mail; Gender and Discourse; Genre; Identities and Discourse; Internet, Technology of; Language Varieties; Linguistic Pragmatics; Linguistics; Logfile Analysis; Personal Communication by CMC; Research Ethics, Internet Research; Technologically-Mediated Discourse; Verbal Aggressiveness; Virtual Communities

REFERENCES AND SUGGESTED READINGS

Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics*, 10(4), 419-438.

Baron, N. S. (2000). *From alphabet to email: How written English evolved and where it's heading*. London: Routledge.

Cherny, L. (1999). *Conversation and community: Chat in a virtual world*. Stanford, CA: Center for the Study of Language and Information.

Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.

Danet B., & Herring, S. C. (eds.) (2007). *The multilingual Internet: Language, culture, and communication online*. New York: Oxford University Press.

Georgakopoulou, A. (2003). Computer-mediated communication. In J. Verschueren, J.-O. Östman, J. Blommaert, & C. Bulcaen (eds.), *Handbook of Pragmatics* (2001 installment). Amsterdam: John Benjamins, pp. 1-20.

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

Herring, S. C. (2004). Computer-mediated discourse analysis: An approach to researching online behavior. In: S. A. Barab, R. Kling, & J. H. Gray (eds.), *Designing for Virtual Communities in the Service of Learning*. New York: Cambridge University Press, pp. 338-376.

Herring, S. C. (2007). A faceted classification scheme for computer-mediated discourse. *Language@Internet*, article 761.

Stein, D. (2006). Language on the Internet. In E. K. Brown (ed.), *Encyclopedia of language & linguistics*. Amsterdam: Elsevier.

Susan Herring is Professor of Information Science and Linguistics at Indiana University, Bloomington. She has a Ph.D. in Linguistics from the University of California at Berkeley. Her research focuses on the linguistic and social aspects of communication mediated by new technologies, especially with regard to gender, interaction management, classification, and discourse analysis methodologies. She is the editor of *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives* (Benjamins, 1996), *Computer-Mediated Conversation* (Hampton, in press), and co-editor of *The Multicultural Internet: Language, Culture, and Communication Online* (Oxford, 2007).