Grammar and Electronic Communication

Susan C. Herring
Indiana University, Bloomington
herring@indiana.edu

Introduction

How new digital communication technologies are affecting language and language change has excited considerable speculation and inspired a growing body of scholarship. At the heart of this question is the notion of grammar, or the set of structural rules that govern the composition of words, phrases, and sentences in a language. Language purists worry that electronic communication (also known as computer-mediated communication, or CMC) is corrupting the grammar of their language. The features of “Netspeak” (see below) are for the most part grammatical innovations, although scholars do not generally view their effects as detrimental.

To begin, grammar in electronic language must be understood somewhat differently from its usual sense as applied to speech. While the grammar of spoken language includes phonology, morphology, and syntax, in text-based CMC phonology is largely irrelevant; typography and orthography take over the functions of sound. Grammar in CMC also differs in that electronic language, as a new and still emergent phenomenon, has not yet had time (nor attained the requisite social status) to become formalized in “rules;” rather, it exhibits patterns that vary according to technological and situational contexts (Herring, 2007).

This article is primarily concerned to describe these patterns. Specifically, its focus is structural characteristics in text-based CMC at the utterance level (for an overview of discourse-level features, see Herring, 2001). It favors the English language, since most language-focused CMC research to date has analyzed English; however, a section on other languages is included. The structure of the lexicon is excluded, except insofar as it involves word formation processes that originated in CMC. Due to space constraints, the coverage is necessarily representative, rather than exhaustive, of the findings that have amassed since computer-mediated language first attracted research attention more than 25 years ago.

E-grammar in English

As a convenience, the set of features that characterize the grammar of electronic language is referred to henceforth as e-grammar, although the use of this term is not intended to imply that there is a single grammar for all varieties of computer-mediated language. One linguist, David Crystal (2001/2006), has suggested as much (using the term “Netspeak” to refer to CMC as a single language variety), but considerable empirical evidence points to e-grammar as varying systematically across languages, contexts, users, and technological modes (e.g., Bieswanger, 2007; Herring, 2007; Johanyak, 1997).
The presentation that follows is organized according to the traditional hierarchy of grammatical phenomena in linguistics. It proceeds from the micro-levels of typography and orthography, through morphology at the word level, to syntax at the utterance level.

**Typography**

In text-based CMC, typography refers primarily to the use of non-alphabetic keyboard symbols such as numbers, punctuation, and special symbols such as <, $, and @. It also includes nonstandard capitalization (including ALL CAPS, lack of initial capitalization, alternating upper and lower case, and so-called 'camel case,' or writing compound words or phrases together with MedialCapitals), as well as emoticons, or sequences of keyboard characters that prototypically imitate facial expressions (e.g., :-D and :P represent a laughing face and a face sticking out its tongue, respectively). Western-style emoticons are viewed at a 90-degree angle, whereas Asian-style emoticons are viewed straight on, e.g., 0_o (a confused face) and <.< (eyes looking sideways); both types are used in English CMC. Emoticon use is widely claimed to be one of the defining typographical characteristics of electronic language (e.g., Crystal, 2001/2006), although studies of emoticons in English CMC report that they occur less often than popularly believed, that the overwhelming majority are simple ‘smilies’ :-) or ‘winkies’ ;-), and that emoticon use varies according to CMC mode and user gender (for an overview of emoticon research, see Dresner & Herring, 2010).

Other typographic characteristics of CMC include repeated punctuation (!!!, ?! ...) and the substitution of numbers or letters for words or parts of words (e.g., 4 ‘for,’ 2day ‘today,’ ur gr8 ‘you’re great’). This latter usage is also sometimes classified as nonstandard spelling; indeed, there is considerable overlap between nonstandard typography and nonstandard orthography in CMC, and the two often co-occur.

Except for emoticons, which may be composed entirely of typographic symbols – and excluding as outside the scope of grammar drawings composed of keyboard characters, such as @}--‘--- to represent a rose (Werry, 1996) – the most conceptually extreme manipulation of typography in plain text English CMC is Leetspeak or Leet. In Leetspeak, some or all letters of a word are replaced by non-alphabetic symbols based on graphic resemblance; thus, Leet becomes 1337 or 133+, and @$$! is a common profanity. The name Leet comes from ‘elite,’ and the style of writing originated among early Internet users as a kind of secret code, especially among hackers exchanging pirated files; it has since become popular in a number of online gaming communities. Some adepts claim it is a language variety with its own vocabulary, morphology, and grammar (Wikipedia, 2010a).

More common in CMC in general is the occasional substitution of words or parts of words with numbers or letters to save keystrokes and/or to symbolize a playful communication style or social identity. Repeated punctuation is also common –
although it makes messages longer, not shorter – to express affect (repeated letters, an orthographic strategy, functions similarly). Such nonstandard typography is especially common in Short Message Service text messaging (SMS), where the message buffer is limited to 160 characters, followed by synchronous CMC modes such as chat and instant messaging (IM), in which exchanges are typically rapid and social in nature; it is also found in email, even in professional contexts (Anis, 2007; Cho, 2010; Murray, 2000).

Orthography

Nonstandard orthography is widely considered to be a defining characteristic of computer-mediated language, and indeed, e-communication often manifests spelling practices that suggest loosened orthographic norms. These include abbreviation (acronyms, clippings, vowel omission as in pls for ‘please,’ etc.); phonetically-motivated letter substitutions (e.g., z for ‘s’); spellings that imitate casual or dialectal pronunciations (e.g., wassup? for ‘what’s up?’); eye dialect (e.g., sez for ‘says’); and spellings that represent prosody or nonlinguistic sounds, such as a ‘calling voice’ (helloooo), laughter, and other (nonhuman) noises.

Language prescriptivists typically view such practices as misspellings or errors (Thurlow, 2006). In fact, the proportion of nonstandard spellings that are unintentional is very low among native speakers. Abbreviations save keystrokes, as do some phonetic spellings, and representing speech in writing is a manifestation of the “orality” of much text-based CMC (e.g., Cho, 2010). Moreover, representations of non-language sounds enrich CMC in the absence of auditory cues. They often accompany other kinds of textual performance, including indications of actions, for which an entire set of orthographic/typographic conventions has developed, such as *w* for a wink (e.g., Cherny, 1999; Werry, 1996).

Early studies such as Cherny’s and Werry’s emphasized the playfulness and creativity driving these phenomena, especially in recreational chat environments. However, recent research suggests that a relatively small number of nonstandard spellings (e.g., u ‘you,’ msg ‘message,’ wanna ‘want to’) have become conventionalized and occur most often in mainstream online contexts, while unique formations are less common (Kapidzic, 2010).

A counterexample to this trend is ‘lolspeak,’ the fractured text that accompanies images of ‘lolcats,’ which were popularized on the Internet several years ago by a photograph of a cat captioned I can haz cheezburger? Lolspeak remains popular in the image forums where it originated, and it is sometimes used (minus images) for humorous effect in other CMC contexts, including Facebook status updates. Another counterexample is the language use of fans of the character ‘Ali G.,’ as created by the (white) British comedian Sasha Baron Cohen, on websites and discussion boards; fans imitate Ali G.’s spoken style, which is a mix of Jamaican Creole and Southern British English street language, through creative spelling and unconventional grammar and lexis (Sebba, 2007). Like Leetspeak, these cases represent special
registers in which a concept (substitution of letters with other symbols; cats’ imaginary ‘bad English;’ a comedian’s humorous language style) generates unbridled orthographic (as well as morphological and syntactic) creativity.

Morphology

Relatively fewer descriptions of computer-mediated language mention morphology. Those that do mainly note the emergence of a few productive word formatives (e.g., e-, cyber-, hyper-) and the outcomes of word formation processes such as clipping (e.g., nick from ‘nickname’), blending (e.g., netizen from ‘network citizen’), acronyms (e.g., lol ‘laugh out loud,’ jk ‘just kidding,’ OMG ‘oh my god,’ wtf ‘what the fuck’), semantic shift (e.g., flame ‘unleash invective on a computer network,’ from flame ‘to act conspicuously homosexual’; spam ‘Internet junk mail,’ from spam ‘a type of canned meat’), and conversion from one part of speech to another (e.g., text as a verb; spam as a verb). These processes are not unique to CMC, but they have been especially productive on the Internet, generating many new words that are increasingly making their way into dictionaries of Standard English.

Less commonly attested word formation processes include outright neologisms, such as newbie (sometimes clipped to noob or n00b, ‘an inexperienced person’), and conventionalization of frequently-occurring typographical errors, such as teh (‘the’). Leetspeak claims both of these, along with several productive derivational and inflectional suffixes, including –age (e.g., flamage, from the verb ‘flame’), -zor (parallel to Standard English –er/or), and –zorz (which intensifies the meaning of a verb, as in pwnzorz, ‘really defeat,’ from pwn, itself a conventionalized misspelling of the word ‘own’). These suffixes are not in general use.

The most creative examples of e-morphology have been reported in playful, self-contained contexts, for example in a multiplayer online game (MOG) (Nilsson, 2009); in a social MUD (Multi-User Dimension), a type of a text-based virtual world (Cherny, 1999); and in emails exchanged in a private sibling code (Rowe, forthcoming, 2011). MOGs, in which interaction can be intense and fast-paced, generate numerous acronyms and shortened forms specific to the game context, e.g., gg (‘good game’), wtb (‘want to buy’), and lvl (‘level’). Examples from the social MUD, which Cherny describes as a relatively closed community of computer geeks, include the productive derivation of verbs from interjections (e.g., Mike cools, meaning ‘Mike says “cool”’) and verb reduplication (e.g., nodsnods) to indicate a repeated or emphatic action. The sibling code, which originated between two sisters in childhood but increased in use when they started emailing as adults, produced such novel words as the clipping immuze (‘immunizations’) and fibin’ (acting like a type ‘five’ (fibe), i.e., avoiding), which involves both semantic shift (from 5 as a numeral to 5 as a personality type) and conversion (from a noun to a verb).

As these examples illustrate, more than one morphological process may operate on a single word, and unconventional morphology sometimes combines with unorthodox typography and/or orthography.
Syntax

The syntax of computer-mediated English, when it deviates from standard syntax, is sometimes described as ‘telegraphic’ and fragmented. Parts of speech such as articles and subject pronouns may be elided in informal style, and messages that do not contain a complete grammatical clause (with a subject and finite predicate) are common, especially in CMC modes characterized by brief, informal messages, such as chat, IM, SMS, and microblogging. The usual reason given for elision is to save keystrokes (e.g., Murray, 2000), whereas sentence fragments may be caused by people typing speech-like utterances and/or the requirement in some CMC systems that messages be brief, which can lead users to break longer utterances into several messages (e.g., Baron, 2010).

CMC syntax also diverges from the standard when users attempt to represent a nonstandard language variety, such as African American Vernacular English (often via copula deletion or invariant ‘be’) or Ali G’s fractured style (Sebsa, 2007). Moreover, special registers of CMC sometimes evolve productive syntactic strategies not found in other CMC modes, such as preposition deletion in the MUD register described by Cherny (e.g., John laughs Lynn), the double-inflected-modal can haz construction in lolspeak, and nominalization of verbal predicates for emphasis in Leet (e.g., Au5t1N is t3h r0xx0rz, literally ‘Austin is the rocks-er,’ meaning ‘Austin really rocks’; Wikipedia, 2010a).

A syntactic innovation particular to CMC environments is 3rd person singular present tense performative utterances, also called ‘emotes.’ Emotes are commands that cause the user to perform a social action by logging a description of that action into a chat window; in a MUD, for example, when Kim types /waves, the message Kim waves is displayed (Cherny, 1999). Emotes started as pre-programmed shortcuts in online game environments, where they are still popular. Until recently, Facebook promoted the use of similar-looking 3rd person present tense utterances by providing the default prompt [Username] is ... for 1st person status updates. Both systems sometimes give rise to inconsistent pronoun and tense usage, as in Susan waves goodbye and puts on my hat and Jim Cosmo is just saw a meteorite. For the most part, however, such mixed constructions are assumed to be unintentional and (at least mildly) infelicitous.

Related to emotes are predications that can function alone as complete performative utterances, such as *waves*, <grin>, *confused*, and *in a bad mood*. Semantically, such ‘performative predications’ may involve virtual actions or states. Structurally, they may be inflected or uninflected, and they are often set off by typographic brackets. Rather than being pre-programmed, they are typed out by the user, although a number are highly conventionalized. A common type expresses vocalizations, such as *sigh(s)/sob(s)/laugh(s)/lol(s)*, that could also be expressed via typed representations of sound (e.g., hahaha) or emoticons. Note that the acronym lol may be inflected like a verb; it may also represent a pulse of laughter...
is characterized by in the case of 'Greeklish Khalil alphabet substitutions 2007) characters in other languages) due to often (2008) favored in different languages and used in different examples, for Italian SMS there is ample evidence that CMC users in First, nonetheless languages raise issues (such as code-mixing) and/or be spelled like it is pronounced (e.g., lawl, lulz), especially in teen chat (Kapidzic, 2010). Most popular in social chat environments, where they originated, performative predications are also common in other CMC modes.

Many researchers have measured the frequency of grammatical function words, such as pronouns, determiners, modal auxiliaries, and negation, in electronic corpora. CMC can be distinguished from traditional genres of speech and writing according to these measures; typically it falls between the two extremes, with synchronous chat closer to casual speech and asynchronous modes such as email closer to formal writing (e.g., Ko, 1996; Yates, 1996). Correlations have also been reported between frequency of grammatical markers and nonlinguistic phenomena, such as lying in CMC (e.g., Hancock et al., 2008), gender of blog authors (Herring & Paolillo, 2006), and bloggers’ psychological response to trauma (Cohn et al., 2004).

**E-grammar in Other Languages**

Structural features of CMC in languages other than English have been less studied (cf. Danet & Herring, 2007), although this varies by language: German CMC, for example, has received considerable attention (in German), and the typography and orthography of French CMC have been described in several works by the late Jacques Anis (in French). At the other extreme, the structure of many languages used online remains virtually undescribed (e.g., Russian, Korean, African languages), at least in venues accessible to Western readers. Furthermore, typography and orthography have been better described than morphology and syntax, and some languages raise issues (such as code-mixing) that blur these levels of analysis. Nonetheless, several general findings can be noted.

First, contrary to Crystal’s (2001/2006) claim, “Netspeak” is not universal. While there is ample evidence that CMC users in other languages play with typography and orthography (see, e.g., Anis, 2007 for French SMS; Herring & Zelenkauskaite, 2009 for Italian SMS; Nishimura, 2003 for Japanese webboards); different strategies are favored in different languages and used in different proportions, as Bieswanger (2008) demonstrated through a systematic comparison of English and German SMS.

Second, and relatedly, languages that make use of non-Roman writing systems have often had to use ASCII (plain text) in CMC environments (although this is changing due to Unicode and programs that convert QWERTY keyboard symbols to characters in other languages), requiring creative adaptation (Danet & Herring, 2007). Some languages exhibit both sound-based and graphic-based character substitutions. For instance, the Greek letter ‘theta’ is written in plain text CMC both as ‘th’ (representing its sound) and ‘8’ (approximating its appearance in the Greek alphabet) (Tseliga, 2007; for an analogous practice in Arabic, see Palfreyman & Khalil, 2003). These Romanized forms may in turn create new language varieties, as in the case of ‘Greeklish,’ which in addition to being written in the Roman alphabet is characterized by less sentence-initial capitalization, more emoticons, and more
English borrowings (e.g., ‘message,’ ‘pc,’ ‘sorry,’ ‘sex’) than Greek written in the Greek alphabet (Tseliga, 2007).

This leads to the third observation, which is that English influences other languages used online, through borrowing of computer-related terms and e-grammar conventions. Lee (2007) found, for instance, that Cantonese in Hong Kong dyadic chat, which is written in a mix of Romanization and Cantonese and Mandarin characters, includes many English “Netspeak” features, along with the use of numerals to represent sounds (e.g., 99 ‘nighty night’ in Chinese English pronunciation). Code mixing of a local language with English or another regional lingua franca occurs often in bi- and multilingual environments. For example, in webforums and chat rooms for Greeks, Turks, and Persians living in Germany, mixing of the native languages and German takes place on multiple linguistic levels: lexical, phrasal, propositional, and discourse-pragmatic (Androutsopoulos, 2007a).

English is not the only language with special online registers. German rappers post messages on web forums in which they imitate ‘hiphop speech’ in a mix of German and English (Androutsopoulos, 2007b). Russian padonki (an intentional misspelling of podonki ‘riff-raff,’ ‘scum’) is a subcultural phenomenon reminiscent of lolspeak, characterized by erratic spellings and gratuitous use of profanity for comic effect (Wikipedia, 2010b). The Fakatsa style popular among Israeli girl bloggers, which replaces Hebrew letters with graphically-similar numbers and Roman letters, somewhat resembles Leetspeak (Vaisman, 2011).

Nonstandard language use online poses problems for machine translation. Climent et al. (2003) found this to be true even in academic newsgroup postings in two structurally similar languages, Catalan and Spanish. The available research suggests that structural irregularities are prevalent in CMC, even if for the most part e-grammar adheres to standard language norms (cf. Thurlow, 2006).

**E-grammar and Language Change**

More than 25 years ago, Baron (1984) predicted that the Internet would change language for the worse. More recently, Stein (2006) predicted that it would accelerate the rate of language change. What limited diachronic studies of online language are available suggest that e-grammar tends to become, if anything, more standard over time (e.g., Herring, 1998), but that in subcultural and interpersonal contexts of intensive, self-contained interaction, such as the private sibling code analyzed by Rowe (forthcoming, 2011), e-grammar may evolve at an accelerated pace.

Many e-grammar innovations have been adopted by the wider community of Internet users. Features such as emoticons and certain acronyms and spellings have been taken up across the Internet, including being borrowed into other languages’ CMC. Others, such as emotes and ‘performative predicates,’ have also extended productively beyond their source contexts. Even features of special registers such as
Leet, lolspeak, and online rapper language are attested outside their source contexts, although such usage is often tied referentially to its original contexts of use.

The evidence is less clear as regards the diffusion of e-grammar into language use offline and into languages themselves, as codified in grammars and dictionaries. Dictionaries are certainly including more CMC and computer-related terms all the time. Yet evidence of the spread of other CMC practices to a language as a whole tends to be anecdotal. Several years ago, National Public Radio in the U.S. broadcast a segment about teens speaking “Netspeak” abbreviations (e.g., pronouncing lol as /lawl/ and I <3 u as ‘I heart you’). Marketers sometimes use nonstandard online forms, including from lolspeak and padonki, in billboards and advertisements (see, e.g., Wikipedia, 2010b). Such uses seem humorous, however, and self-consciously Internet referential.

At the same time, self-report studies indicate that young people are increasingly using e-grammar in their offline writing (e.g., Pew, 2009). Rather than this causing a decline in their language skills, Plester et al. (2009) found that greater knowledge of SMS abbreviations was associated with higher word reading, vocabulary, and phonological awareness measures. While this research is limited, and many gaps remain, it suggests that language change is being affected and effected by Internet communication, and that if anything, e-grammar enriches rather than impoverishes language users and languages themselves.

Cross-references

SEE ALSO: Analysis of Mediated Interaction; Computer-Mediated Communication and Second Language Use; Grammar and the Media; Massively Multiplayer Online Games; Multilingualism and the Internet; Orthography; Pragmatics of Asynchronous Computer-Mediated Communication; Pragmatics of Chat; Pragmatics of Short Message Servicing; Prescriptive and Descriptive Grammar

References


Rowe, C. (Forthcoming, 2011). Genesis and evolution of an e-mail-driven sibling code. *Language@Internet, 8.*


Suggested Readings


