Grammar and Electronic Communication

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

Abstract

How digital communication technologies are affecting language, literacy, 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. Grammatical usage in computer-mediated communication (CMC), referred to here as e-grammar, often deviates from prescriptive norms of written language. This entry describes nonstandard structural characteristics of e-grammar at the levels of typography, orthography, morphology, and syntax in text-based English CMC. Separate sections address emoji grammar, e-grammar and language change, and e-grammar and literacy.

Keywords: CMC, creativity, emoji, grammar, language change, linguistic structure, literacy, morphology, nonstandard, orthography, syntax, typography

Introduction

How digital communication technologies are affecting language, literacy, 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. Electronic communication (also known as computer-mediated communication, or CMC) often deviates from prescriptive norms of written language, causing language purists to worry that “netspeak” (see below) is detrimental to language skills and to the language itself (Thurlow, 2006). However, CMC scholars do not generally view nonstandard language use in CMC in a negative light.

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 relatively recent phenomenon, has not yet become formalized in “rules”; rather, it exhibits patterns that vary according to technological and situational contexts (Herring, 2007).

This entry is primarily concerned to describe these patterns. Specifically, its focus is nonstandard structural characteristics in text-based CMC at the utterance level in English (for an overview of discourse-level characteristics of electronic language, see Herring & Androuloupoulos, 2015). Contextual and demographic factors associated with electronic language use are excluded due to space constraints. The coverage is necessarily representative, rather than exhaustive, of the findings that have amassed since computer-mediated English first attracted research attention 35 years ago.
E-Grammar in English

As a notational convenience, the set of features that characterize the grammar of electronic language is referred to henceforth as e-grammar. 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), and other terms, such as “textese”, “chatspeak”, and “weblish”, are also sometimes used in a general sense. However, there is considerable empirical evidence that e-grammar varies systematically across technological modes, contexts, users, and languages (e.g., Bieswanger, 2007; Johanyak, 1997), although such variation is not addressed here.

The presentation that follows is organized according to the traditional hierarchy of grammatical phenomena in linguistics. It proceeds from the microlevels 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 nonalphabetic keyboard characters such as numbers, punctuation, and special symbols such as <, $, #, and @. It also includes nonstandard capitalization (including lack of utterance-initial capitalization, ALL CAPS, alternating uPpEr AnD lOwEr case, and so-called “camel case”, or writing compound words or phrases together with MedialCapitals). Emoticons, or sequences of keyboard characters that prototypically imitate facial expressions (e.g., 😊:D represents a laughing face), have been claimed to be one of the defining typographical characteristics of electronic language (e.g., Crystal, 2001/2006). In recent years, however, emoticon use has declined, eclipsed by the popularity of emoji (see below).

Other typographic characteristics of CMC include omitted punctuation (especially at the end of messages containing a single utterance), 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.

A 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 nonalphabetic 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. Some adepts claim it is a language variety with its own vocabulary, morphology, and grammar (Wikipedia, 2018).

More commonly, e-grammar involves a combination of standard and nonstandard forms. The most common typographic manipulations are omission of initial capitalization and final punctuation to save keystrokes (Drouin & Driver, 2014). Repeated punctuation is also common—although it makes messages longer, not shorter—to express affect (repeated letters, an orthographic strategy, functions similarly; Kalman & Gergle, 2014). The occasional replacement of words or parts of words with numbers or letters saves keystrokes and also symbolizes a playful communication style or social or individual identity (e.g., Roeder et al., 2019). Such nonstandard typography is especially common in
short message service text messaging (SMS), followed by CMC modes such as instant messaging (IM) and synchronous chat, in which exchanges are typically rapid and informal in nature. It is also sometimes found in email, even in professional contexts (Cho, 2010; Murray, 2000). The @ and # symbols have taken on conventional functions, indicating, for example, addressivity and message topic, respectively, on Twitter and other social media sites (Honeycutt & Herring, 2009; Page, 2012).

**Orthography**

Nonstandard orthography is the main defining characteristic of e-grammar. CMC often manifests spelling practices that suggest loosened orthographic norms, including abbreviation (acronyms, clippings, vowel omission as in *pls* for ‘please’, etc.); phonetically motivated letter substitutions (e.g., *x* for ‘*ks*’); and other irregular spellings (e.g., *2moz* for ‘tomorrow’). Also common are accent stylizations—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” (*hellowoo*), laughter, and other (nonhuman) noises.

Language prescriptivists typically view such practices as errors (Thurlow, 2006). However, while misspellings due to carelessness or low literacy skills (see below) do occur, native English speakers’ use of orthographic manipulations of the above types is usually intentional. 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 nonlanguage sounds enrich CMC in the absence of auditory cues. These representations often accompany other kinds of textual performance, including indications of actions via “performative predicates” (Herring, 2013) such as *waves*, <grin>, *confused*, and *in a bad mood*, for which an entire set of orthographic/typographic conventions has developed (e.g., Cherny, 1999).

Early studies such as Cherny’s emphasized the playfulness and creativity driving these phenomena, especially in recreational chat environments. More 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). Moreover, technological affordances such as predictive spelling and speech-to-text software have reduced the frequency of nonstandard orthography and typography (Turner et al., 2014). Many texters still generate nonstandard forms, but they may need to invest extra effort to do so, evidence of their conscious intent to be playful or creative.

Playful varieties of online language exist that are based in part on orthographic manipulation. “LOLspeak” is the fractured text that accompanies images of “LOLCats”, popularized on the internet several years ago by a photograph of a cat captioned *I can ha cheezburger?* Another example is the language use of fans of the character “Ali G”, as created by the (White) British comedian Sacha Baron Cohen, on websites and discussion boards; fans imitated 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 (replacement 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

Descriptions of computer-mediated English that mention morphology note the emergence of productive word formatives (e.g., e-, i-, cyber-) and the outcomes of word formation processes such as clipping (e.g., blog from ‘weblog’), blending (e.g., netizen from ‘network citizen’), affixation (e.g., defriend, textese), compounding (e.g., flash mob, kthxbye ‘ok thanks, bye’), acronyms (e.g., lol ‘laugh out loud’, jk ‘just kidding’, wtf ‘what the fuck’), semantic shift (e.g., troll ‘one who makes controversial comments in internet forums with the intent to provoke a knee-jerk reaction’, from troll ‘a mythical, cave-dwelling being’), and conversion from one part of speech to another (e.g., text as a verb; mouse as a verb). The popular acronym lol may itself be inflected like a verb (e.g., I LOLed); it may also be reduplicated to represent pulses of laughter (lololol) or be spelled as it is pronounced (lawl) and pluralized (lulz). These processes are not unique to CMC, but they have been especially productive on the internet, generating many new forms.

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, ‘unleash invective in an internet forum’), -or (parallel to Standard English -er/-or), and -orz (which intensifies the meaning of a verb, as in pwnorz ‘really defeat’, from pwn, a conventionalized misspelling of the word ‘own’). These suffixes are not in general use.

Some of the most creative examples of e-morphology have been reported in playful, communication-intensive contexts, for example in a multiplayer online game (MOG) (Nilsson, 2009); in a social MUD (multi-user dimension), a type of text-based virtual world (Cherny, 1999); and in emails exchanged in a private sibling code (Rowe, 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 described as an active 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 dramatically 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 casual style, and messages that do not contain a complete grammatical clause (with a subject and finite predicate) are common in CMC modes characterized by brief, informal messages, such as SMS, IM, chat, and microblogging. The
usual reason given for elision is to save keystrokes (e.g., Murray, 2000). Sentence fragments may be caused by people typing speech-like utterances, or by the requirement on some CMC platforms that messages be brief, which can lead users to break longer utterances into several messages (e.g., Baron, 2010).

Social media have recently given rise to more deeply disruptive syntactic innovations. For example, unlike in Standard English, the complementizer because can be followed by a bare noun, an adjective, or an interjection, e.g., ‘I’m going to bed now, because sleep/exhausted/aaargh’. The eliptical expression ‘I can’t even’ and derivatives such as ‘I am unable to can’ and ‘I’ve lost the ability to even’ represent innovations in which the modal auxiliary can and the intensifying adverb even are relexicalized as finite verbs. McCulloch (2004) suggests that these constructions indicate “stylized verbal incoherence mirroring emotional incoherence”.

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 (Sebba, 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 has 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, 2018). Doge, an online language variety similar to LOLspeak which humorously represents the imagined language of dogs, features word pairings that violate the semantic co-occurrence restrictions of Standard English, such as very biscuit and much wow (McCulloch, 2014).

Researchers have measured the frequency of grammatical function words, such as pronouns, articles, prepositions, auxiliary verbs, and negation, in electronic corpora, using automated tools such as Linguistic Inquiry and Word Count (LIWC; e.g., Pennebaker, 2011). According to these measures, different registers of CMC can be distinguished (Biber & Egbert, 2018), and CMC can be distinguished from traditional genres of speech and writing; 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). The frequency of grammatical markers in CMC also correlates with nonlinguistic phenomena, such as personality traits (Pennebaker, 2011), lying in dating profiles (Toma & Hancock, 2012), gender of blog authors (Herring & Paolillo, 2006), and bloggers’ psychological response to trauma (Cohn et al., 2004).

Nonstandard language use online poses problems for automated language analysis, such as part-of-speech tagging (Eisenstein, 2013), machine translation (Climent et al., 2007), and profanity detection (Sood, Antin, & Churchill, 2012). The available research suggests that structural irregularities are prevalent in CMC, even if for the most part e-grammar adheres to standard language norms.

Emoji Grammar

Emoji are small, colorful graphical icons that represent facial expressions, objects, actions, and symbols. Some render automatically when a user types the keyboard sequence for an emoticon; more elaborate emoji are selected from menus provided by mobile phone providers and social media platforms. Individual emoji can substitute for letters, (parts of)
words, phrases, and entire utterances. They can be repeated to indicate plurality (’es: ‘babies’) and adverbial modification (ʼfly quickly’). In combination, they form sequences that can resemble sentences with subject, object, and verb, such as 🍕❤️️(msg ‘I love pizza’). Other sequences of emoji collectively represent concepts without regard for linear order, such as 🌴㸆📅 ‘Christmas’. One popular concept (Sexual intercourse’) shows evidence of grammaticalizing as an adjective, as in 🪤dddd ‘f**king crazy’. Given such evidence, some scholars (e.g., Danesi, 2016; Ge & Herring, 2018) maintain that emoji are evolving into a kind of graphical language.

As emoji use has increased, emoticon use has decreased. Simultaneously, the use of standard language has increased at the expense of nonstandard language (Pavalanathan & Eisenstein, 2016). This suggests that emoji are taking over some of the expressive and pragmatic functions of textual “netspeak”.

E-Grammar and Language Change

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 are available indicate that online language has become significantly more informal (Bourlai, 2018; Herring, 1998), as well as more emotional and negative (Bourlai, 2018), in a period of less than 25 years. E-grammar evolves especially rapidly in contexts of intensive, playful interaction, such as the LOLcats community (Gawne & Vaughan, 2011) and the private sibling code analyzed by Rowe (2011).

Accompanying the general trend toward informality has been a loosening of standard language norms. However, even though features of e-grammar are pervasive in certain contexts (such as texting and commenting in internet forums), their use remains stigmatized. Although they may use e-grammar features themselves, US university students evaluate such use by others negatively (e.g., Roeder et al., 2019). On the internet, nonstandard orthography, such as typographic errors and writing your for ‘you’re,’ is often met with censure from “grammar nazis,” individuals who scold or harass others for their language errors (Harris & Hiltunen, 2014; Sherman & Švelch, 2015).

Nonetheless, many e-grammar innovations have been adopted by the wider community of internet users. Features of special registers such as Leet and LOLspeak, including spelling words with nonalphabetic characters and the has construction, have entered general internet culture. (Typo)graphical features such as emoticons and emoji, acronyms such as lol and wtf, and spellings such as wanna and msg are widespread in CMC. The expressions because [reasons] and I can’t even are popular internet memes, and they can also be heard in offline speech.

E-grammar has diffused into offline language use and into languages themselves. Standard dictionaries accept more internet coinages every year. Alongside words referring to technology, such as email, cybercafe, and iPhone, internet-based terms are being used with broader meanings (e.g., to ping someone, meaning ‘get in touch’). Iconic “netspeak” features such as letter/number homophones grace storefronts, billboards, clothing, and gift items (e.g., Lee, 2015). In 2015, a major US car maker issued a press release entirely in emoji (Green, 2015). Moreover, although they are fundamentally visual phenomena, e-grammar typography and orthography are sometimes pronounced in speech by internet-
savvy individuals, e.g., /lal/ or /el oh el/ for lol, and /hart/ for the heart emoticon (<3) or emoji ❤, as in ‘I heart this song.’

**E-Grammar and Literacy**

In a recent survey, 40% of teachers in the US said that digital technologies make students more likely to “use poor spelling and grammar” (Purcell, Buchanan, & Friedrich, 2013). Young people are increasingly using e-grammar in their offline writing, yet whether this corresponds with poor reading and writing skills depends on which features of e-grammar are used. Creative uses of typography and orthography such as numbers for letters and accent stylization enhance metalinguistic awareness and have been found to be positively associated with literacy skills (Turner et al., 2014). In contrast, features motivated by ease of production, such as lack of initial capitalization and omission of final punctuation, as well as typos and spelling errors, are negative predictors of literacy (Drouin & Driver, 2014). Rather than e-grammar causing low literacy, however, it could be that people with lower literacy skills gravitate toward more casual, lazy uses, and people with higher literacy gravitate toward more creative uses. Lending plausibility to this view, Kern et al. (2014) found that Facebook users who use shorthand text, misspellings, and emoticons exhibit the personality traits of low conscientiousness and low openness to experience.

This research suggests that e-grammar per se does not reduce literacy skills; creative use of e-grammar might actually be a mark of growth and proficiency in literacy rather than a deficit. Meanwhile, electronic language continues to evolve in response to developments in CMC technology.

**References**


Baron, N. S. (2010). Discourse structures in instant messaging: The case of utterance breaks. *Language@Internet, 7*, article 4,


Cho, T. (2010). Linguistic features of electronic mail in the workplace: A comparison with memoranda. *Language@Internet, 7*, article 3,


**Suggested Readings**


