Orthographic variation of (lol)

Joshua McNeill

March 30, 2021

Abstract

[TBD]

1 Introduction

Studies of orthographic variation are not unheard of in sociolinguistics. This has been especially true since the widespread adoption of the internet as computer-mediated communication (CMC) presents both a wealth of relatively easily accessible data and social contexts that are less subject to overt social controls than what would have historically been the case for written forms as they had previously been largely relegated to educational and literary contexts. However, CMC studies and studies of orthographic variation in general have been focused mostly, though not always, either on situating CMC along a continuum between spoken language and older forms of written language or on the connection between phonology and spelling.

With some notable exceptions, an area of orthographic variation that has not been explored as much is how it functions on its own terms, independent of phonology. As such, the aim of the present study is to give an in depth examination of the potential social and pragmatic associations of variants of an orthographic variable that cannot reasonably be linked to either spoken language or educational and literary writing: (lol). Specifically, we will look at (lol) as used on the social media platform Twitter, whether certain spelling variants are associated with certain Twitter communities or certain positions within communities and whether certain spelling variants are associated with particular sentiments.

[Insert examples from data with explanation of the variable]

The rest of this introduction will be structured as follows. Section 1.1 will cover the general nature of CMC and any special considerations that are applicable to the present study. Section 1.2 will review work that has been done on orthographic variation. Finally, a thorough review of work that has been done on (lol), whether as a lexical or orthographic variable, will be presented in section 1.3.

1.1 Computer-mediated communication

With the advent of the internet, many new mediums of communication have entered daily life, collectively referred to as CMC. Likewise, by the end of the 1990s, research focused on these new mediums began in ernest. For instance, Cherny (1996) examined multi-user dungeons (MUDS), which are online, text-based, multiplayer games, whereas Yates (1996) analyzed language in computer conferencing systems, which are perhaps best thought of as progenitors of discussion forums, and Paolillo (2001) looked at chatrooms on internet relay chat (IRC). These were all early mediums, and so studies have progressed to other mediums as they have been developed, which include instant messaging (IM) (e.g., Baron, 2004; Tagliamonte & Denis, 2008), the social media photo-sharing platform Instagram (e.g., Stewart et al., 2017), and both commonly and importantly for the present study, the social media micro-blogging platform Twitter (e.g., Bamman et al., 2014; Eisenstein, 2013; Hong et al., 2011; Ilbury, 2020; Jones, 2015; Kim et al., 2014). Schneier (2021) also looked at text messaging on cell phones, though his data included use of any communication application on a cell phone.

Androutsopoulos (2008a) described the goals of these researchers of CMC as fitting into two separate "waves",[1] the first focused on the impact of the constraints placed on users of different CMC mediums on the language they produce, and the second focused more on analyses of what happens in CMC pragmaticly and sociolinguisticly (pp. 1-2), though one might also add that there has been a consistent interest in characterizing the relationship between CMC and both speech and older written mediums throughout both of these waves. A particularly useful tool for the present study – and indeed any CMC study – that was developed during the first wave is a typology for CMC mediums.

In order to compare results from a study centered on Twitter, as we are doing here, with results from previous work, it is important to have a framework for classifying different mediums. As Paolillo (1999) recognized early on CMC is not identical to face-to-face conversation (p. 1). Yates (1996) had previously gone even further, suggesting that CMC occurred in the absence of any field à la Halliday other than the text itself (pp. 45-46). Baron (2004) echoed these general sentiments, adding to them that the medium may change "the character of language produced in that medium," leading her to formulate perhaps the most prevalent typological framework in CMC for mediums, which includes two dichotomous parameters: synchronicity versus asynchronicity, and one-to-one versus one-to-many interactions (p. 398). The first parameter, synchronicity versus asynchronicity, refers to whether there is a reasonable expectation between the interlocutors that messages will be received and responses made immediately, as if in a face-to-face conversation. The second parameter, one-to-one versus one-to-many interactions, refers to whether locutors are sending messages that are meant to be received either by one person

[1] This is not to be confused with Eckert's (2012) description of three "waves" of sociolinguistics, though there comparison is apt.

or by many people.

Baron's (2004) typological framework for CMC is useful for understanding how comparable data produced through different mediums is as there may be different limitations placed on the forms of messages. For instance, Baron (2004) suggested that chatrooms, MUDs, and IM are all synchronous CMC mediums, whereas text messaging is asynchronous as one may not expect immediate replies to text messages (p. 398). The implication is that data collected from chatrooms and IM are more comparable than data collected from chatrooms and text messages.

Indeed, there are clear examples of the synchonocity or asynchronicity of a medium having an effect on how conversations progress. Baron (2004) herself noted a phenomenon in instant messaging in which multiple topics overlap as a result of both interlocutors being able to construct messages simultaneously without interrupting each other (p. 400), which would not happen in a less synchronous medium such as e-mail. Somewhat relatedly, she observed a prevalence of multiturn sequences, where each sent message is considered a turn. Baron (2004) argued that interlocutors face pressure to break up their messages into multiple shorter turns in order to hold the floor (p. 417), the idea being that, since messages are not seen until sent, longer single turn messages provide more time for the other party to chime in before the whole statement is ready to be sent. These sorts of phenomena suggest that one must be cautious when comparing the language used in one CMC medium versus another.

However, the boundaries between what should be considered synchronous and asynchronous are somewhat fuzzy. While chatrooms would presumably yield situations where interlocutors immediately reply to each other, this is not necessarily the case. IRC still exists today, though the way it is used may have changed over the last 20 years. For instance, at the time of this writing, the chatroom #nlp on the freenode network, which offers discussion and help with natural language processing, has a topic that states that it may take two to three hours to get a response to a question. The reason is that it is typical to be connected to an IRC chatroom without monitoring it closely, making it less synchronous than might be expected. Indeed, Baron (2004) herself acknowledged that IM users in her data were sometimes multitasking while connected and so did not always respond immediately (p. 419). Even in clearer cases of asynchronicity, such e-mail, there is still some imperfection in the classification as many people now own cell phones that receive e-mail notifications instantly wherever they are, allowing for relatively quick responses, though quick responses are not necessarily expected. This parameter is still useful, of course, since a medium can be thought of as more or less synchronous in terms of expectations – in our case, Twitter is relatively asynchronous but can be used from a cell phone that provides immediate notifications still – but there are caveats to keep in mind.

It is also important to consider the audience for messages. Baron's (2004) framework accounts for this with the one-to-one versus one-to-many parameter. Protypical examples of each might be e-mail for the former and blogs for the latter, although here again one finds some fuzziness in the categorization.

While e-mail might typically be a one-to-one medium, one-to-many messages are not unknown.

This same difficulty in categorization is present on Twitter, as well, where the most common sort of messages may be one-to-many, but there exist public directed messages and private directed messages. Again, the parameter here is still useful, but in this case, there is also an alternative: Bell's (1984) audience design framework. Indeed, Pavalanathan and Eisenstein (2015) employed this framework on Twitter, which yielded interesting results, as will be discussed further in the Methods section. Likewise, the audience design framework would have been useful when Androutsopoulos (2008b) had found "hip-hop slang" to be used more on German hip-hop discussion forums than on German hip-hop homepages and online magazines (p. 293). Both of these venues are indeed one-to-many when approached from Baron's (2004) typology, but the results are not the same. One could instead argue from the audience design framework that the audience for a homepage or magazine is far broader than for a discussion forum, leading to more standard language. What is to be taken away from this is that audience appears to impact the character of the language produced on Twitter, which can inform how we interpret results in the present study.

1.2 Orthographic variation

From early on in CMC research, it has been acknowledged that CMC is not only different from face-to-face language, as Paolillo (1999) recognized, but that it is also different from older forms of writing. Indeed, Tagliamonte and Denis (2008), speaking of instant messaging, described it as a "hybrid register", somewhere between speech and older forms of writing (p. 5). Part of this hybridity perhaps stems from CMC taking place outside of the auspices of social institutions, such as schools and book publishers, that extert overt social control on what is acceptable. This makes CMC a fruitful area for locutors to creatively manipulate linguistic features to various ends, one of those features being the focus of the present study: orthography. It is thus useful to review the literature on orthographic variation to better understand what functions and associations it has, as that is what we would like to analyze for ⟨lol⟩. We group these under four broad categories: those related to 1) grammar, 2) audience, 3) community or subsets of a community, and 4) pragmatics.

Before moving on to the possible functions and associations of orthographic variation, it should first be noted that orthographic variation is unlikely to simply be the result of poor spelling capability. Varnhagen et al. (2010), in their examination of instant messaging, compared participants scores on spelling tests to their use of non-standard spellings in IM. They found no relationship between the two, suggesting that non-standard spelling is not the result of a poor grasp of standard spelling. In fact, they found that non-standard spelling norms were acquired quite readily. For example, ⟨shoulda⟩ 'should have' was found but never forms like ⟨shulda⟩ (Varnhagen et al., 2010, p. 731). The implication is that non-standard spellers may actually be very good spellers if the

qualification of "good" in this case is defined as 'accurately follows norms'.

1.2.1 Connection to grammar

It is not unusual for variant spellings to be constrained by grammar. A very standard and overt example of this comes from French verbs. For most verbs, the singular first, second, and third person present forms of verbs are phonetically identical, yet the written forms vary anyway to agree with the subject. *Parler* 'to speak' is pronounced [paʀl] for all three singular subjects in perhaps most varieties of Hexagonal French,[2] but is written ⟨parle⟩ for first and third person and ⟨parles⟩ for second person. It is thus useful to consider other potential grammatical constraints on orthographic variation.

One such case can be found in Hinrichs and White-Sustaíta's (2011) work on Jamaican Creole, which has English as its superstrate. They found that speakers of the language living outside of Jamaica would vary between spellings such as ⟨mi⟩ and ⟨me⟩, both pronounced [mi], based on the syntactic function within the sentence, using the former as the subject pronun and the latter in all other cases. Hinrichs and White-Sustaíta (2011) argued that ⟨mi⟩ was limited to the subject pronoun function because this was a stereotypical feature of Jamaican Creole as English would have *I* in these cases. However, they note that it could also be said that this variation indicates a complex relationship to English and perhaps non-Creole-speakers (Hinrichs and White-Sustaíta 2011, as cited in Eisenstein, 2015, p. 165), making it not only grammatical constrained but possibly socially motivated.

Hinrichs and White-Sustaíta's example from Jamaican Creole is somewhat similar to Tatman's (2016) report of variation between ⟨work⟩ and ⟨werk⟩ among drag queens online. She compared collocations for the two and found that ⟨werk⟩ was typically used to express approval and ⟨work⟩ to express the more typical notion of doing work, leading her to conclude that these are in fact different lexical items (Tatman, 2016, pp. 163-164). Determing whether this involved two lexical items versus one polysemous lexical item is well beyond the scope of the present study, and it would be admittedly a weak argument to claim that this variation is purely constrained by syntax, but there is enough in this example to make the more grammatical interpretation possible, wherein ⟨werk⟩ is limited to particular syntactic positions.

Clearer examples of grammatically constrained orthographic variation come from Eisenstein's (2015) work on Twitter. He examined two orthographic variables that, on the surface, would appear to have a spoken phonological connection: (ing) as ⟨ing⟩ or ⟨in⟩[3] and (th) fortition, which he refers to as *th*-stopping. In both cases, grammatical constraints were found. The variable (ing) was more likely to be realized as ⟨in⟩ for verbs (p. 176), and the variable (th) was more likely to be realized as a glyph ⟨d⟩ when the spoken form would typically be voiced as opposed to having it realized as ⟨t⟩ when the spoken form would

[2] French spoken in France.

[3] This presumably would include ⟨in'⟩, though Eisenstein (2015) does not explicitly give that variant.

be voiceless (Eisenstein, 2015, pp. 170-171). There were other social factors found to be constraining these variables, but there is a strong argument here that linguistics factors were important for both.

1.2.2 Connection to audience

Another constraining factor that Eisenstein (2015) found for the realization of (ing) was the audience. As was already discussed in section 1.1, the intended or expected audience for a message seems to have an impact or the character of language used on CMC just as it does in spoken language, but the examples from Pavalanathan and Eisenstein (2015) and Androutsopoulos (2008b) were related to lexical variation instead of orthographic variation. Eisenstein (2015), on the other hand, found this some relationship for orthographic variation where ⟨in⟩ was the more frequent variant for @-messages on Twitter, meaning messages that were directed at a particular user instead of posted as general public statements (p. 176).

Additionally, Androutsopoulos (2008b) did generally look for respellings in his data, as well, also finding evidence of the importance of audience. On the German hip-hop website webbeatz.de, for example, he noted more colloquial spellings (i.e., those non-standard spellings that are related to colloquial speech) than in other areas of the same site (p. 297). The implication is that discussion forums are expected to be read by the more engaged members of the community whereas general web pages are more likely to be viewed by broader passers by, so to speak.

1.2.3 Connection to community membership

Perhaps the most common associations examined for linguistic variables in variationist studies are associations between variants and particular communities or particular segments of communities, and indeed, these associations exist for orthographic variation. For instance, Sebba (1998) examined spelling in British Creole and suggested that non-standard spellings are used both because they distance the language from its English superstrate and because there is no orthographic norm for British Creole. The very idea of non-standard in this case, then, is 'not as would be done for English'. Sebba (1998) provided examples such as ⟨Jameka⟩ and ⟨kool⟩ where standard English orthography would have ⟨Jamaica⟩ and ⟨cool⟩ (as cited in Androutsopoulos, 2000, p. 515). The idea is not simply to buck the superstrate but to make a statement against speakers of the superstrate,[4] to assert an independent identity from those speakers.

Androutsopoulos (2008a) also presented an example of using a particular orthographic variant on German hip-hop websites to signal membership in the

[4]Of course, it is possible that many British Creole speakers are also English speakers, but the presumption is that speakers identities are bound up in their language preferences.

particular community. In this case, ⟨z⟩ would be written where ⟨s⟩ was expected. What makes this example's importance particularly clear is that Androutsopoulos (2008a) quoted one of his informants as explicitly claiming that such ⟨z⟩ signaled an extreme dedication to hip-hop (pp. 12-13). It appears, then, that orthographic variation can be quite salient.

Eisenstein's (2015) once again proves useful in also showing community membership can be a determining factor in orthographic variation. When examining (ing), he found that tweets emanating from US counties with high population density and/or high Black populations were more likely to have tokens of ⟨in⟩ than those emanating from other counties (Eisenstein, 2015, p. 176). It is important to note that he did not know the claimed racial identities of these Twitter users, but the presumption is that they are likely to identify as Black or to at least be highly exposed to those who identify as Black in their daily lives.

Eisenstein's (2015) finding also begins to highlight the important of geographic location. Indeed, one should not confound virtual communities, those formed through consistent interaction online around shared interests (Castells 2000, as cited in Androutsopoulos, 2008b, p. 283), and physically centered communities, but this does not mean that a connection between the two is impossible. As McNeill (2018) showed for virtual communities and geographically defined communities in the Maritime Provinces of Canada, those who live near each other tend to converge online, as well (pp. 88-91).

As such, Jones (2015) found that the choice of non-standard spelling between ⟨nuttin⟩ and ⟨nun⟩ on Twitter, both attempts to represent *nothing* spoken with an intervocalic glottal stop, was constrained by geographic. Those tweeting from the north in the US preferred ⟨nuttin⟩, and those tweeting from the south in the US preferred ⟨nun⟩ (p. 424). It might thus be expected that northerners and southerners form separate virtual communities, as well, where different norms are established, as the spoken pronunciations in both areas would be the same.

There are also cases, though, where one might say that a virtual location is the constraining factor. Cherny (1995) offered somewhat anecdotal evidence of this in the early days of the internet. She analyzed the use of ⟨u⟩ and ⟨r⟩ for ⟨you⟩ and ⟨are⟩, respectively, finding that while these non-standard spellings did appear in MUDs, players of MUDs considered them to be forms that originated in IRC chatrooms (as cited in Paolillo, 1999, p. 2). If MUDs and IRC chatrooms are conceptualized as separate locations, then a important determinant here was likely virtual location.

Age also appears to be a determinant for the realization of orthographic linguistic variables. Schnoebelen (2012) analyzed the use of different emoticons, faces essentially drawn using alphanumeric characters. He looked particularly closely at those that included noses, such as ⟨:-)⟩, versus those that did not, such as ⟨:)⟩. He found that noseless emoticons were preferred more by older users of Twitter more so than younger users, though noseless emoticons were also associated with users who disregarded various orthographic standards by doing things such as repeating letters or leaving out apostophes (pp. 122-124).

Baron (2004) also analyzed emoticons in instant messaging with the addi-

tion of analyzing abbreviations and acronyms that she considered unique to CMC. She found that emoticons were almost exclusively limited to female participants in her data (pp. 415-416). This result was reproduced in Varnhagen et al.'s (2010) work on instant messaging, though the results were not as quite as lopsided as in the earlier study (pp. 728-729).

1.2.4 Connection to pragmatics

Androutsopoulos (2000) long ago acknowledged that orthographic variants can perform pragmatic work. He argued that they "signal certain attitudes or evoke certain frame of interpretation by establishing a contrast to the text's spelling regularities or to the default spelling of a linguistic item" (Androutsopoulos, 2000, p. 517). An example of such a contrast can be found in his study of German punk fanzines, essentially low budget magazines made by enthusiasts. Androutsopoulos (2000) noted that the typical spelling of the term *fanzine* was ⟨fanzine⟩, which in fact bucks against standard German spelling in which all nouns are capitalized. Nevertheless, this was the established norm within fanzines themselves. As a result, those familiar with the medium would sometimes produce hypothetical quotations from Germans who were not familiar that included the spelling ⟨Fähnziehn⟩, much more in line with standard German orthography, with the result being mockery of the latter's ignorance (Androutsopoulos, 2000, p. 526).

The example of a pragmatic factor in the realization of orthographic variables given in this section as well as the examples of other factors from the preceding sections provide some insight into what the possible determinants variables are. Unsurprisingly, the range is as broad as that which can be found for linguistic variables in spoken language. While it will not be possible to look at each and every factor in analyzing (lol) due to the limitations of what we know about Twitter users without directly interacting with them, we will endeavor to include as many factors as possible in line with the exploratory nature of this study.

1.3 Previous work on (lol)

The orthographic variable being analyzed in the present study is (lol), which at least originally was an acronym that stood for 'laugh out loud'. This acronym is thought to have originated in English language chatrooms in the 1980s (McCulloch 2019, as cited in Schneier, 2021, p. 4). However, it has found its way into other languages, as well. Liénard (2014) documented (lol) being used by early adopters of the internet in Mayotte, an island nation near Reunion in Africa. What is notable about this case is that the internet had only effectively been accessible started in 2012 (p. 154), and English was not a local language nor an official language in Mayotte, those being Shimaroe and Kibushi for the local languages and French for the official language (p. 158). Likewise, McNeill (2018) documented significant use of *lol* in what would otherwise be viewed as French-language tweets on Twitter. It appears possible then that (lol) has

become something of an internet-language acronym rather than an English-language acronym, though almost all the work done on it has been focused on English.

Despite its penetration even into other languages, *lol* is not overall a frequent lexical item when compared to other lexical items. In Baron's (2004) IM data, *lol* made up only 0.6% of the total words. Similarly, in Tagliamonte and Denis's (2008) IM data, it made up 0.41% of the words, and 0.35% in Schneier's (2021) cell phone data. Schneier (2021) also found *lol* to be most frequent turn-initially and turn-finally (p. 14). This low overall frequency might be damning for quantitative analyses, but it is perfectly in line with Zipf's law, which claims that the most frequent lexical items in any corpus will be exponentially more frequent than those ranked even only slightly lower in frequency, and indeed *lol* is highly frequent when compared to other CMC-specific items.

2 Methods

3 Results

4 Discussion

References

Androutsopoulos, J. (2000). Non-standard spellings in media texts: The case of German fanzines [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9481.00128]. *Journal of Sociolinguistics*, *4*(4), 514–533. https://doi.org/https://doi.org/10.1111/1467-9481.00128

Androutsopoulos, J. (2008a). Potentials and Limitations of Discourse-Centred Online Ethnography. *Language@Internet*, *5*(8), 1–20. http://www.languageatinternet.org/articles/2008/1610

Androutsopoulos, J. (2008b). Style online: Doing hip-hop on the German-speaking Web. In P. Auer (Ed.), *Style and Social Identities: Alternative Approaches to Linguistic Heterogeneity* (pp. 279–317). De Gruyter, Inc. Retrieved February 7, 2019, from http://ebookcentral.proquest.com/lib/ugalib/detail.action?docID=364724

Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, *18*(2), 135–160. https://doi.org/10.1111/josl.12080

Baron, N. S. (2004). See You Online: Gender Issues in College Student Use of Instant Messaging. *Journal of Language and Social Psychology*, *23*(4), 397–423. https://doi.org/10.1177/0261927X04269585

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204. https://doi.org/10.1017/s004740450001037x

Cherny, L. M. (1996). The MUD register: Conversational modes of action in a text-based virtual reality., 1. Retrieved February 12, 2021, from https://www.elibrary.ru/item.asp?id=5621106

Eckert, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*, *41*(1), 87–100. https://doi.org/10.1146/annurev-anthro-092611-145828

Eisenstein, J. (2013). Phonological Factors in Social Media Writing. *Proceedings of the Workshop on Language in Social Media (LASM 2013)*, 11–19.

Eisenstein, J. (2015). Systematic patterning in phonologically-motivated orthographic variation. *Journal of Sociolinguistics*, *19*(2), 161–188. https://doi.org/10.1111/josl.12119

Hong, L., Convertino, G., & Chi, E. H. (2011). Language Matters in Twitter: A Large Scale Study. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 4.

Ilbury, C. (2020). "Sassy Queens": Stylistic orthographic variation in Twitter and the enregisterment of AAVE. *Journal of Sociolinguistics*, *24*(2), 245–264. https://doi.org/10.1111/josl.12366

Jones, T. (2015). Toward a Description of African American Vernacular English Dialect Regions Using "Black Twitter". *American Speech*, *90*(4), 403–440. https://doi.org/10.1215/00031283-3442117

Kim, S., Weber, I., Wei, L., & Oh, A. (2014). Sociolinguistic Analysis of Twitter in Multilingual Societies. *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, 243–248. https://doi.org/10.1145/2631775.2631824

Liénard, F. (2014). Les communautés sociolinguistiques virtuelles. Le cas des pratiques scripturales numériques synchrones et asynchrones mahoraises. *Studii de lingvistică*, *4*, 145–163.

McNeill, J. (2018). *LOL sur Twitter: Une approche du contact de langues et de la variation par l'analyse des réseaux sociaux* (Master's thesis). Université du Québec à Montréal. Montreal, QC. Retrieved December 8, 2018, from https://archipel.uqam.ca/11948/

Paolillo, J. C. (1999). The virtual speech community: Social network and language variation on IRC. *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, *Track2*, 10 pp.–-. https://doi.org/10.1109/HICSS.1999.772680

Paolillo, J. C. (2001). Language variation on Internet Relay Chat: A social network approach. *Journal of Sociolinguistics*, *5*(2), 180–213. https://doi.org/10.1111/1467-9481.00147

Pavalanathan, U., & Eisenstein, J. (2015). Audience-Modulated Variation in Online Social Media. *American Speech*, *90*(2), 187–213. https://doi.org/10.1215/00031283-3130324

Schneier, J. (2021). Digital Articulation: Examining Text-Based Linguistic Performances in Mobile Communication Through Keystroke-Logging Anal-

ysis [Publisher: Frontiers]. *Frontiers in Artificial Intelligence*, *3*. https://doi.org/10.3389/frai.2020.539920

Schnoebelen, T. (2012). Do You Smile with Your Nose? Stylistic Variation in Twitter Emoticons. *University of Pennsylvania Working Papers in Linguistics*, *18*(2), 116–125. https://repository.upenn.edu/pwpl/vol18/iss2/14

Stewart, I., Chancellor, S., De Choudhury, M., & Eisenstein, J. (2017). #Anorexia, #anarexia, #anarexyia: Characterizing online community practices with orthographic variation. *2017 IEEE International Conference on Big Data (Big Data)*, 1–9. https://doi.org/10.1109/BigData.2017.8258465

Tagliamonte, S. A., & Denis, D. (2008). Linguistic Ruin? Lol! Instant Messaging and Teen Language. *American Speech*, *83*(1), 3–34. https://doi.org/10.1215/00031283-2008-001

Tatman, R. (2016). "I'm a spawts guay": Comparing the Use of Sociophonetic Variables in Speech and Twitter. *University of Pennsylvania Working Papers in Linguistics*, *22*(2), 161–170. http://repository.upenn.edu/pwpl/vol22/iss2/18

Varnhagen, C. K., McFall, G. P., Pugh, N., Routledge, L., Sumida-MacDonald, H., & Kwong, T. E. (2010). Lol: New language and spelling in instant messaging. *Reading and Writing*, *23*(6), 719–733. https://doi.org/10.1007/s11145-009-9181-y

Yates, S. J. (1996). Oral and Written Linguistic Aspects of Computer Conferencing: A Corpus Based Study. In S. C. Herring (Ed.), *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives* (pp. 29–46). John Benjamins Publishing Company. Retrieved February 11, 2021, from http://ebookcentral.proquest.com/lib/ugalib/detail.action?docID=680383