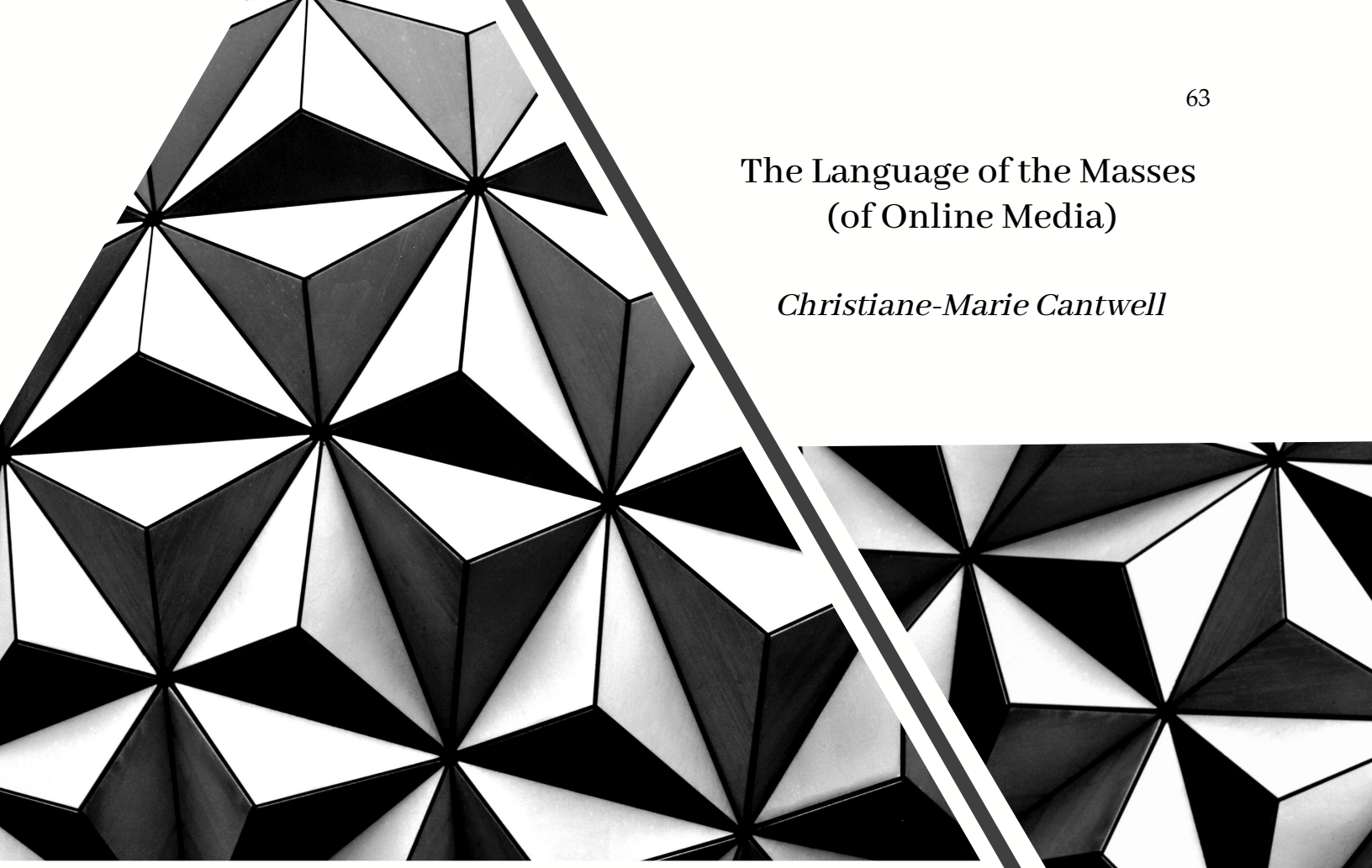
The Language of the Masses (of Online Media)

Christiane-Marie Cantwell



Abstract

This paper aims to counter the belief that the language used in technology-based mass media is rapidly becoming a homogenous entity. Through a review of existing literature on web-based language complemented with folk examples, this paper illustrates that linguistic diversity is fostered in technology-based mass media in two ways. Firstly, online media creates new virtual communities of practice that develop their own linguistic variants. Secondly, technology-based media, such as social media, allows individuals to transcribe local variants of spoken language as markers of identity, thereby transferring offline linguistic variation online. Once it is established that linguistic variation is maintained in online linguistic forms, Squires's theory of enregistrement is used to explain the existence of the assumption that online languages change faster than offline languages. This theory, in conjunction with previous research on the relationships between spoken language and virtual language in a set community of practice, is then applied to counter the assumption and illustrate that technology-based mass media do not increase the speed of language change.

Essential to the field of sociolinguistics is the tenet that human language constantly changes. However, this does not necessarily mean that such changes occur at the same rate. Indeed, through technology, language seems to change quickly and drastically; new linguistic forms emerge, different from their standard counterparts, and conventionally, these rapidly evolving globalized forms of language are understood to limit linguistic variation. In this essay, I challenge the idea that homogeneous language forms rapidly supersede all regional and social variants online. Instead, I show that technologically-based mass media in fact encourages linguistic variation at a normal rate, demonstrated through three themes: the advent of new internet-based communities of practice, users' transcription of regional varieties of language online as virtual markers of identity, and the rate of change of language online being equal to the rate offline.

The idea of community of practice is integral to the development of several variations of language online. Communities of practice are "a collection of people who engage on an ongoing basis in some common endeavour" (Eckert, 2006). Conducting sociolinguistic research on the basis of communities of practice rather than speech communities allows for a fluidity in the members' identity, who define themselves through their groupings, and can belong to many groups at once (Eckert, 2006). The traditional view of communities of practice, especially in Eckert's study, relegates them to the world offline, usually seeing their effect on speech[1]. However, the idea can easily transfer to virtual mass media. Take, for example, Twitter or Tumblr, both microblogging sites, which operate based on spontaneous creation of content and the subsequent interactions between users. Fandoms, as defined as groupings of people online positioned around common interests, such as television shows, video-games, movies, etc., develop specific lexical tokens that relate to their web-based linguistic community, which people outside of the group would not use. Importantly, online communities of practice originate in the virtual world, with members having little to no physical contact with each other, at least initially. Their membership is performed solely on technologically-based mass media. Eventually, a community will grow to such an extent that the contact begins to occur off-line, as does the specific language they have been using.

A key example of a fandom constituting a sociolinguistic group is "Bronys", adult (usually male) fans of the show "My Little Pony" (MLP), who have developed an extraordinary community of practice online. This community is entirely focused around mass media (i.e. the television show, and the products of it on the Internet such as memes) and their language reflects this, especially lexically. Self-styling as "Bronyspeak", their communication is conducted almost exclusively in English, with references to the show mixed in. It is in no way a new language, but just a different register, solely created online. According to Whatisabrony.com (2017), a fan-run website, words specific to this community include many references to MLP, like "20% Cooler", which can either be a marker of a positive or negative situation. This illustrates a certain semantic shift depending on the situation in which the utterance occurs. It also includes euphemisms, such as "hay" for "hell", and portmanteaus like "brofist" which indicates an online fist bump between two members of the community and, indeed, "brony", a mix of "bro" and "pony". This linguistic variation is promoted by the virtual mass-media, since this terminology was first only used

[1] see Eckert (1989) for the study of social groups in a Detroit high school

online. However, Bronys have taken their variation of language off-line and into spoken communication. The community began on the Internet, but grew to such extents that fans started to have conventions—much like "Trekkies", superfans of Star Trek. The most popular is BronyCon, an annual conference, which reached peak attendance in 2015 with 10,001 attendees (BronyCon, 2019). In this setting, the language preferred online by this community transfers into spoken language. One can hear this linguistic variation in footage of the convention; in one video in particular by LittleshyFiM on YouTube, there are multiple instances of a "brohoof", and even attendees correcting each other when one person uses a non-Brony word where a Brony one is available (for example, correcting "bagpipe" to "lluviduphone", a term which makes reference to a specific episode of MLP). The latter illustrates norm-enforcement mechanisms in this group and lends credence to the idea that they are indeed speaking a variant of Standard English, which has its own norms. Further, the people in this video know each other, but only virtually. They are not from the same region and come together only for this event, yet their lexical entries are remarkably homogeneous, having been influenced by the community of practice. Even further, their address forms are their online profile names, which often do not reflect their given names. In these ways, we can truly see how technologically-based social media creates new specific communities of practice.

Another instance of an online community of practice is illustrated in Paolillo (1999). In this case, the setting is a global online chat platform, Internet Relay Chat (IRC), under the channel "#india", where members are united by their engagement and interest in this tag. This study is now twenty years old, and technology has much evolved since then, yet Paolillo's data illustrates strong communities of practice online, even towards the beginning of digital mass media. Paolillo analyzes the online linguistic interactions of the people in the "#india" channel to measure any correlation between network strength and non-standardized use of language, both English and Hindi, in the vein of Milroy & Milroy (1978). He found that there was no simple link between the two and argues that there are actually sub-groups within the channel which are partitioned primarily based on their core-ness, as measured by the number of operators (IRC moderators) in that subgroup.

One could expect that under the channel "#india", there would be a linguistic trend towards a unified form of Hindi, yet this is not the case. Instead, each sub-group follows different code-switching norms, and though the author does not explicitly state it, these sub-groups are moving away from the concept of social networks and towards that of communities of practice. The discourse data suggests that each sub-group, governed independently by its own vernacular rules, belongs to a different community, like Indian expats or nationals. This could intrinsically explain the inherent variation based on established linguistic patterns within these groups in spoken language, though the author does not capitalize on this. Indeed, through communities of practice, the variation can be more easily explained. Unknowingly, Paolillo illustrates how people group themselves based on common interest, even within another group, to develop their own linguistic norms and identity with no external influence (as the members of each sub-group do not know each other offline). Though these data were collected early in the age of the Internet, there is an emerging pattern: the idea of social networks as sociolinguistic units is one that has little to no standing in the digitized – and importantly, anonymous – world of today, where creating social ties is much harder. There are none of the classic norm enforcers, like knowing a

person's name, face, or occupation to govern their creation. All users know about each other are their usernames and the language they are using, and as such both become key players in defining identity online. Essentially, Paolillo exposes communities of practice who are incidentally defined by a geographical location, but may or may not be physically there.

In fact, regional linguistic variation has transferred into technologically-based mass media, allowing users to affirm this part of their identity and creating more variation in linguistic forms. Through the rise of social media and geotags, researchers are now able to track both the language employed, as well as the place the content was created in. Gonçalves and Sanchez (2016) mapped Spanish dialects through a corpus of 106 million tweets in Spanish, all geolocated, from 2010 to 2015. They identified 2 supra-dialects, a more popular, relatively homogeneous one used across the world in urban areas (A) and a heterogeneous one which is popular in more rural areas (B). The latter is divided geographically into four zones: Spain, North America, South America (general), and along the Andes. What is remarkable about the findings of the second supra-dialect are that they match the commonly drawn dialect boundaries for Spanish (Gonçalves and Sanchez, 2016: 70). This goes directly against the idea that online media is destroying linguistic diversity. Though Gonçalves and Sanchez argue that dialect A is a globalized form of Spanish, erasing the distinct regional character of cities because of overt prestige, they fail to recognize that the usage of dialect B may not represent a slowness to adopt dialect A in the rural regions, but rather an act of covert prestige. Using dialect B may be a conscious choice in preserving regional character and promoting it online as a way for people to mark their identity on the Internet, where other social cues, such as clothes, accent, and mannerisms, are lost.

This effect is not limited to Spanish and can be seen in English as well. For a folk example, "Scottish Twitter", a social media phenomenon, is the Scottish dialect of English transcribed as it would be spoken. By doing so, users of the variant are establishing their regional identity—and it is not simply a few people doing so, but a large community. A tweet by @sdel6795 (2018), a young man from Hamilton, Scotland, is representative of the variant:

> "Ma das just caught me running up the stair wae a couple of the good chocolate biscuits n he's complaining tae ma maw saying 'he thinks he's a major player in this hoose' Hahahaha"

It is clearly not Standard English spelling, and there are several components that differ from any standard form of English. Gonçalves and Sanchez (2016) saw that Spanish online varies based on the existing boundaries of existing regional dialects, and Scottish Twitter illustrates the same finding for English. The use of "biscuit" for "cookie", for example, clearly indicates the regional dialect as separate from North American English. Further, the usage of "ma" for "my", "hoose" for "house", and "tae" for "to", illustrates the phonetic difference in the dialect. In this tweet, the user highlights his spoken usage of a schwa in the word; he would not say (tu), but (tə), and to represent that, he changes the spelling to best (one assumes) represent his dialect.

Past this quick folk analysis of one tweet, there have been systematic studies of English on Twitter: Grieve (2018), which explores lexical innovation and diffusion on English Twitter, and Eisenstein & O'Connor, Smith, Xing (2014), which maps lexical tokens. Both studies make use of the geolocation of tweets, and gathered an enormous corpus—980 million and 170 million tweets respectively. They relate in depth the methodology of analyzing this enormous corpus and map lexical differences. Eisenstein's findings were that language online mirrors the existing

dialectal differences in American English. By tracking the use of "ion" (instead of "I don't") and "ard" (instead of "alright") among 3 other lexical tokens in American-generated tweets, they were able to not only reproduce existing dialectal boundaries (like "ard" being used only in Philadelphia and "ion" used in the South-East), but also found that race affects online language, much like it does spoken language. African American Vernacular has transferred to Twitter, and areas which are demographically similar are more likely to be linguistically similar. On social media, people are using language that, while written, still divides them into groups of existing English dialects—evidence of linguistic variation in digital mass media. These findings are reproduced in Grieve (2018). This study focuses more on the diffusion of new words on Twitter, finding that new tokens, like "bruh", meaning "bro", and "on fleek" do not originate online, but are diffused geographically through technologically-based mass media, following consistent diffusion patterns in which words will most often only be diffused in the dialectal region from which they emerge. Grieve identifies five regions in the United States within which new words can occur. Words will usually be restrained to that region, unless the word, such as in the case of "bae", meaning "significant other", gains extreme popularity. The five identified regions are the West Coast, from California to Arizona; the Deep South, centered around Atlanta; the Northeast, where New York is a hub; the Mid-Atlantic, around Washington D.C. and Baltimore; and the Gulf Coast, positioned in Louisiana and Texas. These are all dialect regions, with specific virtual manifestations. The author of this paper does not compare these proposed dialect boundaries with those from the Atlas of North American English (Labov, Ash & Boberg, 2006), which could have bolstered the claim. Indeed, the Mid-Atlantic and Northeast region match exactly with the finding from the Atlas (Labov, Ash & Boberg, 2006). It should be noted, however, that the regions do not align in the South: Grieve singles out the Gulf Coast as different from the rest of the South, whereas the Atlas considers it part of the General South. Yet, it must be understood that Grieve has only studied lexical entries and does not have access to phonological data, unlike the Atlas. Nonetheless, words diffuse regionally online like spoken language does; on social media, users are still able to differentiate themselves based on their language, promoting linguistic variation, especially lexically.

Though these last two studies do support the idea that technologically-based mass media is not responsible for establishing a homogeneous standard variety of English, they forget to account for an important subtlety that may have skewed their data. Neither consider the authenticity of the tweets themselves. Though the fraction may be small, it is important to account for the nuance that among the millions of tweets analyzed, the use of non-standard language is likely to, in some cases, not be an affirmation of identity, but rather something intended for comedic effect. By analogy, in sociolinguistic research of spoken AAVE, a white man who uses African American Vernacular English (AAVE) for a stand-up routine and not as his actual dialect would not be taken as an authentic speaker of AAVE, and therefore his speech would not be taken to be a true representation of the language. A study of language on the Internet should not treat this issue differently. There must be a certain amount of attention spent to this nuance, and it is one of the downfalls of having millions of data points. It would be extremely time-consuming to make this distinction for each tweet, but researchers could at least acknowledge that it could affect their results, which neither Grieve (2018) or Eisenstein & O'Connor, Smith, Xing (2014) do.

Contrastingly, Moll (2018) constantly asks whether or not their results are authentic in a thorough study of Jamaican Creole (JC) and its usage in web discussion forums. This is especially important for a language that has a history of being co-opted and reproduced in mass media, like in a 2013 Volkswagen commercial[2]. Moll (2018) finds that there is a conventionality in the non-standard orthography of JC in web-based media, much like Androutsopoulos (2000) remarked in the study of German punk zines, where orthography is systematically non-standard, through norms of covert prestige in specific communities of practice. Much like Androutsopoulos (2000), Moll (2018) identifies variables that are consistently realized as non-standard forms, but which have become the standard in JC forums. For example, (aw) is used to indicate the phoneme that would be /ɔ/ or /ɑ/ in Standard North American English, either word-finally or before a rhotic consonant, as in "lawd" (meaning "lord") and "naw" ("not"). This illustrates the difference between the creole and its lexifier, English: the creole only has 5 monophthongs, and there is no difference between /ɔ/ or /ɑ/ (Harry, 2006). Further, it also illustrates the non-rhoticity of JC mapped to written language. JC is governed on the Internet by community-chosen conventions, not by supranational entities (like the Académie Française for French) or language planning agencies. Social media have not forced speakers of JC to use another language to communicate online; in fact, they have created a platform for speakers to declare their identity in the faceless void of the Internet.

As has been demonstrated above, the Internet fosters linguistic variation by creating new communities of practice who develop their own linguistic forms and by providing platforms where regional variations and underrepresented languages can be written down and used as identity markers. Yet, there is a common negative opinion of web-based linguistic forms, wherein there is variety, yet these are appearing far too quickly and seem to supersede the standard. Squires (2010) argues that this is because society has gone through the process of enregisterment of "web" languages far faster than it does with spoken forms. Enregisterment is the mechanism by which new sets "of linguistic features conceived as distinctive, imbued with social meaning linked to social personae, and linked to what are perceived as distinct varieties of language" are acknowledged and accepted into the norms of a language (Squires, 2010: 3). It is the process by which the new features of online languages are internalized by speakers. Squires illustrates how, because digital forms are written, they are defined much more quickly as sub-standard variants of the norm because of the ease in comparing permanent forms of language to one another. Unfortunately, Squires assumes that any language on the Internet is a completely new sub-standard dialect, especially in English. However, that is not the case. As seen above, many regional forms written down online are forms that exist – and originate – in spoken language. Outside of mass media, they are not necessarily thought of as being rapidly changing or even degraded forms. It is then the medium in which the language is presented that alerts speakers' perception of a change, without them realizing that these changes are actually occurring simultaneously offline. The language used in mass media is oftentimes a non-standard form, which for the first time is appearing in written communication. These languages have existed for years, but it is the fact that these differences are presented on mass media that allows people to recognize them faster. There is no accelerated shift.

[2] see Lopez & Heinrichs (2017) for the commodification of JC stereotypes in media

In this vein, Tagliamonte & Denis (2008) show that speech generated by online communication is actually more conservative than what is spoken. In a study of teenagers in Toronto, Tagliamonte tracked speech through instant messaging (IM). The results show that in all four grammatical variables analyzed in the study (intensifiers, quotative systems, future temporal reference, and modals of necessity), compared to results of Tagliamonte's past research into their spoken forms (Tagliamonte, 2008; Tagliamonte & D'arcy, 2004a; Tagliamonte & D'arcy, 2007b; Tagliamonte, 2007), the English used in IM was more conservative, varied, and standardized than what the teenagers would actually produce in speech. Tagliamonte's study is important in highlighting that indeed, through technology, language change towards homogenization has not been sped up, but variation has actually been preserved. By studying young adults who are already friends and belonging, one can imagine, to the same social groups and from the same region, Tagliamonte does not have to account for differences in online language created by the need to linguistically reinforce identity. Yet, herein lies one of the weaknesses of the study as well, where the results represent only the online language of one community of practice, in one dialect region. Further, it only considers private communication between a small group of people, so it cannot readily apply to all Internet-based communication, such as social media, which is more formal due to its public nature. Still, the core results remain stable: language online is a representation of what is actually spoken and it may, in effect, be more conservative. The pervasive idea that technologically-based mass media is swiftly homogenizing language is but a reaction to the rapid enregisterment of web languages, as put forth by Squires (2010).

There is no one dialect that will supersede all others and erase any kind of linguistic variation in technologically-based mass media. The Internet allows people who do not live close to one another to form virtual communities of practice, which increase linguistic variation. It prompts speakers of non-standard forms of languages to assert their linguistic diversity through non-standard orthography and lexical terms, which can be shown to follow pre-existing dialect boundaries. Further, language is not changing faster than it ever has, though it is changing. It is simply the novelty of the medium which causes people to notice the change. Language is an identity marker, whether online or offline. Yet in technologically-based mass media, the way one communicates takes a much bigger role, for there is not much else by which someone can define themselves. The new forms of mass media which have appeared in the last two decades have allowed people to communicate faster and better than ever before without being forced to compromise their specific linguistic identity through standardized writing forms. Though technology has led to globalization on many scales, language seems to not be one of them.

References

Androutsopoulos, Jannis. (2000). Non-standard spellings in media texts: The case of German fanzines. *Journal of Sociolinguistics* 4 (4): 514-533. DOI: 10.1111/1467-9481.00128

Delany, Sean [@sdel6795]. (2018, April 4). *Ma das just caught me running up the stair wae a couple of the good chocolate biscuits n he's complaining tae ma maw saying 'he thinks he's a major player in this hoose' Hahahaha*[Tweet]. Twitter. https://twitter.com/sdel6795/status/981581823959592961

Eckert, Penelope. (2006). *Encyclopedia of Language and Linguistics (2).* Edited by E. K. Brown and Anne Anderson. Boston: Elsevier.

Eckert, Penelope. (1989). *Jocks and Burnouts : Social Categories and Identity in the High School.* New York: Teachers College Press Columbia University.

Eisenstein J, O'Connor B, Smith NA, Xing EP. (2014). Diffusion of Lexical Change in Social Media. *PLOS ONE* 9(11). https://doi.org/10.1371/journal.pone.0113114

Gonçalves, Bruno, and David Sanchez. (2016). Learning about Spanish Dialects through Twitter. *Revista Internacional De Lingüística Iberoamericana*14, 2 (28): 65-75. http://www.jstor.org/stable/26379776

Grieve, Jack, et al. (2018). Mapping Lexical Innovation on American Social Media. *Journal of English Linguistics*46 (4): 293-319. https://doi.org/10.1177/0075424218793191

Harry, Otelemate G. (2006). Jamaican Creole. *Journal of the International Phonetic Association*36 (1). https://doi.org/10.1017/S002510030600243X

Labov, William and Charles Boberg and Sharon Ash. (2008). *The Atlas of North American English : Phonetics, Phonology and Sound Change*. Berlin: Walter de Gruyter. http://www.atlas.mouton-content.com/

[LittleShy FiM]. (2018, July). *BronyCon 2018 -- Vlog & Footage Reel* [Video]. YouTube. https://www.youtube.com/watch?v=khK9h4rkinI

Hinrichs, Lars and Quiana Lopez. (2017). "C'mon, Get Happy": The Commodification of Linguistic Stereotypes in a Volkswagen Super Bowl Commercial. *Journal of English Linguistics*.45 (3): 1-27. DOI: 10.1177/0075424217702106

Moll, Andrea. (2015). Jamaican Creole Goes Web : Sociolinguistic Styling and Authenticity in a Digital "Yaad." *Creole Language Library.* Amsterdam: John Benjamins Publishing Company.

Tagliamonte, Sali. (2008). So Different and Pretty Cool! Recycling Intensifiers in Toronto, Canada. *English Language and Linguistics*12 (2). Cambridge University Press: 361–94.

Tagliamonte, Sali and Alexandra D'Arcy. (2004). He's like, She's like: The Quotative System in Canadian Youth. *Journal of Sociolinguistics* 8: 493-514.

Tagliamonte, Sali, and Alexandra D'Arcy. (2007). The Modals of Obligation/Necessity in Canadian Perspective. *English World-Wide*28: 47–87.

Tagliamonte, Sali and Derek Denis. (2008). Linguistic ruin? LOL! Instant Messaging and Teen Language. *American Speech*83 (1): 3–34. https://doi.org/10.1215/00031283-2008-001

Squires, Lauren. Enregistering Internet Language. *Language in Society*39, no. 4 (2010): 457-92. http://www.jstor.org/stable/40925792

Whatisabrony. (2017, October 29). Popular Brony Terms, Phrases, and Words – Glossary. *WIAB Blog*. whatisabrony.com/blog/2017/09/03/popular-brony-terms-phrases-and-words-glossary/