



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

“Sassy Queens”

Citation for published version:

Ilbury, C 2020, "Sassy Queens": Stylistic orthographic variation in Twitter and the enregisterment of AAVE', *Journal of Sociolinguistics*, vol. 24, no. 2, pp. 245-264. <https://doi.org/10.1111/josl.12366>

Digital Object Identifier (DOI):

[10.1111/josl.12366](https://doi.org/10.1111/josl.12366)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Journal of Sociolinguistics

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



“Sassy Queens”: Stylistic orthographic variation in Twitter and the enregisterment of AAVE

Christian Ilbury 

Queen Mary University of London, UK

Correspondence

Christian Ilbury, Queen Mary University of London, Mile End Road, London, UK, E1 4NS.

Email: c.i.ilbury@qmul.ac.uk

Funding information

Economic and Social Research Council, Grant/Award Number: ES/J500124/1

Abstract

Recent computational sociolinguistic analyses of social media have emphasized the potential of using orthographic variation as a proxy for speech, thereby permitting macro-level quantitative studies of regional and social variation (e.g. Eisenstein, 2015). However, the extent to which stylistic variation may affect these analyses remains largely unexplored. In this paper, I explore how authors use variant spellings stylistically to deploy personae and characterological figures (Agha, 2003), by examining the presence of African American Vernacular English (AAVE) features in a corpus of 15,804 tweets extracted from the timelines of 10 gay British men. I argue that the stylization of AAVE signals the development of a very specific persona—the “Sassy Queen”—which relies on an essentialized imagining of Black women as “fierce” and “sassy.” Concluding, I emphasize the value of micro-level analyses in complementing quantitative analyses of linguistic variation in social media.

KEYWORDS

AAVE, computational sociolinguistics, orthographic variation, social media, stylization, Twitter

1 | INTRODUCTION

In recent years, the social sciences have witnessed a methodological shift towards utilizing computational tools to automatically extract and analyse complex datasets. Consequently, researchers have been

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Author. *Journal of Sociolinguistics* Published by John Wiley & Sons Ltd.

able to examine social phenomena across populations far greater in size than previously thought possible. These developments have not bypassed sociolinguistics. Increasingly, Natural Language Processing (NLP) tools have been used to examine linguistic phenomena in “big” datasets, including multimillion word corpora scraped from social media sites Twitter and Reddit (e.g. Wieling et al., 2016).

Along with the increasing convergence of sociolinguistics and computer sciences is the emergence of a distinct theoretical enterprise—“Computational Sociolinguistics” (henceforth CS) (Nguyen, Seza Doğruöz, Rosé, & de Jong, 2016)—which seeks to integrate the methodological approaches of Computational Linguistics (henceforth CL) and the theoretical frameworks of the variationist paradigm. Unlike in CL where variation is characterized as “noise,” CS conceptualizes variability in text as “social and cultural data” (Nguyen et al., 2016: 537), which can be analysed in relation to the author's attributes, such as their age or gender. In this line of inquiry, sociolinguistic theories and concepts developed for spoken language phenomena have been applied, often unproblematically, to written data (e.g. Eisenstein, 2015). These analyses have tended to identify correlations between the use of a particular orthographic feature and some macro-level category, suggesting that written variation patterns in ways that are systematically similar to speech (e.g. Eisenstein, 2015). As a consequence, scholars have emphasized the potential of using social media data as a proxy for spoken language, permitting the creation of dialect maps based on geotagged tweets and analyses which examine the orthographic representation of some spoken language feature across different locations (e.g. Eisenstein, 2015, 2017; Grieve, Nini, & Guo, 2018; Jones, 2015; Tatman, 2015).

In this paper, I explore and assess the extent to which macro-level social factors can explain patterns of written variation in social media. Specifically, I examine the orthographic representation of features typically found in African American Vernacular English (henceforth AAVE) in 15,804 tweets extracted from the social media site Twitter, which originate from the profiles of ten White gay men based in the UK. I show that, across all linguistic levels, features typically found in AAVE are frequently used in tweets by these individuals.

In taking a micro-level approach in analysing orthographic variation, I demonstrate that the practice typical of CS which identifies correlations between orthographic variants and macro-level factors is unable to account for such patterns. Rather, I claim that users, in this case gay men, use non-standard spellings which are ideologically associated with a particular identity to project personae, stances, and related characterological figures (Agha, 2003) to achieve certain interactional ends. Specifically, my argument is that by selectively using features of AAVE, the users draw on stereotypes of Black women to present themselves as “sassy” or “fierce”—qualities which have become appreciated within certain subcultures of the gay community (Barrett, 2017; Calder, 2019a,b). The appearance of AAVE features in the dataset can therefore be explained in terms of “stylization” (Coupland, 2007), wherein the use of this style indexes a valuable persona that has become recognizable in mainstream gay culture which I refer to here as “the Sassy Queen,” where “queen” refers to the gay slang term denoting a flamboyant, effeminate gay man (Graf & Lippa, 1995).

In the sections following, I briefly introduce the sociolinguistic research on Computer Mediated Communication (CMC) and the subsequent development of CS before introducing the notion of stylization. I then go on to examine the intersection of sexuality, AAVE, and style in the corpus. Finally, I explore the implications that this analysis has for a “sociolinguistics of social media” (Androutsopoulos, 2016) and CS.

2 | CMC AND COMPUTATIONAL SOCIOLINGUISTICS

Early sociolinguistic analyses of CMC tended to focus on patterns of online communication that could be attributed to broad social categories, such as identifying gender differences in language choice

(Herring, 2000). As part of this endeavour, which Georgakopoulou (2006) refers to as “first-wave” CMC research, scholars focussed mainly on the level of orthography, developing taxonomies of “distinctive” CMC features that characterized the “language of emails” or “features of Instant Messaging” (Baron, 1998; Crystal, 2004). These accounts often provided deterministic explanations of digital communication, with the medium perceived to be the main constraint on language use (Squires, 2010).

Contemporary sociolinguistic CMC research, however, has moved beyond this determinist agenda, instead seeking to understand the diversity of ways in which users appropriate the multimodal affordances of platforms for specific interactional purposes (Georgakopoulou & Spilioti, 2016). In this vein, scholars have sought to examine the discourse-level practices of users to examine how meaning is discursively created in digital contexts (e.g. Androutsopoulos, 2015).

Research on digital contexts in CL, on the other hand, has mostly sought to model the structural and informational dimensions of language (Nguyen et al., 2016). Within recent years, however, an increasing interest amongst CL scholars in examining textual patterns in social media data has given rise to the emergence of a distinct research enterprise in the form of CS.

Unlike CL, which has had “little to contribute to our understanding of language’s social dimension” (Krishnan & Eisenstein, 2015: 1616), CS research combines sociolinguistic theories and computational methods to examine the social factors that condition linguistic variation in text (Hovy & Johannsen, 2016). The development of CS can be largely attributed to an increasing interest amongst CL scholars in developing NLP models that are sensitive to the sociolinguistic variability in text in order to increase the reliability and productivity of those models (Jørgensen, Hovy, & Sjøgaard, 2016). In this vein, models have been built to predict the author’s age (Burger & Henderson, 2006; Rosenthal & McKeown, 2011; Schler, Koppel, Argamon, & Pennebaker, 2006), gender (Bamman, Eisenstein, & Schnoebelen, 2014; Rao, Yarowsky, Shreevats, & Gupta, 2010; Schler et al., 2006), and ethnicity (Eisenstein, Smith, & Xing, 2011).

Given the obvious intersections between the two areas of research, CS scholars have emphasized the potential of using NLP tools in sociolinguistics (Eisenstein, 2017; Hovy & Johannsen, 2016). Traditionally, sociolinguists have relied on fairly organic yet modest datasets composed of recorded conversations and interviews. Here, the advantages of “big data” in sociolinguistics are clear. Vast datasets composed of conversations from numerous individuals can be created far more easily in less time than “traditional data” (Eisenstein, 2017; Nguyen et al., 2016).

It is, therefore, perhaps unsurprising that social media datasets are becoming commonplace in studies of language variation and change (e.g. Wieling et al., 2016). In these studies and in the absence of speech, researchers examine the orthographic representation of some spoken language feature. Overwhelmingly, data are extracted from Twitter—a social media site where users post messages restricted to 280 characters¹ known as “tweets.” Users elect to “follow” others and receive an aggregated stream of updates in their “newsfeed.” Unlike most other social media sites, Twitter grants researchers access to its Application Programme Interface (API), permitting the extraction of tweets from millions of users. Helpfully, metadata such as the time and date of the tweet and the geographical location of the device from which it was sent can also be scraped, allowing for inferences to be made about the social characteristics of the user.

In a number of studies, researchers have used this metadata to examine the social conditioning of an orthographic feature in relation to its spoken language counterpart. A case in point is Eisenstein’s (2015) analysis of orthographic coda deletion in Twitter. Examining 114 million tweets from 2.77 million users, Eisenstein explores the orthographic representation of two spoken language features: *t/d* deletion and (ING), as in *<hol up>* for *<hold up>*, and *<walkin>* for *<walking>*. His analysis shows that, overwhelmingly, the orthographic representations of these features are constrained by similar linguistic and social factors to those identified in speech (cf. Thomas, 2007). Specifically, he observes a statistical

correlation between the variable realization of *t/d* and (ING) and the user's inferred ethnicity based on the location from which the tweet was sent, noting that there are "higher frequencies of non-standard spellings in counties that include more individuals who identify as African Americans, and more standard spellings in counties that include more individuals who identify as White" (Eisenstein, 2015: 181).

Similarly, Jones (2015) uses a dataset of geotagged tweets to identify AAVE dialect regions by querying Twitter's API for non-standard representations of the variety (e.g. *sholl* "sure"). By plotting the geolocation of the tweets, Jones identifies unique patterns of orthographic variation, each corresponding to four distinct AAVE (spoken) dialect regions that map directly onto patterns of movement during the Great Migrations.

Whilst these findings demonstrate the analytical potential of using social media data in variationist sociolinguistic research, they also reveal one substantial limitation of the CS approach. Specifically, the extent to which the socio-demographic factors inferred from metadata can be used to explain socially meaningful patterns of orthographic variation.

Typically, CS researchers have used metadata to explore population-level trends, examining the correlation between an orthographic variable and a given social factor. In these studies, social factors are generally operationalized at the macro-level, with users classified into broad demographic categories such as "male" and "female." For instance, in their analysis of user attributes in Twitter, Rao and colleagues (2010) identify a correlation between specific orthographic variables and the authors' inferred sex. They find that women were more likely to use emoticons, ellipses (...) and alphabetic character lengthening (*waaay*), and forms of excitement (!!) than men.

Although the macro-level approach typically used in CS has been useful in examining population trends of language use, by extracting over more micro-level patterns, CS analyses have tended to characterize social factors (e.g. sex) as latent qualities of the individual, with the contribution of those variables analysed independently. Consequently, CS risks encountering similar issues to those discussed in relation to "first-wave" sociolinguistic and CMC research (cf. Eckert, 2012; Georgakopoulou, 2006), whilst remaining largely resistant to third-wave perspectives on the social meaning of variation.

In an attempt to provide a more nuanced account of identity and orthographic variation in social media, Bamman and colleagues (2014) examine gender patterns on Twitter from a more performative perspective. They show that users whose linguistic behaviour differs from population statistics can be accounted for in terms of the homophily of social networks, in that users with significantly fewer same-gender social connections exhibit patterns of language use that diverge more strongly from the classifiers' model of gender.

Whilst Bamman and colleagues' (2014) analysis demonstrates the potential of combining qualitative and quantitative methods to examine the more performative aspects of language use online, the bulk of CS research continues to examine the relationship between variation and independent macro-level social factors (Nguyen et al., 2016). As a consequence, CS research risks what Jones refers to as a "serious confound" (2015: 411) in that features may be erroneously attributed to the dialect/social factor under study. In the context of his own analysis, Jones cites the White mainstream adoption of AAVE terms as a case in point, acknowledging that such instances can no longer be "exclusively attributed" to the variety (2015: 411).

Although similar processes are no doubt relevant to spoken language (e.g. Cutler, 1999), these issues are potentially heightened in digital contexts. In social media, where geographically disparate users forge connections with users beyond their own physical community, the potential for non-local linguistic styles to be (stylistically) appropriated is arguably increased. Thus, whilst macro-level analyses may be able to account for orthographic variation at the population level, it is unlikely the approach typical of CS will be able to account for patterns of orthographic variation that are constrained by micro-interactive factors, such as stylization. Given that a wealth of research has demonstrated

the importance of stylization in spoken communication (Coupland, 2001; Rampton, 1995), I argue that it is necessary to explore this phenomenon in social media.

3 | STYLIZATION

Contemporary sociolinguistics has tended to conceptualize intraspeaker variation beyond a model of “attention paid to speech” (cf. Labov, 1972), instead seeking to understand not only how speakers reactively adapt their speech styles, but also the ways in which speakers actively deploy linguistic features to communicate some given personae and achieve certain interactional ends. Consequently, a great deal of work has examined the notion of “stylization”—speech events in which the speaker produces “specially marked and often exaggerated representations of languages, dialects, and styles that lie outside their own habitual repertoire” (Rampton, 2009: 149; see also Coupland, 2007). For instance, in his analysis of English-language national radio broadcasts in Wales, Coupland (2001) shows how the presenter, Roy Noble, evokes images of Welshness by deploying phonological features that are considered “stereotypically Welsh.” Coupland suggests that by stylizing features of the dialect, Noble evokes Welsh cultural stances and practices that lay beyond his own habitual identity.

Importantly, what makes stylization possible is that linguistic variables do not exist in isolation, but instead accrue some indexical association with a specific style and/or identity. Agha (2003) has termed this process “enregisterment,” referring to the procedure in which linguistic forms become ideologically related with social identities. Agha refers to Received Pronunciation (RP) as a case in point. He shows that RP, once a local variety associated with the south-east of England, has since become enregistered as a prestige variety associated with “a specific scheme of cultural values” (2003: 231)—a change fuelled by prescriptivist metadiscourses that were prevalent in the eighteenth and nineteenth centuries.

As enregistered styles, linguistic features become associated with socially recognizable personae, or “characterological figures” (Agha, 2003: 243)—defined as essentialized personifications of the imagined typical user, linked with non-linguistic attributes, such as social class, conduct, and other mental and aesthetic qualities. Thus, in deploying a specific linguistic feature, the user can evoke the characterological figure(s) associated with that style and capitalize on its interactional potential. This is the case in the “crossing” observed by Rampton (1995). In that analysis, Rampton argues that adolescent speakers stylize features of British Asian English to perform a characterological figure of the essentialized “Other” to achieve specific interactional ends, such as to undermine White authority figures and increase social bonds amongst peers.

Whilst theories of style and stylization have been developed (primarily) with reference to spoken phenomena, Barrett (2017) examines stylization in posts made to the “Bear Mailing List” (BML) by members of the “bear” gay subculture—an identity that is often defined in terms of a larger stature, appreciation of body hair, and displays of hyper-masculinity. He shows that the use of non-standard spellings are central to the formation of a “bear” identity, with users stylizing orthographic features that index a “White Southern” dialect, juxtaposing these with French borrowings such as “savoir faire” and stereotypically “gay” lexical items such as “fabulous” and “extravaganza” (Barrett, 2017: 113). The use of these non-standard forms does not suggest that contributors to the BML are attempting to claim “working-class southerner” or “French speaker” identities, but rather, as Barrett argues, members of this subculture use these forms to exploit the indexical value of these varieties—masculinity and the aspirations of the bourgeoisie—without explicitly claiming membership to either, thereby distinguishing the unique “bear” style.

In this paper and in the sections following, I make a very similar argument. Here, I focus on a specific type of persona, well established in the gay community but having had little academic treatment—which I refer to as the “Sassy Queen” (cf. Podesva, 2007, on the closely related “Gay Diva” style, and Calder, 2019b, on the “fierce queen” style).

4 | METHOD

4.1 | Data collection and coding

The data considered in this paper are taken from a corpus of tweets posted between May 2015 and April 2016. Tweets were extracted from 10 different users’ public timelines through Twitter’s streaming API using the *twitteR* package in R (Gentry, 2016; R Core Team, 2016). The users were individually identified by randomly selecting public accounts of those who follow LGBT lifestyle and culture pages, such as gay magazine *Attitude* and well-known (UK) drag queens.

Unlike other CS analyses which analyse vast numbers of tweets from millions of users, the intentionally small sample of users was selected in order to make more nuanced inferences regarding the authors’ characteristics. Only users who present as White, aged between 18 and 25, based in the south of England, and who openly identify as gay men were selected. These demographics were determined based on the content of the users’ timeline/tweets and/or was explicitly stated in their profile information.

Whilst it is not possible to determine how these characteristics relate to users’ offline identities, it is important to note that all of these individuals present (at least) as gay, White, young male adults living in the south-east of England. Thus, based on the available demographic information, it is therefore likely that the user’s spoken habitual dialect is a variety of Southern British English (SBE). Whilst they may have had some mediated exposure to American English varieties (including AAVE), most likely through US TV shows, it is highly unlikely that they are native speakers of AAVE given the absence of this speech community in the UK.

This sample of individuals was selected for two reasons. First, White southern British users were selected to maximize the distance between the style analysed here (AAVE) and the user’s assumed habitual dialect (SBE) based on the linguistic differences between the two varieties (cf. Lindsey, 2019). Second, the tweets of gay men were examined because of the well-established use of stylized features of AAVE in spoken interactions amongst members of certain subcultures of the gay community, where such elements are often borrowed via drag culture (e.g. *RuPaul’s Drag Race*; Barrett, 2017; Calder, 2019b; Mann, 2011). Importantly, unlike other ethnic and local varieties of British English, AAVE features prominently in LGBT culture. For instance, many elements of contemporary “gay-slang” appear to have been appropriated from the variety (Street, 2018), and the style often features in the performances of drag queens in the UK.

As is typical in CS (e.g. Eisenstein, 2015; Wieling et al., 2016), and to avoid duplication, retweets and advertisements were removed from the dataset. In addition, the maximum number of tweets was limited to 3,000 per user to avoid an uneven skew across the dataset. The final corpus comprises 15,804 tweets. Whilst I acknowledge that both the number of users and the size of dataset are modest, particularly in comparison to the large-scale datasets obtained by other CS scholars (cf. Eisenstein, 2015), the intentionally restricted dataset allows for close qualitative analysis of orthographic variation and the related interactional stances that these features afford.

To examine stylization in the dataset, the corpus of 15,804 tweets was manually trawled for instances of orthographic variation. All tweets containing non-standard spellings were extracted and were individually examined for the presence of AAVE features by cross-referencing the feature to existing accounts of the spoken variety (cf. Green, 2002; Thomas, 2007).² Unlike in CS, where variable patterns are

analysed automatically using computational tools, the features in this paper were manually analysed. The purpose of this is two-fold: first, to increase the accuracy of the coding process and, second, to examine the specific interactional contexts in which the feature occurs (see also Georgakopoulou, 2006).

In the following analysis, I report absolute totals as it is not possible to conduct a traditional variationist analysis of these variables since, as I will argue, the orthographic representation of AAVE features occur in very specific interactional contexts. Thus, unlike other CS approaches which examine the linguistic constraints of a particular feature (e.g. Eisenstein, 2015), I argue that the envelope of variation here is instead constrained by the interactional context in which the variable occurs (cf. Moore & Podesva, 2009; Rampton, 1995).

4.2 | African American Vernacular English

For the purposes of this paper, I define African American Vernacular English (AAVE) as an ethnolinguistic repertoire (Benor, 2010) which is typically spoken by (some) working-class Black Americans (Green, 2002; Winford, 2015). Here, I follow other scholars in distinguishing between African American English (AAE)/African American Language (AAL), labels which describe the speech of African Americans more generally, and AAVE, which refers specifically to the vernacular variety discussed herein (see Mufwene, 2001; Thomas, 2007: 451). When defined in these terms, AAVE can therefore be considered a subsystem of AAE/AAL, characterized by a number of distinctive lexical, phonological, and morphosyntactic features, including completive *done* and the monophthongization of /aɪ/. Owing to its long history in the South, several features characteristic of AAVE are also shared with other southern varieties of American English.

Due to constraints of space, it is not possible to provide an exhaustive list of the AAVE features that were coded in the analysis. Rather, in Table 1, I introduce some of the most recognizable and, arguably, salient features of the variety (Green, 2002; Thomas, 2007; Thomas & Bailey, 2015). For comparison, these features are provided alongside their General American English (GAE) counterpart.

5 | ANALYSIS

5.1 | Distributional analysis

Of the 15,804 tweets extracted from the 10 users, 307 (1.9%) individual tweets contain AAVE features. Of these, 29/307 tweets (9.4%) contain more than one feature. This represents a total of 336 unique AAVE tokens which are considered in this analysis.

Table 2 shows the variability across all users; whilst all 10 users exhibit some use of AAVE features, rates vary considerably across the dataset. User 4 dominates, using at least one AAVE feature in 6.8% (n=133) of his 1,958 tweets. Whilst the overall rates of usage are relatively low, it is important to remember that these individuals are highly unlikely to be native AAVE speakers. Thus, unlike in

TABLE 1 Example features of AAVE (Green, 2002; Thomas, 2007; Thomas & Bailey, 2015)

| AAVE feature | GAE realization | AAVE realization |
|-------------------------------------|-------------------------------|---------------------------------|
| <i>Monophthongization of /aɪ/</i> | bye /baɪ/, time /taɪm/ | /ba:/, /ta:m/ |
| <i>Copula absence of is and are</i> | she is going who are you? | she going who you? |
| <i>Substitution of /ð/ and /θ/</i> | that /ðæt/, bath /baθ/ | /dæt/, /baf/ |
| <i>Completive done</i> | she used up all the good ones | she done used all the good ones |

TABLE 2 Raw frequencies of the lexical, phonological, and grammatical AAVE features in the dataset

| User ID | Lexical | Phonological | Grammatical |
|----------------|---------|--------------|-------------|
| User 1 | 3 | 26 | 3 |
| User 2 | 2 | 10 | 2 |
| User 3 | 0 | 0 | 6 |
| User 4 | 90 | 54 | 11 |
| User 5 | 2 | 2 | 1 |
| User 6 | 12 | 11 | 0 |
| User 7 | 19 | 14 | 7 |
| User 8 | 7 | 14 | 6 |
| User 9 | 2 | 20 | 0 |
| User 10 | 8 | 1 | 3 |
| TOTALS: | 145 | 152 | 39 |

Black Twitter (cf. Jones, 2015), the users in this study do not appear to be faithfully representing their own linguistic system, but rather seem to be selectively, and intentionally, deploying a subset of AAVE features in specific contexts. The somewhat low rate of variation is therefore expected.

Comparing the rates of usage across linguistic domains, Table 2 shows that there is some preference towards phonological and lexical features of AAVE, as opposed to grammatical (i.e. morphosyntactic) features. This pattern holds across all users except User 3, who rarely uses AAVE in his tweets and whose usage is limited to copula deletion, invariant *be*, and demonstrative *them*—three features that are often considered stereotypical of AAVE (Green, 2002). The fact that we see a preference for lexical and phonological features is, perhaps, unsurprising given that other studies of stylized dialect use have shown that speakers exhibit a command of phonological and lexical features but generally do not acquire the morphosyntactic system (e.g. Cutler, 1999). Similarly, as noted, only 9.4% ($n = 29/307$) of tweets with AAVE features contain more than one feature of the variety. This distribution therefore seems to mark this language use as inauthentic, such that the users in this study are not attempting to depict themselves as AAVE speakers or Black, but rather appear to be appropriating specific elements of the variety in specific contexts.

Tables 3–6 show the raw frequencies of AAVE features across different linguistic domains. Whilst it would be possible to subject the data to a variationist analysis, I argue here that this would be uninformative. My claim is not that the AAVE features occur with any dialectological systematicity as we would expect native speakers to exhibit. Rather, the methodological and analytical decision to provide absolute frequencies coincides with my argument that I offer in later sections: that the appearance of AAVE features is conditioned by the specific interactional context. I examine each level in turn, drawing attention to the diversity of variables across the dataset.

5.2 | Lexis

Table 3 shows the AAVE lexical features in the dataset. As discussed previously, it is necessary to acknowledge that some lexical items are shared with other (Southern US) varieties (e.g. *y'all*; Green, 2002), whilst others have been appropriated into other varieties of English (Jones, 2015). With these caveats in mind, it is not possible to definitively argue that the particular word has been appropriated directly from the AAVE community, nor can I establish whether users were aware of the association of these forms with AAVE. Rather, my decision to include the lexical features discussed here is

TABLE 3 Lexical features in the dataset

| Feature | Example | Count |
|----------------|--|-------|
| <i>Y'all</i> | @User @User @User regardless of her colour, she's using her celebrity for good to raise awareness. Y'all haters corny.Done. | 47 |
| <i>Yaas</i> | YAAAAAAS @User! Greatest athlete in the world? Of course that's how @User rolls. #LEMONADE | 34 |
| <i>Basic</i> | THERES MORE CHANCE OF ME TURNING BLACK THAN TURNING ON YOUR BITCH ASS #INSTAGRAM NOTIFICATIONS!!! STOP BEGGING, YOU BASIC ASS BITCHES!!!! | 32 |
| <i>Thirsty</i> | @User @User y'all need to have a drink. Lawd knows y'all thirsty ! | 9 |
| <i>Hunty</i> | @User hunty it's not clever if it's aided. Try parking a four-seater convertible with no sensor and no rear visibility. Xo | 7 |
| <i>Ratchet</i> | @User @User @User oh I will be. And I'll be sure to drag this ratchet bitches ass name through the mud like y'all have. Xo | 6 |
| <i>Bae</i> | When I wake up and there no text from bae ! And then suddenly remembering the fact there i have no bae | 5 |
| <i>Squad</i> | #squad goals | 3 |
| <i>Momma</i> | @User oh totally. Momma didn't raise no pussy bitch. An ex is an ex for a reason. I wish them all the best | 2 |

TABLE 4 Phonological features in the dataset

| Feature | Example | Count |
|--|--|-------|
| <i>(ING)</i> | Sat Sippin' ma Starbucks in Soho Square with @User #london #soho #sohosquare #relax | 45 |
| <i>Substitution of dental fricatives</i> | Post swim, gym and steam.... Feeling pumped for another night in the big city working dat podium! | 23 |
| <i>Monophthongization of /aɪ/</i> | Story of ma' fuckin life | 8 |
| <i>Centralization of /ɛ:/ → /ə/</i> | My huur done gone CRAY! | 9 |
| <i>R-lessness</i> | @User@User y'all need to have a drink. Lawd knows y'all thirsty! | 4 |
| <i>Consonant cluster reduction</i> | @User don't stan near me please xox | 4 |

TABLE 5 Eye dialect features in the dataset

| Feature | Example | Count |
|------------------|---|-------|
| <i>Gurl</i> | @User oooh gurl not even I get my arse out on a sun bed... And I'm not of fair skin. | 33 |
| <i>Werk/werq</i> | Y'all have a cracking week mofos !!!!! #mondaymadness #werk #happymonday | 11 |
| Miscellaneous | @User boy bye... I ain't thinking bout chu . | 12 |
| <i>Fuq/fucq</i> | DaFUQ @User | 3 |

motivated by a methodological concern to enable a comparison with other lexical-based CS accounts of AAVE (e.g. Eisenstein, 2017; Grieve et al., 2018: 298–299)³ and to demonstrate that these items often co-occur with other features of the variety which, together, suggest an awareness of the invoked persona as “typically AAVE” (e.g. see *thirsty*).

TABLE 6 Morphosyntactic features in the dataset

| Feature | Example | Count |
|--|--|-------|
| <i>Copula absence (are)</i> | @User @User oh you nasty | 16 |
| <i>Invariant be</i> | @User Filters be a gay's best friend. | 6 |
| <i>Demonstrative them</i> | My bitches working them boots #Gucci #glamsquad #dubai #mydubai #friends #fashionista | 6 |
| <i>Absence of third person singular -s</i> | You a strong, black, independent woman, who don't need no man! @User @User #Beyoncé #Lemonade #BeyHive #Bey | 5 |
| <i>Copula absence (is)</i> | I miss @User where she at? | 4 |
| <i>Completive done</i> | Done bought me a candyfloss machine! | 2 |

Table 3 shows that, whilst AAVE lexical items such as *yaas*⁴ and *basic* commonly occur in the dataset, they generally do so with other standard (and relatively formal) lexical items, such as *suddenly* and *regardless*. In this sense, the use of these lexical features appears to be highly tokenistic, such that we do not see that all words which have an AAVE counterpart are systematically substituted for that lexical feature.

5.3 | Phonology

Turning to the orthographic representation of AAVE phonology, we see that, perhaps unsurprisingly, the most frequently depicted variants are those which occur in multiple varieties. For instance, the orthographic representation of (ING) as in <in> for <ing>, which resembles the spoken realization of the apical nasal [ɪn] in place of the standard velar nasal [ɪŋ], far exceeds other features. However, it would not be possible to argue that every instance of (ING) can be attributed to AAVE since this variable occurs in numerous dialects, and indeed in casual speech (Thomas, 2007). Rather, as with the lexical features, it is the combination of (ING) with other features in particular contexts which suggests that this variable is used to invoke a reading of the message as “stereotypically AAVE.”

Whilst orthographic representations of (ING), the substitution of the interdental fricatives, and even to some extent the monophthongization of /aɪ/ to [a:], such as in the spelling <ma>, may be unsurprising given their tokenistic status and the possibility that they are more likely to be “performed” in stylization (e.g. Cutler, 1999), features that are less salient stereotypes of the variety are also represented. For instance, the relatively recent development of the centralization of the NEAR VOWEL /ɪə/ and the SQUARE VOWEL /ɛ/, to a near-merger with the NURSE VOWEL /ɜ:/ (Thomas, 2007: 466), is represented as <hurr> for <hair> and <thurr> for there.

Nevertheless, as with the features previously discussed, there appears to be a lack of adherence to the phonological constraints that AAVE speakers generally exhibit. For instance, the orthographic representation of the monophthongization of /aɪ/ is largely restricted to the lexeme “my” (n=7/8) and is variably realized as <ma> and <my> by the same users (cf. AAVE, Green, 2002; Thomas, 2007). Similarly, whilst the substitution of the dental fricative /ð/ with [d] can occur in any word position in AAVE (Green, 2002), in the dataset this variant is limited to word initial position in the function words *this* <dis>, *that* <dat>, *the* <da>, and the pronoun *them* <dem>. This observation coincides with Woolard's suggestion that word initial stopping in words with definite reference (e.g. this, that, the, them) has become a salient stereotype which makes this feature “ripe for social semiotic and stylistic work” (2008: 443). These observations suggest that the users are not attempting to accurately represent the systematicity of AAVE phonology, but rather they have stylistically appropriated a subset of features from the variety.

5.4 | Eye dialect

A subsection of tweets categorized as phonological features of AAVE seem to attempt to mimic some phonological representation, yet it is not entirely clear what that particular feature is (see Table 5). I distinguish here between direct representations of AAVE phonology and “eye dialect” forms since I argue that the latter category rests entirely on the reading of the form as distinctly “non-standard” (Sebba, 2007) as opposed to those forms discussed previously which attempt to represent some AAVE phonetic and phonological feature directly.

In Table 5, we see that the eye dialect forms generally concern two features: the representation of schwa /ə/ as <u> as in <gurl> for *girl* and the substitution of <k> for <q> as in <fuq> for *fuck*. The miscellaneous category includes features which simply represent general phonological processes such as the coronal palatization of /j/ to /tʃ/ in <chu> for *you*. As with all features discussed previously, there is little consistency in how these words are represented. For instance, *work*, which is variably spelled as <werk>, <work> and <werq>. Since eye dialect respellings are often used to destabilize notions of standardness (Sebba, 2007), the lack of consistency in the representation of these forms is expected.

5.5 | Morphosyntax

Table 6 shows the distribution of AAVE morphosyntactic features in the dataset. Note that, although some of these features occur in numerous dialects (e.g. copula absence), less frequent features such as completive *done* are also found in the data. Whilst morphosyntactic features are generally less frequent than phonological and lexical variables, this cline is to be expected given that there is evidence to suggest that grammatical and aspect systems are more difficult to stylistically acquire than other features (Bucholtz & Lopez, 2011; Cutler, 1999).

Although the tweets reflect a diverse range of AAVE morphosyntactic features, as with the other features examined thus far, there appears to be little systematicity in the ways in which these non-habitual users represent these forms. For instance, habitual *be*, which is perhaps one of the most salient features of the dialect, fulfils a very specific function in AAVE, where it references repeatedly occurring events. The use of invariant *be* in this data does not always adhere to this function. For instance, in the example given in Table 6, the meaning of *be* is not transparent. The message is a response to a compliment on the user's photo, admitting that the specific photo was greatly improved by the addition of photographic filters. This reading does not evoke a habitual reading. If the intended meaning is “filters are *always* a gay's best friend,” however, a habitual reading would be possible. Nevertheless, the use of invariant *be* in contexts that are not explicitly habitual appears to be stylistic, marking the text as stereotypically “Black.” As Green notes in her analysis of the feature in media, habitual *be* “can conjure up images of [B]lack speech even if it is not used according to rules of AAE” (Green, 2002: 207).

6 | AAVE IN INTERACTION

In the analysis above, I have shown that orthographic representations of AAVE features are frequent across all linguistic levels. The lack of dialectal systematicity, however, and the obvious discrepancies between the user's (inferred) habitual dialect seem to suggest that users are not faithfully replicating their habitual spoken repertoire (c.f. Black Twitter: Jones, 2015). I would therefore argue that the appearance of this non-local style cannot be explained by macro-level correlations (cf. Eisenstein, 2015), but rather must be understood as “stylization”—or, more specifically, “crossing” (Rampton, 1995)—with the appearance of AAVE features influenced by the specific communicative context

in which they occur and the affordances of using that style in achieving certain interactional ends (Coupland, 2007).

Given that Twitter is inherently a *social* media website, one way to examine the interactional affordances of AAVE is to analyse the types of posts that these features occur in. Although Twitter mainly facilitates the posting of one-to-many tweets, authors can “direct” a message at another user by tagging their username with <@> in a message.

Of the 15,804 tweets, just over half of the messages (52.8%, n=8336) are directed. To situate this in the current analysis, 187/307 (60.9%) tweets containing AAVE features are directed—that is, they specify another user's handle in the tweet, creating a “conversation.” This finding alone appears relatively uninformative, suggesting that AAVE features are only slightly more common in directed messages. However, when the intended recipient of the message is considered, we see a strong correlation with the user's sexuality and the use of AAVE features: Of the 187 tweets that are both directed and contain AAVE features, 90.9% (n=170) of those tweets are to other gay male users.⁵

This finding seems to suggest that AAVE features are more likely to be used in interactions with other gay users. Such an observation is largely predicted by Coupland's assertion that stylization relies on an “acculturated audience able to read and predisposed to judge the semiotic value of a projected persona or genre” (2007: 154). Thus, what this correlation suggests is that the use of the non-local style (AAVE) appears to circulate amongst these individuals as an “ingroup code,” fostering relationships between members of the “acculturated audience” who identify with the same gay subculture (cf. Barrett, 2017).

Although this style is arguably most useful in directed messages, this interpretation can also explain the use of this style in non-directed, “one-to-many” messages where the tweet does not form a conversational thread. To interrogate the interactional affordances of this style further, I now turn to an analysis of the individual tweets to examine the contexts in which it is evoked.

1. Work dat pole gurl! @VENUE <link to photo> [User 7]

The tweet in (1) is an example of a non-directed message which includes a hyperlink of a photo depicting a (male) friend pole-dancing in a nightclub, which is presumably the same venue that is tagged in the tweet—a London-based LGBT venue. The tweet refers to the sender's somewhat exaggerated and “lively” appreciation of his friend's dancing, using the AAVE forms <dat> and <gurl>.

Even in the absence of “interaction,” the stylistic use of AAVE features in this tweet appear to suggest that they are being used as part of an ingroup style, as signalled by the references which point to the intended audience. The use of the term <gurl> “girl” is widely used among gay men as a form of address (Graf & Lippa, 1995), the photo of the male friend links both subjects to the use of this style and, by tagging an LGBT nightclub, the user explicitly situates the tweet within a “gay” context. Thus, whilst this tweet is not directed at a specific individual, the author designs his message using semiotic resources (such as the orthographic variants <dat> and <gurl>) for the “imagined audience” (Marwick & boyd, 2010), i.e. members of the same gay subculture, who recognize and value the use of that style. In this tweet, the “value” of this stylization is in its ability to project a persona that is both lively and provocative.

Importantly, however, the successful interpretation of the stylization of AAVE is reliant on a mutual awareness amongst the acculturated audience that the style used and the persona invoked is inauthentic. As noted previously, the users in this study are not attempting to present themselves as AAVE *speakers* (cf. Black Twitter: Jones, 2015), but rather selectively stylize elements of the variety to achieve certain interactional ends. Most frequently, this strategically inauthentic style (Coupland, 2007) is seen to occur at junctures of interactional uncertainty or disagreement, where stylization reduces the speakers' commitment to the pragmatic force of the message (Rampton, 2009).

2. @USER don't come for my queen, bitch. ELSE WE GONNA FALL OUT! And that will have negative consequences for you [User 4]

The tweet in (2) is an example of a directed message to another gay male user, in which User 4 responds to derogatory comments made about music artist Beyoncé, who he refers to as “my queen.” Responding to these comments, User 4 suggests that disrespecting his “queen” would lead to the pair “falling out,” using the AAVE feature of copula deletion (“WE GONNA”), upper-case letters, and expressive punctuation (!), to take an inherently oppositional stance.

However, based on other interactions between the two users, it seems unlikely that this tweet is intended to be perceived as truly aggressive or as a threat. Rather, this exchange appears as part of a larger light-hearted discussion between two users who are seemingly friends. Here, I argue that the stylization plays a central role in establishing the intended meaning of the tweet that this is not a “true” argument by using features of AAVE to evoke a non-habitual identity that is not to be taken on face value—a type of “strategic inauthenticity” (Coupland, 2007). Specifically, I suggest that by performing an identity that is mutually recognized as “Other,” the user deploys a contentious or provocative stance whilst simultaneously avoiding the possibility that his message may be misinterpreted as a true argument or insult.

In (3), a similar type of interaction ensues. In this tweet, the user responds to an interaction where other users are discussing the attractiveness of another gay man. His tweet, like (2), is inherently oppositional, suggesting that his friends are ‘thirsty’ (i.e. desperate). Here, the combination of AAVE lexical features, <y'all>, <Lawd> and <thirsty>, are used as a means to construct a stance that is both provocative and outspoken, whilst avoiding the offensive interpretation of the tweet – that his friends are desperate.

3. @USER y'all need to have a drink. Lawd knows y'all thirsty!

Here, as in (1), it appears that by stylising elements of AAVE, the user evokes a non-local persona that both the sender and recipient mutually interpret as inauthentic (Rampton 1995), allowing him to mitigate the possibility of confrontation by distancing himself from the ‘pragmatic implications of what is said’ (Coupland 2001:366). Thus, the user is able to simultaneously disagree with his interlocutors, whilst preserving normative social relations amongst friends. However, whilst it is possible to discern the interactional affordances of this style from these examples, it is unclear as to why it would be AAVE, specifically, that is stylised by these users. In what follows, I argue that features of this variety have become appropriated to deploy ‘provocative’, ‘confrontational’ and ‘outspoken’ stances because these qualities have become enregistered as components of a recognisable ‘characterological figure’ (Agha 2003) that has been made possible by an essentialised imagining of Black women. It is this discussion that I turn to next.

7 | AAVE AND THE “SASSY QUEEN” PERSONA

As Eckert (2003, 2012) has argued in relation to the “first-wave” of sociolinguistic research, macro-analyses tend to abstract over the social meaning of variation, instead identifying broad correlations between features and social groups. By extension, CS faces similar challenges since, in these analyses, the social meaning of orthographic variation is attributed to broad socio-demographic categories, e.g. “male” vs. “female” (e.g. Rao et al., 2010). In the current analysis, macro-level correlations usually associated with AAVE speakers (Black, working-class, African American) are clearly unhelpful given that the demographics of the users are incongruent with these social factors. Thus, to explain the use of this style, I suggest that we need to look beyond Twitter, to the wider digital context, to observe the “regular patterns of metapragmatic typification” (Agha, 2004: 29) associated with AAVE.

Here, I argue that one way of inferring these associations is by examining prevalent depictions of AAVE and Blackness that are circulated as “internet memes”—those pictorial representations of a cultural artefact (e.g. videos, pictures, and bitmap images such as GIFs) that attain viral status. Widely shared across social media platforms, memes have become “highly valued pillars of a so-called *participatory culture*” (Shifman, 2014: 4, emphasis in original). Since their success is reliant on a collective appreciation of the sentiment or cultural reference of the meme, I propose that examining memes may be helpful in uncovering the “metapragmatic classifications” (Agha, 2004: 23) that users attribute to social identities and their associated linguistic styles.

In the context of this analysis, one of the most celebrated memes that is relevant to the current discussion comes from a 2012 television interview with African American and AAVE speaker Kimberly Wilkins. Appearing live, Wilkins—introduced as “Sweet Brown”—gives a dramatic interview in which she details her escape from a fire in an apartment complex in her hometown of Oklahoma City. The video became an instant viral sensation, with many parodying Wilkins’ claims that she “got bronchitis” from smoke inhalation and her remark that “ain’t nobody got time for that”—phrases which have since become commodified in memes (see Figure 1) and in other user-generated content, including the song “I got bronchitis,” which appeared on the iTunes music store (Know your Meme, 2012a). To date, the “autotune remix” version has been viewed by over 65 million YouTube users.

Along with her viral celebrity status, however, Wilkins’ testimony of a life-threatening situation has instead become characterized as a comical depiction of the ethnic “Other.” Consequently, “Sweet Brown” becomes the personification of an essentialized imagining of Blackness and working-classness, playing into mainstream discourses of the “typical” African American (Winfrey-Harris, 2012). Central to this persona and rhetorical style is her vernacular—AAVE.

Similar themes are evident in Figure 2, a meme widely shared under the guise of the “strong independent Black woman who don’t need no man.” The image depicts a Black woman who is, assumedly, African American—note the absence of third person singular *-s* in the use of *don’t* for *doesn’t*, a well-documented feature of the variety (Green, 2002). So popular is this meme that Sierra (forthcoming) observes references to the “strong independent Black woman” character in conversations amongst American youth. It is also referenced directly in the tweets of several individuals in the study (see Table 6, for example).

As with “Sweet Brown,” this meme does not depict an innocuous Black identity. Rather, it relies on a very specific imagining—the stereotype of the “angry Black woman”—a well-established racial trope which renders Black women as aggressive, obnoxious, and lacking vulnerability and empathy (Harris-Perry, 2011). With her hand on her hip and lips pouted, the subject is characterized as confrontational, fierce, and determined, stances which reify the subject as “sassy”—a term that describes the



FIGURE 1 The “ain’t nobody got time for that” meme (Know Your Meme, 2012a) [Colour figure can be viewed at wileyonlinelibrary.com]



FIGURE 2 The “strong Black woman who don’t need no man” meme (Know Your Meme, 2012b) [Colour figure can be viewed at wileyonlinelibrary.com]

adoption of “impudent, saucy, cheeky; outspoken, provocative; conceited, pretentious; self-assured, spirited, bold” personal qualities (OED, 2018).

Together these memes become enregistered as digital stereotypes—or “cybertypes” (Nakamura, 2002)—of Black women that promote the “sassy Black woman” trope that is frequently depicted in TV and film. As such, qualities associated with “sassiness”—being provocative, self-assured, impudent—become enregistered as facets of a characterological figure (Agha, 2003) of Blackness. As part of this process, it is AAVE—as the vernacular that is associated with these subjects—that becomes the linguistic “vehicle” through which these qualities can be enacted.

Here, I argue that this explanation can go some way in explaining the variation in the users’ tweets examined in earlier sections. Specifically, I suggest that, by stylistically and selectively using elements of AAVE, the users reap the indexical qualities of this style and its association with the stereotype of the “sassy Black woman,” to present themselves as “fierce” or “sassy.” For the users in the study, this stylization is particularly prevalent because the characteristics of “fierceness” and “sassiness” have become appreciated qualities within the gay community (e.g. Barrett, 2017; Calder, 2019b). Indeed, an appreciation of being “sassy” is referenced in tweets by those in the study. For instance, consider (4) and (5):

4. @User There’s a really awful gay guy trying to be sassy. It’s very hashtag awks. [User 6]

5. @User there’s always time to be sassy [User 4]

In (4), the user makes clear that being sassy is a valued quality that someone “tries” to achieve. Achievements are generally thought of as positive lifestyle choices. Failure to achieve the standards of this quality results in the situation being read as “awkward,” such that this situation is one to avoid. Whereas in (5), the user writes that sassiness is an enduring quality—something that persists beyond the limits of time, that should be constantly employed.

Thus, for the gay men in this study, it appears that AAVE has become a form of cultural capital (Bourdieu, 1991) through which they are able to evoke a “sassy” or “fierce” persona. Whilst sassiness could potentially be enacted in other ways, I suggest that it is the stylization of AAVE specifically that fulfils this purpose because of a characterological reification of Black women as “sassy,” which emerges in the memes discussed here. Importantly, however, it must be noted that the users in the study do not explicitly reference “Blackness” in their discussions of “sassiness.” This observation

lends support to my interpretation that the users are not attempting to pass as “Black women” or even “AAVE speakers” (cf. Eisenstein, 2015; Jones, 2015; Tatman, 2015). But, rather, they appear to exploit a particular indexical association of AAVE (“sassiness”) as a strategically inauthentic style (Coupland, 2007) to deploy a recognizable and appreciated persona that circulates in the gay community—what I refer to here as the “Sassy Queen.”⁶

Returning to the interactional analyses presented in (1–3), it is now possible to argue that the use of AAVE features and the related “outspoken,” “provocative,” and “contentious” stances are components of this broader persona. Specifically, what I would suggest is that, in stylizing elements of AAVE in certain contexts, the user evokes a recognizable gay persona—the “Sassy Queen”—that is reliant on a particular essentialized imagining of Black women as “fierce” or “sassy”—qualities that have become appreciated personal characteristics in mainstream gay culture (Barrett, 2017: 48–53; Calder 2019a, 2019b; see also *RuPaul's Drag Race*). In interaction, the performance of the “Sassy Queen” persona is particularly useful for these men because, as a strategically inauthentic style, it permits the user to mitigate the possibility that the “contentious,” “provocative,” or “confrontational” stances evoked through the use of this persona will be taken at face value by other members of the ingroup.

8 | CONCLUSION

Macro-demographic approaches to studying language variation in social media using NLP tools has initiated a fruitful and exciting area of sociolinguistic research. However, by examining variation in relation to macro-level social factors (e.g. Eisenstein, 2015; Jones, 2015; Tatman, 2015), CS risks repeating the shortcomings of first-wave CMC and sociolinguistic research (cf. Eckert, 2012; Georgakopoulou, 2006).

In this paper, I have argued that users of social media utilize the indexical potential of orthographic variation to deploy styles and personae, rather than simply reflecting habitual macro-level identities (cf. Eisenstein, 2015). By exploring variation in a corpus of 15,804 tweets extracted from 10 UK-based gay men, I have argued that the prevalence of AAVE features across all linguistic levels (phonological, morphosyntactic, and lexical) signals that users appropriate the indexical capital of Blackness (which AAVE is directly indexical of). By analysing memes, I have argued that there exists an essentialized imagining of the “Black Other” which is contingent on the property of “sassiness.” The appreciation of this quality can explain the orthographic variation in the data in that the use of AAVE is intended to evoke a characterological figure (Agha, 2003) that is imbued with the aesthetic and physical qualities of the “sassy Black woman”—an identity which is dependent on the “angry Black woman” trope. In using features of AAVE, these users draw on an appreciated and celebrated identity in the gay community—the “Sassy Queen.” Viewing the variation as dialect stylization explains why the features are not employed in a systematic native-like fashion (cf. Cutler, 1999): The persona is only evoked in certain interactional contexts where the “sassy” identity is appreciated and valued, where it is used to reduce the threat of disagreement and increase social cohesion between peers.

Given that the use of non-standard spellings is an overtly conscious activity requiring the flouting of heavily standardized rules (Sebba, 2007), it is important to analyse social media writing as “performance” (Deumert & Vold Lexander, 2013), wherein users appropriate semiotic resources to deploy and construct identities, rather than reflect them. Thus, like Bamman and colleagues (2014), I suggest that micro approaches to studying language variation in social media should complement large-scale macro-methodologies in order to fully explore how sociolinguistic identities manifest in digital settings (Androutsopoulos, 2016).

ACKNOWLEDGMENTS

I would like to thank a number of people who have provided insightful feedback during the development of this paper, in particular, Erez Levon, and the audiences at 'ICLaVE 9' and at Departmental Seminars at Queen Mary. I am also extremely grateful to the reviewers and editors of the *Journal of Sociolinguistics* who provided detailed and incisive comments on earlier versions of this article. I alone am responsible for any remaining errors or shortcomings.

ORCID

Christian Ilbury  <https://orcid.org/0000-0001-9289-271X>

ENDNOTES

¹This was a change implemented by Twitter in 2017. The data in this analysis, and all studies discussed, comprise tweets that are restricted to the original word limit of 140 characters.

²The data were coded by the author of this paper. A second blind-coding process was completed thereafter. Although using multiple annotators as in CL would have been preferable, this is not standard protocol in variationist studies and has not been practised in most CS studies (e.g. Jones, 2015; Tatman, 2015).

³Whilst I cannot verify whether the users are “aware” that these lexemes are AAVE given that several words have been appropriated in “mainstream” English, the majority of the interactional contexts in which these lexemes occur—where they refer to notions of Blackness (see the example of *basic*)—suggests that this indexical association still stands.

⁴The origins of <yaas> are difficult to ascertain and a more conservative analysis could exclude these tokens. However, I follow Grieve and colleagues (2018: 299) in characterizing this feature as AAVE. I code <yaas> as “lexical” as opposed to “phonological,” as <yaas> does not appear to be a direct synonym for “yes.”

⁵The sex/sexuality of the addressee was determined using the same methods as the users in the study basing this decision on timeline/profile content. Sex was unambiguous in all instances, but the user's sexuality was not always clear. In these cases, the recipient was coded as heterosexual.

⁶One potential interpretation of the users' performance of AAVE is that it represents a type of “minstrelsy” (cf. Bucholtz & Lopez, 2011). Whilst I do not explore this possibility owing to constraints of space, it should be acknowledged that the analysis presented here should not be perceived as an attempt to condone or justify “digital Blackface.”

REFERENCES

- Agha, A. (2003). The social life of cultural value. *Language and Communication*, 23, 231–273.
- Agha, A. (2004). Registers of language. In A. Duranti (Ed.), *A companion to linguistic anthropology* (pp. 23–45). Oxford, UK: Blackwell.
- Androutsopoulos, J. (2015). Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism*, 19, 185–205.
- Androutsopoulos, J. (2016). Theorizing media, mediation and mediatization. In N. Coupland (Ed.), *Sociolinguistics theoretical debates* (pp. 282–302). Cambridge, UK: Cambridge University Press.
- Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18, 135–160.
- Baron, N. (1998). Letters by phone or speech by other means: The linguistics of email. *Language and Communication*, 18, 133–170.
- Barrett, R. (2017). *From drag queens to leathermen: Language, gender, and gay male subcultures*. Oxford, UK: Oxford University Press.
- Benor, S. B. (2010). Ethnolinguistic repertoire: Shifting the analytic focus in language and ethnicity. *Journal of Sociolinguistics*, 14, 159–183.

- Bourdieu, P. (1991). *Language and symbolic power*. Trans. John B. Thompson. Cambridge, MA: Harvard University Press.
- Bucholtz, M., & Lopez, Q. (2011). Performing blackness, forming whiteness: Linguistic minstrelsy in Hollywood films. *Journal of Sociolinguistics*, 15, 680–706.
- Burger, J. D., & Henderson, J. C. (2006). An exploration of observable features related to blogger age. *AAAI Spring Symposium: Computational Approaches to Analyzing Weblog* (6: 15–20). Palo Alto, CA: Association for the Advancement of Artificial Intelligence.
- Calder, J. (2019a). The fierceness of fronted /s/: Linguistic rhematization through visual transformation. *Language in Society*, 48, 31–64.
- Calder, J. (2019b). From *Sissy* to *Sickening*: The indexical landscape of /s/ in SoMa, San Francisco. *Journal of Linguistic Anthropology*. <https://doi.org/10.1111/jola.12218>
- Coupland, N. (2001). Dialect stylization in radio talk. *Language in Society*, 30, 345–375.
- Coupland, N. (2007). *Style: Language variation and identity*. Cambridge, UK: Cambridge University Press.
- Crystal, D. (2004). *The language revolution*. Cambridge, UK: Polity.
- Cutler, C. (1999). Yorkville crossing: White teens, Hip-Hop and African American English. *Journal of Sociolinguistics*, 3, 428–442.
- Deumert, A., & Vold Alexander, K. (2013). Texting Africa: Writing as performance. *Journal of Sociolinguistics*, 17, 522–546.
- Eckert, P. (2003). The meaning of style. *Proceedings of the Eleventh Annual Symposium about Language and Society*, 47, 41–53.
- Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, 41, 87–100.
- Eisenstein, J. (2015). Systematic patterning in phonologically-motivated orthographic variation. *Journal of Sociolinguistics*, 19, 161–188.
- Eisenstein, J. (2017). Identifying regional dialects in online social media. In C. Boberg, J. Nerbonne, & D. Watt (Eds.), *Handbook of dialectology* (pp. 368–383). Chichester, UK: Wiley-Blackwell Press.
- Eisenstein, J., Smith, N. A., & Xing, E. P. (2011). Discovering sociolinguistic associations with structured sparsity. *Proceedings of the Association for Computational Linguistics* (pp. 1365–1374). Portland, OR: Association for Computational Linguistics.
- Gentry, J. (2016). TwitterR: R based Twitter client. Available at: <https://cran.r-project.org>. R package version 1.1.9.
- Georgakopoulou, A. (2006). Postscript: Computer-mediated communication in sociolinguistics. *Journal of Sociolinguistics*, 10, 548–557.
- Georgakopoulou, A., & Spilioti, T. (2016). *The Routledge Handbook of Language and Digital Communication*. Oxford: Routledge.
- Graf, R., & Lippa, B. (1995). The Queen's English. In W. Leap (Ed.), *Beyond the lavender lexicon: Authenticity, imagination, and appropriation in lesbian and gay languages* (pp. 227–235). Amsterdam, Netherlands: Gordon and Breach.
- Green, L. J. (2002). *African American English: A linguistic introduction*. Cambridge, UK: Cambridge University Press.
- Grieve, J., Nini, A., & Guo, D. (2018). Mapping lexical innovation on American social media. *Journal of English Linguistics*, 46, 293–319.
- Harris-Perry, M. V. (2011). *Sister citizen: Shame, stereotypes, and Black women in America*. London, UK: Yale University Press.
- Herring, S. (2000). Gender differences in CMC: Findings and implications. *CPSR Newsletter*, 18(1).
- Hovy, D., & Johansen, A. (2016). Exploring language variation across Europe: A web-based tool for computational sociolinguistics. *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (pp. 2986–2989). Paris, France: The European Language Resources Association.
- Jones, T. (2015). Toward a description of African American Vernacular English dialect regions using “Black Twitter”. *American Speech*, 90, 403–440.
- Jørgensen, A., Hovy, D., & Sjøgaard, S. (2016). Learning a POS tagger for AAVE-like language. *Proceedings of NAACL-HLT* (pp. 1115–1120). San Diego, CA: The Association for Computational Linguistics.
- Know Your Meme. (2012a). Sweet Brown/Ain't nobody got time for that. Last accessed April 2019 at <http://knowyourmeme.com/memes/sweet-brown-aint-nobody-got-time-for-that>.

- Know Your Meme. (2012b). Strong Black woman who don't need no man. Last accessed April 2019 at <http://knowyourmeme.com/memes/strong-black-woman-who-dont-need-no-man>.
- Krishnan, V., & Eisenstein, J. (2015). "You're Mr. Lebowksi, I'm the dude": Inducing address term formality in signed social networks. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 1616–1626). Denver, CO: Human Language Technologies.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.
- Lindsey, G. (2019). *English after RP: Standard British pronunciation today*. Basingstoke, UK: Palgrave Macmillan.
- Mann, S. L. (2011). Drag queens' use of language and the performance of blurred gendered and racial identities. *Journal of Homosexuality*, 58, 793–811.
- Marwick, A. E., & boyd, d. (2010). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media and Society*, 13, 114–133.
- Moore, E., & Podesva, R. (2009). Style, indexicality, and the social meaning of tag questions. *Language in Society*, 38, 447–485.
- Mufwene, S. S. (2001). What is African American English? In S. Lanehart (Ed.), *Sociocultural and historical contexts of African American English* (pp. 21–52). Amsterdam, Netherlands: John Benjamins.
- Nakamura, L. (2002). *Cybertypes: Race, ethnicity, and identity on the internet*. London, UK: Routledge.
- Nguyen, D., Seza Dođruöz, A., Rosé, C. P., & de Jong, F. (2016). Computational sociolinguistics: A survey. *Computational Linguistics*, 42, 537–593.
- OED, The Oxford English Dictionary Online*. (2018). Oxford, UK: Oxford University Press.
- Podesva, R. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics*, 11, 478–504.
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.R-project.org>.
- Rampton, B. (1995). *Crossing: Language and ethnicity among adolescents*. London, UK: Longman.
- Rampton, B. (2009). Interaction ritual and not just artful performance in crossing and stylization. *Language in Society*, 38, 149–176.
- Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010). Classifying latent user attributes in Twitter. *Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents* (pp. 37–44). Toronto, ON: Conference on Information and Knowledge Management.
- Rosenthal, S., & McKeown, K. (2011). Age prediction in blogs: A study of style, content, and online behavior in pre- and post-social media generations. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* (pp. 763–772). Portland, OR: The Association for Computational Linguistics.
- Schler, J., Koppel, M., Argamon, S., & Pennebaker, J. W. (2006). Effects of age and gender on blogging. *AAAI Spring Symposium: Computational approaches to analyzing weblog (6)*: 199–205. Palo Alto, CA: Association for the Advancement of Artificial Intelligence.
- Sebba, M. (2007). *Spelling and society: The culture and politics of orthography around the world*. Cambridge, UK: Cambridge University Press.
- Shifman, L. (2014). *Memes in digital culture*. Cambridge, MA: MIT Press.
- Sierra, S. (forthcoming). Linguistic and ethnic media stereotypes in everyday talk: Humor and identity construction among friends. *Journal of Pragmatics*.
- Squires, L. (2010). Enregistering internet language. *Language in Society*, 39, 457–492.
- Street, M. (2018). Do not erase black femmes in your history of gay slang. Last accessed April 2019 at <http://www.papermag.com/gay-slang-history-black-femmes-2578325972.html>.
- Tatman, R. (2015). #go awn: Sociophonetic variation in variant spellings on Twitter. *Working Papers of the Linguistics Circle of the University of Victoria*, 25, 97–108.
- Thomas, E. R. (2007). Phonological and phonetic characteristics of African American Vernacular English. *Language and Linguistics Compass*, 5, 450–475.
- Thomas, E. R., & Bailey, G. (2015). Segmental phonology of African American English. In S. Lanehart (Ed.), *The Oxford handbook of African American language* (pp. 403–419). Oxford, UK: Oxford University Press.
- Wieling, M., Grieve, J., Bouma, G., Fruehwald, J., Coleman, J., & Liberman, M. (2016). Variation and change in the use of hesitation markers in Germanic languages. *Language Dynamics and Change*, 6, 199–234.
- Winford, D. (2015). The origins of African American Vernacular English: Beginnings. In S. Lanehart (Ed.), *The Oxford handbook of African American language* (pp. 85–104). Oxford, UK: Oxford University Press.

- Winfrey-Harris, T. (2012). Move over, Antoine Dodson! The ironically racist internet presents Sweet Brown. Last accessed April 2019 at <http://clutchmagonline.com/2012/04/move-over-antoine-dodson-the-ironically-racist-internet-presents-sweet-brown/>.
- Woolard, A. (2008). Why dat now?: Linguistic-anthropological contributions to the explanation of sociolinguistic icons and change. *Journal of Sociolinguistics*, 12: 432-452.

How to cite this article: Ilbury C. “Sassy Queens”: Stylistic orthographic variation in Twitter and the enregisterment of AAVE. *J Sociolinguistics*. 2020;24:245–264. <https://doi.org/10.1111/josl.12366>