



12-2016

## "I'm a spawts guay": Comparing the Use of Sociophonetic Variables in Speech and Twitter

Rachael Tatman

Follow this and additional works at: <https://repository.upenn.edu/pwpl>

---

### Recommended Citation

Tatman, Rachael (2016) "'I'm a spawts guay": Comparing the Use of Sociophonetic Variables in Speech and Twitter," *University of Pennsylvania Working Papers in Linguistics*: Vol. 22 : Iss. 2 , Article 18.  
Available at: <https://repository.upenn.edu/pwpl/vol22/iss2/18>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/pwpl/vol22/iss2/18>  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

## **“I’m a spawts guay”: Comparing the Use of Sociophonetic Variables in Speech and Twitter**

### **Abstract**

This paper compares the rates of use of phonological variation across speech and, through variant spellings, Twitter data. Speech data came from New York sports personality Mike Francesa and one of his fans parodying him. The fan's tweets, which include a large number of variant spellings, were also analyzed. Two patterns emerged. For the most salient variables, th-stopping and r deletion, the fan overshot Francesa's use in speech and used them at that same higher rate in his tweets. However, for less-salient variables the fan used them at the same rate in speech as Francesa and used them at a far lower rate in their tweets. This suggests that the rate of use of a variable on Twitter can be used to investigate salience.

# “I’m a spawts guay”: Comparing the Use of Sociophonetic Variables in Speech and Twitter

Rachael Tatman\*

## 1 Introduction

Computer mediated communication, especially on Twitter, shows many of the same types of variation observed in speech. This includes variation at every level of language use, from discourse to dialect-based phonological variation (Tatman 2015). However, noting the similarity in variation between speech and Twitter data leads one to wonder: are they directly comparable?

It does seem clear that Twitter users are not encoding every phonetic aspect of their speech in writing. Several factors prevent this. The first is the standardized spelling of English. Especially given the use of modern smart-phones and browsers, where spell-check, auto-correct and auto-complete are ubiquitous, the pressure to conform to standard spellings is hard to ignore. A second factor is simple lack of awareness. While some phonetic variation is highly salient, the definitional lack of awareness of sociolinguistic indicators (Labov 2006) means they won’t be found in Twitter data. As a result, any sociophonetic variation we do see represented on Twitter must be both highly salient and intentionally included. The question I will explore here, then, is not *are* speakers encoding sociophonetic variation, but *when* they encode specific salient features, how faithfully are they mirroring their use of those features in speech?

This project tackles this question by using three separate but comparable sources of data. The first is speech from a well-known sports announcer (Mike Francesa), the second speech by a fan mimicking him and the third are tweets by that same fan, again mimicking Francesa. As a result, we can compare relatively unself-conscious speech from Francesa with both speech and tweets in a performance register. This will allow us to determine how faithfully the tweets are encoding highly salient sociophonetic variables. Since both Francesa and the fan are from New York, there are many salient variables available to them.

## 2 Background

### 2.1 Previous Work on Twitter

Previous work on language use on Twitter has found that it shows much the same patterns of variation as we find in speech. This includes racial variation (Eisenstein 2015), accommodation (Johnson 2013) and style-shifting (Schnoebelen 2012). This paper will focus on another well-established speech feature that is also robust on Twitter: dialectal variation. While other work has mainly focused on syntactic and lexical variation (Grieve 2009, Russ 2012), there is also evidence that sociophonetic variation is encoded on Twitter through variant spellings (Tatman 2015). Once it had been established that sociophonetic variation showed up on Twitter at all, the obvious next question was this: how closely does it pattern with speech data?

## 3 Methods

While some varieties of English, such as Scots, have a robust variant spelling system, most do not. In order to find a rich source of variant spellings that specifically target phonological variation, it is usually necessary to find examples of reported speech from a nonstandardized dialect. This type of “eye dialect” has a long history (Krapp 1919, Preston 1985) and is prolific on Twitter, especially when representing the speech of specific celebrities.

---

\*Work Supported by NSF Grant DGE-1256082. I would also like to extend my thanks to Nicole Chartier, the UW Sociolinguistics brownbag and everyone who gave me comments at NWAV44.

This creates an excellent environment for directly comparing speech and Twitter usage. Audio from the celebrity can be used as acoustic data, while variant-spelling-heavy tweets from their fans can be used as the written data. However, these aren't precisely parallel data sets; the celebrity's speech is (presumably) unself-conscious, while the tweets are a specific type of evocative performance register. In order to have really parallel data, it would be ideal to have four separate data sets: 1) speech from a celebrity who uses a nonstandardized dialect, 2) tweets from that celebrity, 3) speech from someone mimicking the celebrity and 4) tweets where that person mimicked the celebrity. While I was unable to find an instance of this, the data discussed here cover groups 1), 3) and 4). In addition, they have the advantage of being very parallel in subject matter, intended audience and style.

### 3.1 Sources

#### 3.1.1 Mike Francesa

The celebrity in question is Mike Francesa, a popular New York sports radio personality. He is a white, 61-year-old (born 1954) man from Long Beach, NY. He also has a strong New York identity and is a speaker of New York English. One way in which he expresses his identity as a New Yorker is in his in-depth coverage of New York sports teams, such as the Rangers (a National Hockey League team).

Francesa is best known for hosting a call-in sports radio show on WFAN, which has a very loyal fanbase. The show was originally called "Mike and the Mad Dog show" but is currently in its second iteration as "Mike's On: Francesa on the FAN." The show was also broadcast nationally on the Fox Sports television channel in 2014 and 2015. Many segments from the show have been uploaded to YouTube by fans of the show, which is where the audio data discussed here come from.

#### 3.1.2 Mongo Nation and Zaunisms

What makes Francesa a particularly good celebrity to focus this study on is his large and active fan community. This is a densely-connected community of practice with a large presence on Twitter. The community self-identifies as the Mongo Nation, and members make use of a number of style markers and hashtags. Among the latter are #mongonation and #mikezaun. In addition, some fans briefly made use a cluster of hashtags that followed the form #[cityName]Zaun to parody Francesa as if he were from different cities in the US while still maintaining his use of New York English features.

The use of highlighting phonological NYE features through variant spellings is of central interest to this project. Fans refer to these variant spellings as "zaunisms." Along with the hashtags mentioned above, this is one of the most common markers of membership in this community. Skillful use of zaunisms is a necessary skill for central members of this community, and central members are often asked to arbitrate on them or praised for their ability to produce them. Some particularly committed fans even incorporate zaunisms into their Twitter handles. While there is a handful of set forms, such as "rain jizz" for "Rangers," proficient users of this genre consistently produce novel ones.

#### 3.1.3 Mike Zaun

One of these proficient users, and a central member of the Mongo Nation community, is Bill Buchanan (@BigActionBill on Twitter, hereafter Mike Zaun or Zaun). Zaun is probably best known in the community for having released a series of parody videos in the persona of "Mike Zaun," a pun on "Mike's On." In these videos he parodies Francesa, offering political and social commentary at different points in history, including 1776, 1862 and 1943. These were very well received within the community and many members commented on how well he was able to mimic Francesa's speech patterns. An example of this, which also includes a zaunism, can be seen in (1).

- (1) I'm so jealous that you sound so much like Mike without even trying. I'd be doing Zaun impressions at the supermawket [supermarket]. (RNs.Funhouse 2015)

Though Zaun is much younger than Francesca (26, born 1989), he shares many other demographic characteristics with him, including race, gender and geographic location. Zaun is from Massapequa, NY, less than twenty miles from Francesca's Long Beach location. As a result, they also share many dialect features.

In addition, as mentioned above, he is an active Twitter user and commonly uses zaunisms. Almost 50% of his tweets include at least one zaunism, which makes him a particularly rich source for these variant spellings.

The audio data was taken from one of the parody videos discussed above, and the tweet data was downloaded from his Twitter account. These data sets are very parallel: they represent one person engaging in the same type of mimicry behavior across mediums.

### 3.2 Twitter Data

Twitter data was collected via the Twitter API (application program interface) using an R script written with the TwitteR package (Gentry 2015). The 100 most recent tweets (on May 22, 2015) were downloaded from Zaun's public Twitter account. Both the script used to collect them and the tweets themselves are available on the author's GitHub.<sup>1</sup>

The tweets were then annotated by hand. Each tweet was annotated for several characteristics: whether or not it included variant spellings, whether or not it included each of the features discussed below and, for each feature, how many possible environments there were for that variable to surface. For instance, (2) would first have been annotated as containing variant spellings. Then, for each variable of interest, it was marked for the proportion of use. For  $\delta$  stopping (2) was marked 0/2, since out of two possible environments ("the" and "this") it was encoded in neither. The same procedure was applied to each feature in Table 3.5.

- (2) befoah [before] we get to the graduates I'd like to touch on this rainjizz [Rangers] story (4)

The following conventions for representing tweets are used in the remainder of the paper. Standardized spellings are in square brackets and are the author's annotations. All @ tags have been removed for ease of reading, though they remain in the original data set. And, finally, for tweets which are not followed by a citation, the values after a tweet in parentheses are the index of that tweet in the database.

### 3.3 Variant Spellings

One potential wrinkle in looking at variant spellings in Twitter data is that it is entirely possible that a variant spelling might actually be a different lexical item. An example of a variant spelling which has a distinct meaning from the word it was based on is "werk," which no longer means the same thing as "work." (Although some non-fluent users of the form, particularly Twitter accounts run by marketing departments, don't adhere to this otherwise robust usage distinction.) "Werk" is an expression of approval, especially for female-coded gender performance. This can be seen in (3) and (4), where the term is applied to drag queens and a female professional wrestler. This quite different from the use of "work" to describe labor, as in (5).

- (3) oh yeah! but I'm gunna treat it like I'm watching Rupaul's drag race and yell things like 'YAAass Kween', 'Werk!' and 'Slaay Mama' (Sammie.Hickman 2016)
- (4) Oh HAAAYY!!! There's my lil firecracker [two explosion emoji] @AlexaBliss.WWE! Looking so damn fierce!! #WERK #WWENXT (NiaJaxWWE 2016)
- (5) Americans work more, take fewer vacations, & have worse benefits than any other industrial country. The #FightFor15 is more than justified. (DrJillStein 2016)

<sup>1</sup><http://github.com/rctatman>.

We can contrast this with the difference between “working” and “workin,” which can be thought of as different forms of the same word. The variant spelling of “werk” indexes a different meaning as well as a distinct social identity. The variant spellings in the second pair, however, are meant to encode difference in pronunciation. While this may also entail a shift in style, it does not affect the semantics. This can be seen by comparing tweets such as (6) and (7), where the meaning of the two terms seems to be almost exactly the same.

- (6) hope u guys know how hard im workin for this.. means the world 2 see all of u caring so bad, love the heck outta u guys (mackenziebourg 2016)
- (7) im really proud of all my teammates and really sorry i couldn't show up well ill keep working hard for next split zz (LiquidDardoch 2016)

This can be further verified by comparing how these words are used in a larger sample of tweets. One thousand English language tweets containing each of the four target words (“werk,” “work,” “working” and “workin”) were sampled using the Twitter API. (Non-English tweets were specifically excluded because “werk” is a word in Danish and German.) For each group of 1000 tweets, the 500 most frequent words, excluding common function words, were calculated. The figure below summarizes how many of those five hundred most commonly co-occurring words were shared between sets of tweets containing each pair of words. In other words, it is a very rough approximation of how similar the usage of each pair of words is. While this is a simplistic measure of content, it does pattern as expected: “werk” shares fewer commonly co-occurring words than the other three. This supports the assertion that tweets with the various forms of “work” in them pattern together more closely than those with “werk,” suggesting that the latter is in fact a different word entirely.

workin-	100	183	165	0
working-	116	217	0	165
work-	112	0	217	183
werk-	0	112	116	100
	wèrk	work	working	workin

Figure 1: How many of the 500 most commonly co-occurring words overlap between tweets containing each word. The darker the square, the fewer shared words, excluding the center diagonal. Note that tweets with “werk,” in the first column and last row, share fewer co-occurring words with tweets with all forms of “work.”

However, it is not sufficient to assume that each variant spelling is entirely in one or another camp. It is necessary to carefully examine the context and community in which the word is used. Take “jawn.” In (13) (see Section 4.3) we see it used as a spelling variant for “John,” which is clear from context; Zaun is parodying Francesca making a commencement speech at a college and unfavorably comparing it to St. John’s, Francesca’s *alma mater*. However, “jawn” is also a separate lexical item. The term is found mainly in and around Philadelphia, where it is strongly associated with regional identity (Eisenstein 2015). Without a careful examination of both the tweet itself and the community it was produced for it would no have been possible to distinguish between these two uses.

Ultimately, the clearest and most accurate way to determine whether a variant spelling is a new lexical item is to closely observe the community which uses it. This does require considerable time investment, certainly more than even large-scale text mining, but is necessary in order to correctly characterize the use of variant spellings.

### 3.4 Speech Data

Audio data was taken from YouTube videos. For parallelism, recordings which were approximately the same length and which were recorded temporally close to one another were chosen (Francesca 2013, Buchanan 2013).

The first clip, which was recorded and uploaded by one of his fans, was taken from Francesca's radio show. It follows the usual format of the show, where listeners call in and discuss sports with Francesca. In this clip, the subject was the Yankees baseball team, which had been performing poorly, and specifically the performance (and speculative future performance) of Robinson Canó. The caller disagreed with Francesca's assessment and the two talked over each other for a few minutes before the caller was cut out and Francesca continued to talk alone. He was particularly upset with fans who had sold their season seats due to the team's poor performance.

The second clip is a parody of Francesca's show set in 1776. In it, Mike Zaun, who is portrayed as a royalist dressed in a red coat and powdered wig, takes calls and discusses issues of the day. He suggests the rebels should trust Benedict Arnold, that King George will crush the rebellion and takes calls from General Cornwallis and a prank caller who makes a joke about John Hancock. Though the subject matter is different from the first clip, they are stylistically very similar.

The recordings were transcribed at the sentence level by hand in Praat, and then forced aligned at the phone level using FAVE-align (Rosenfelder et al. 2011). These alignments were hand-checked. Then, for each variable which was represented in the Twitter data, all the possible environments for it were marked as a new tier in the TextGrid. Finally, each possible environment was marked for whether or not the variable of interest was observed there. This allowed for a calculation of rate of use in the speech data, parallel to the measures made for the tweets.

### 3.5 Observed Variables

An initial content analysis of Twitter data revealed a large number of phonological variables (stereotypes) associated with New York English (Labov 2006), which are summarized in Table 3.5. Note that, as there were 100 tweets collected, the numbers in the center column are also the percent of collected tweets which included each feature.

Variable	# of Tweets	Example
[ɪ] deletion	38	beah [beer] (9)
ð Stopping	9	duh [the] (8)
Backed /aɪ/	7	woyld [wild] (25)
Distinct [ɔ]	5	nawt [not] (14)
g-dropping	4	swarmin (40)

Table 1: Stereotyped NY features represented in Twitter dataset.

Though not a dialect marker, g-dropping did show up in both Twitter and speech data and was included for completeness. It patterns with the backed /aɪ/ and distinct [ɔ], as seen in Figure 3.

## 4 Results

### 4.1 [ɪ] Deletion and ð Stopping

Both of these features are stereotypes of New York English and, in these datasets, they showed the same overall patterns of use.

ð stopping may have originated with European immigrants, and remains an ethnic marker (Newlin-Łukowicz 2013), but is generally more associated with working-class and regional New York identity.

[ɹ] deletion in post-vocalic positions is both one of the best-known and mostly widely studied stereotypes of New York English. It is socially stratified, with the highest levels of [ɹ] deletion in working class speakers (Labov 2006). The high levels of r-lessness from both speakers is thus not surprising. Especially given the subject matter (sports), the use of a traditionally working-class feature serves as a way of garnering covert prestige and an air of authentic regional identity. This is perhaps particularly important given the high proportion of New York residents who weren't born in the city.

Both of these features were extensively represented in the Twitter data, as can be seen in the following tweets. ð stopping is generally represented by replacing “th” in the spelling with “d.” While there is more variation in how [ɹ] deletion is represented in these variant tweets, it almost always involves removing the letter “r,” and sometimes replacing it with an “h.” These spellings are pretty faithful representations of how the variable is used in speech. Note that in (8) the r's in “Truce Torre [True story]” have not been removed in the orthography. It is only when the r's are in post-vocalic position, as in “nevah” and “aih” that they are removed from the variant spelling.

- (8) Truce Torre [True story]. I once provided ordio [audio] that was gonna be yoozed [used] as a funny bit awn da show. Nevah made it on thee aih. (32)
- (9) Weah [we're] not doin da mets today!! Get lawst! (36)

The rate of use of both r-lessness and th-stopping was greater in the parody and Twitter data than in Francesa's speech. This overshoot was statistically significant ( $\chi^2(2, N = 82) = 8.11, p < 0.05$  and  $\chi^2(2, N = 106) = 6.32, p < 0.05$ , respectively). This suggests that Zaun is playing up these features intentionally as part of his dialect performance.

#### 4.2 Backed /aɪ/

Backed /aɪ/, as in (10) through (12), is very commonly represented in the tweets. In speech it is characterized by a very backed nucleus, often surfacing as /aɪ/. In the Twitter data it is usually represented by inserting a rounded vowel, such as “o” or “u” in the variant spelling.

This is an established feature of New York English (Gordon 2004), but it is interesting that this feature is highlighted by this community while the more stereotyped NURSE diphthongization, as in “toity toid,” is not. This may be due to the relative frequencies of each environment.

- (10) ...then we go to go to adventure land...and went on awl the rides with no loynes [lines] ok (18)
- (11) was very koynd [kind] ok (37)
- (12) I'm a spawts guay [sports guy]...a better spawts guay than you #KING (26)

In their speech, both Zaun and Francesa show a high level of use of the backed variant. The nucleus was noticeably backed over 80% of the time by both speakers. However, in the Twitter data there's a much lower rate of use of this variable. It is represented in variant spellings under 60% of the time. In particular, Zaun's lower rate of use of this variable in tweets than speech was statistically significant,  $\chi^2(2, N = 52) = 5.96, p < 0.05$ .

#### 4.3 Distinct [ɔ]

A distinct [ɔ] vowel, sometimes discussed as the “cawfee” vowel, is another well-known feature of NYE (Gordon 2004), and was used extensively in tweeting. It was usually written by including a “w” in the variant spelling, perhaps to indicate that it should be rounded. However, the use was far from categorical. Indeed, even within the same tweet, such as (8) and (9) above, one [ɔ] will be marked in the variant spellings (“lawst,” “awn”) while another (“on”) will not be changed. This suggests that,



while it is certainly a feature that tweeters are aware of, it is sometimes slipping beneath their radar. In fact, a variant spelling of [ɔ] was only used in 50% of the possible environments for it.

- (13) listen folks um I hate to break dis to u but dis is a nice college not a great one. It is not st jawns [John’s] ok (2)

In the speech data, the vowels were first annotated by hand by a phonetically-trained listener without the low back merger. Due to the complexity of the low-back vowel space (Newman and Kelly 2015), only words which belonged to the historical THOUGHT and LOT classes were included in this stage of analysis. The annotations were double-checked by plotting the target vowels in an F1-F2 space. As can be seen in the figure below, the vowels that were annotated as distinct also fell into two groups acoustically. Both speakers had a categorical split between [ɔ] and [ɑ].

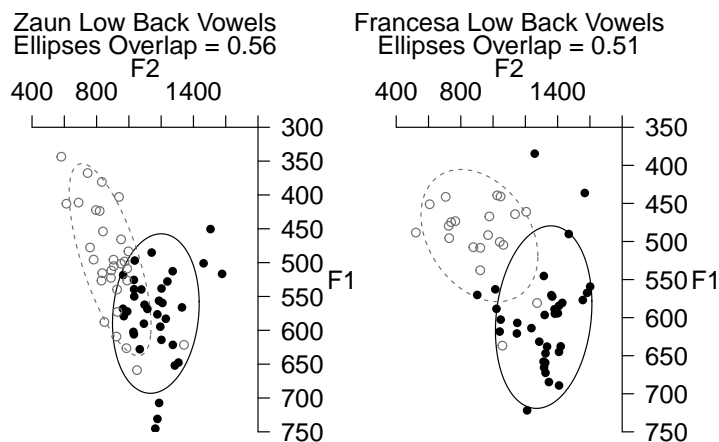


Figure 2: Vowels from LOT and THOUGHT class words produced by both Zaun and Francesa. The color indicates the annotation of each vowel: light gray for [ɔ], black for [ɑ]. Ellipses overlap was calculated, not with the 80% confidence interval shown, but with ellipses that included all data points from each category.

So, as with backed /aɪ/, both speakers almost always use the variable in their speech. However, the rates of use are much lower in the Twitter data. Again, Zaun is much less likely to use this variable in his tweets than his speech,  $\chi^2(2, N = 48) = 9.86, p < 0.05$ .

### 5 Discussion

To summarize the discussion of individual variables above: there are two ways that the use of a variable can pattern across speech and tweets. In the first, as with [ɪ] deletion and ð stopping, Zaun used a variable at a higher rate in his speech than Francesa did. If this was the case, then the variable was used at a comparable level in his tweets as his speech. On the other hand, if Zaun used a variable at the same level as Francesa in his speech, then he used the variable at a far lower rate in his tweets. This is summarized in Figure 3. G-dropping is also included in this figure, and it patterns with the latter type.

So are tweeters using variables at the same rates in speech and tweets? It depends, and what it depends on seems to be salience. Very salient variables are used at the same rate in mimicked speech and tweets, but at higher rates than they are in non-mimicked speech. This high use of very salient variables patterns with previous work on performance registers (Schilling-Estes 1998). (Note that performance register here refers specifically to *dialect* performance. It could certainly be argued that Francesa is performing in his radio show, but that would not necessarily fall under the heading of “performance register.”) This is especially appropriate given that Zaun and Francesa are from the same dialect area; Zaun isn’t adopting a new dialect, but rather playing up his use of certain features in service of a persona.

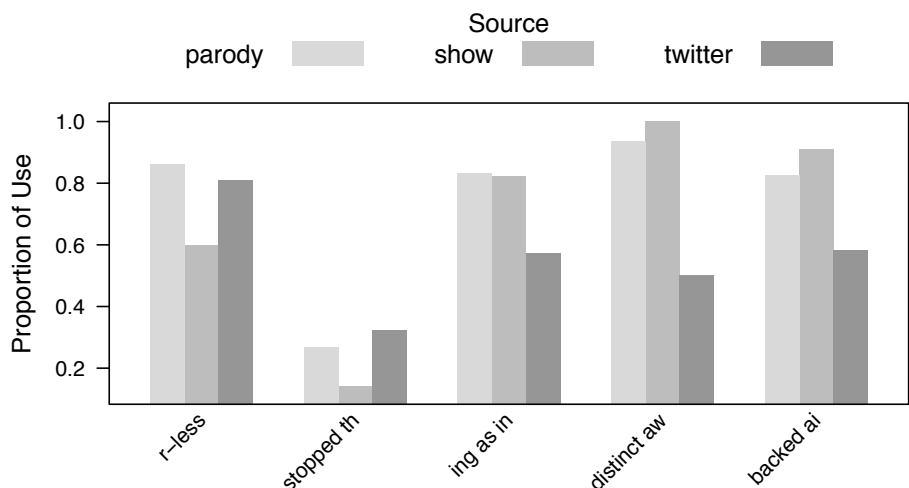


Figure 3: Proportion of use of each variable across three different sources. Note the two distinct patterns of use.

On the other hand, even though the registers are the same, less salient (but still stereotyped) variables are used at lower rates in tweets than speech. However, the rates of occurrence of these variables are the same in parodying and non-parodying speech. This patterns with earlier findings by Tagliamonte and Denis (2008) and Honeybone and Watson (2013).

We can thus propose a two-tier system of salience. The top tier is made up of very salient features which are used at higher rates in both performance speech and tweets. The second tier has features which are still very salient, but which seem to fly under the radar at least some of the time. While speakers use them at the same level in performance and non-performance registers, it can't be said that this is due to ignorance of these features since they are still highlighted in writing.

### 5.1 What Can We Learn from Variant Spellings?

Variant spellings are not a faithful phonetic transcription of speech. This should not be surprising, especially given the discussion in section 3.3. However, variant spellings on Twitter do seem to be encoding some information about sociophonetic variation. Specifically, they allow us to determine which features a speaker considers especially relevant for performing a dialect.

While they can not be used to supplant traditional sociolinguistic interviews, sociophonetic insights from Twitter can supplement them. In particular, they may be useful for identifying stereotypes of a particular language variety. It would be especially interesting to see if similar information could be obtained by including a written component during data collection, perhaps by asking participants to write like someone talks or transcribe a recording.

### 5.2 Directions for Future Work

This paper has focused on the relative use of variables which appear in both speech and Twitter. What it has not discussed are those variables which appear in speech but not in tweets, and this will need to be addressed in future work. It seems likely that variables which are not included in writing may simply be less salient, but it's equally possible that some stereotyped variables are not written due to being hard to write. Durham (2015), for example, notes that while prosody is often discussed on Twitter as an important feature of Welsh English, it is rare to see someone attempting to depict it in writing.

Another area which has not been touched on here is comparing how a Twitter user represents their own dialect as opposed to others' dialects. Zaun was a native user of the dialect he was attempting to portray in writing; how might less fluent users of the dialect represent it?

Since Twitter data includes the time and date that it was produced, and stretches back several years, it would also be interesting to look at longitudinal data, especially for variant spellings which have become stable, such as “rain jizz.” Do multiple possibilities emerge before one is adopted by a community? This would be in keeping with other work looking at longitudinal linguistic change in online communities (Danescu-Niculescu-Mizil et al. 2013), but has not been investigated with regard to variant spellings.

The relative ease of data collection via Twitter means a Twitter analysis can be conveniently appended to existing studies, particularly those that focus on previously undescribed variables. While the lack of a feature on Twitter does not mean that it is not salient, it can be argued that any socio-phonetic feature represented in Twitter data is salient to that Twitter user. In addition, comparing the use of the same variable across speech and tweets can help researchers determine how salient it is.

## References

- Buchanan, Bill. 2013. Mike Zaun 1776. URL <https://www.youtube.com/watch?v=ptPBCSsPJR4>.
- Danescu-Niculescu-Mizil, Cristian, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. No country for old members: User lifecycle and linguistic change in online communities. In *Proceedings of the 22nd International conference on World Wide Web*, 307–319.
- DrJillStein. 2016. Americans work more, take fewer vacations, & have worse benefits than any other industrial country. the #fightfor15 is more than justified. URL <https://twitter.com/DrJillStein/status/720681211757088768>.
- Durham, Mercedes. 2015. Representations of Welsh English online: What can tweets tell us about salience and enregisterment? Paper presented at NWA 44, Toronto.
- Eisenstein, Jacob. 2015. Systematic patterning in phonologically-motivated orthographic variation. *Journal of Sociolinguistics* 19:161–188.
- Francesca, Mike. 2013. Mike argues with a “fair-weather Yankee fan”, tells him to get lawst. URL <https://www.youtube.com/watch?v=JpH%5F31XBW%5F4>.
- Gentry, Jeff. 2015. *Twitter: R Based Twitter Client*. R package version 1.1.9.
- Gordon, Matthew J. 2004. New York, Philadelphia, and other northern cities: Phonology. In *A Handbook of Varieties of English, Volume 1*, ed. E. W. Schneider, K. Burridge, B. Kortman, R. Mesthrie, and C. Upton, 282–299. Berlin: Mouton de Gruyter.
- Grieve, Jack. 2009. A corpus-based regional dialect survey of grammatical variation in written Standard American English. Doctoral dissertation, Northern Arizona University.
- Honeybone, Patrick, and Kevin Watson. 2013. Salience and the sociolinguistics of Scouse spelling: Exploring the phonology of the contemporary humorous localised dialect literature of Liverpool. *English World-Wide* 34:305–340.
- Johnson, Ian. 2013. Audience design and communication accommodation theory: Use of Twitter by Welsh-English biliterates. In *Social Media and Minority Languages: Convergence and the Creative Industries*, ed. E. H. G. Jones and E. Uribe-Jongbloed, 99–118. Buffalo, NY: Multilingual Matters.
- Krapp, George Philip. 1919. *The pronunciation of Standard English in America*. New York: Oxford University Press.
- Labov, William. 2006. *The social stratification of English in New York City*. Cambridge, UK: Cambridge University Press.
- LiquidDardoch. 2016. im really proud of all my teammates and really sorry i couldn't show up well ill keep working hard for next split zz. URL <https://twitter.com/LiquidDardoch/status/718997189721108480>.
- mackenziebourg. 2016. . @macbourgs hope u guys know how hard im workin for this.. means the world 2 see all of u caring so bad, love the heck outta u guys. URL <https://twitter.com/mackenziebourg/status/720011550875107332>.
- Newlin-Lukowicz, Luiza. 2013. TH-stopping in New York City: Substrate effect turned ethnic marker? In *University of Pennsylvania Working Papers in Linguistics 19.2: Selected Papers from NWA 41*, ed. A. Freeman, 151–160.
- Newman, Michael, and E. Brian Kelly. 2015. LOTs of THOUGHTs on the endangered PALMs of New York. Paper presented at NWA 44, Toronto.

- NiaJaxWWE. 2016. Oh HAAAYY!!! There's my lil firecracker [two small explosion emoji] @AlexaBliss.WWE! Looking so damn fierce!! #WERK #WWENXT. URL <https://twitter.com/NiaJaxWWE/status/720406517825060865>.
- Preston, Dennis R. 1985. The Li'l Abner syndrome: Written representations of speech. *American Speech* 60:328–336.
- RNs.Funhouse. 2015. @BigActionBill I'm so jealous that you sound so much like Mike without even trying. I'd be doing Zaun impressions at the supermawket. URL [https://twitter.com/RNs\\_Funhouse/status/600817401664565248](https://twitter.com/RNs_Funhouse/status/600817401664565248).
- Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, and Jiahong Yuan. 2011. FAVE (forced alignment and vowel extraction) program suite. URL <http://fave.ling.upenn.edu>.
- Russ, Brice. 2012. Examining large-scale regional variation through online geotagged corpora. Paper presented at ADS Annual meeting, Portland.
- Sammie.Hickman. 2016. @CapnCole oh yeah! but I'm gunna treat it like I'm watching Ru-paul's drag race and yell things like 'YAAass Kween', 'Werk!' and 'Slaay Mama'. URL [https://twitter.com/Sammie\\_Hickman/status/719601605323190273](https://twitter.com/Sammie_Hickman/status/719601605323190273).
- Schilling-Estes, Natalie. 1998. Investigating “self-conscious” speech: The performance register in Ocracoke English. *Language in Society* 27:53–83.
- Schnoebelen, Tyler. 2012. Do you smile with your nose? Stylistic variation in Twitter emoticons. In *University of Pennsylvania Working Papers in Linguistics 18.2: Selected Papers from NWAV 40*, ed. H. Prichard, 117–125.
- Tagliamonte, Sali A, and Derek Denis. 2008. Linguistic ruin? LOL! Instant messaging and teen language. *American speech* 83:3–34.
- Tatman, Rachael. 2015. #go awn: Sociophonetic variation in variant spellings on Twitter. In *Working Papers of the Linguistics Circle 25.2: Proceedings of the 31st annual NorthWest Linguistics Conference*, ed. S. Onosson and M. Huijismans, 97–108.

Department of Linguistics  
 University of Washington  
 Seattle, WA 98195-2425  
 rctatman@uw.edu