

Language and *Latinidad* in US Spanglish hip hop

An account from YouTube Comments



Matt Garley [mgarley@york.cuny.edu]
York College / CUNY & The Graduate Center / CUNY
LASSO 2017 XLVI, Las Cruces, NM, Oct. 7, 2017



Introduction

- A fair amount of work has been done examining hip hop (and related musical genres) in the US-Latinx community from a cultural studies perspective.
- McFarland (2008, 2013) explores Chicano rap and hip hop from a largely sociopolitical standpoint
- Flores (2000) and Rivera (2003) provides background on Nuyorican identities and inclusion in hip hop
 - Rivera's distinction between island-affiliated and New York-affiliated artists:
 - those who have a stronger identification with the island and Spanish-dominant *reggaeton*, vs
 - those who more strongly identify as part of the 'core' multi-ethnic New York hip hop scene (e.g. Big Pun) and perform generally monolingually in English

Introduction

- Not much language-centered work, or work performed from a sociolinguistic standpoint considering questions about language mixing, language change, or multilingualism/translanguaging practices

Introduction

- Hip hop is particularly language-focused as a set of cultural practices, and the language of hip hop has been identified as a *resistance vernacular* (Potter 1995)
- Lyrics are extremely interesting and a good place to start as a hip hop primary text, but the culture does not only reside with the artists
- Fan practices and responses are crucial (and as a linguist who digs 'natural' language data, YouTube comments are about as messy as it gets).

Introduction

- Analyzing language in computer-mediated communication:
- Pros:
 - digital preservation of ‘low-stakes’ linguistic production,
 - interesting orthographic features and innovations
 - a space for linguistic creativity
 - relative ease of data collection
- Cons:
 - “demographically lean” (Iorio 2009)
 - less structured discourses
 - no phonology, prosody (except as represented in orthography)

Research Questions

- What are Latinx hip hop artists and fans doing with varieties of English and Spanish in their lyrics and in reactions to those lyrics?
- What is the connection between artist and fan (code-mixing/translanguaging) practices?
- In what ways do Latinx hip hop artists and fans of artists from different US regions and of distinct Latinx heritages differ in their use of Spanish, English, hybrid, and creative features?
- How are linguistic practices of Latinx hip hop in the 2010s changing from previous decades (particularly the 1990s and 2000s)?

Previous work

- Garley (in preparation) qualitatively examines lyrics and YouTube comments in individual videos from Chingo Bling, the Beatnuts, and Sick Jacken to examine revealed language ideologies and orientations toward language
- Found that this set of artists and fan communities too different in decade, style, and 'scene' for appropriate direct quantitative comparison, but suggested avenues for future research, e.g.:

Previous work

- Attitudes/ideologies/functions surrounding Spanish as used by non-Latinx artists like The Game or Kendrick Lamar (also in prep)
- Examination of mediatization/enregisterment of features, metalinguistic discourses, attitudes and ideologies in interviews (published and/or elicited) with Latinx hip hop fans and artists
- Direct quantitative comparison of linguistic behaviors in YouTube comments with larger corpora comprised of several videos

Previous work

- Attitudes/ideologies/functions surrounding Spanish as used by non-Latinx artists like The Game or Kendrick Lamar (also in prep)
 - Examination of mediatization/enregisterment of features, metalinguistic discourses, attitudes and ideologies in interviews (published and/or elicited) with Latinx hip hop fans and artists
- **Direct quantitative comparison of linguistic behaviors in YouTube comments with larger corpora comprised of comments from multiple videos [YOU ARE HERE]**

Methods

- Combination of quantitative and qualitative methods
 - Qualitative/exploratory work → what's out there? what sort of behaviors and situations can be found?
 - Quantitative work → what's most common? how often does it occur?
 - More qualitative work → what exactly is going on with these behaviors or situations?
 - Qualitative/ethnographic work → why though?
 - Quantitative work → can we see these reasons in the data?
 - (repeat and remix as necessary)

Methods – Artist selection

- Selected three artists to center individualized corpora
 - Criteria: US-based, code-switching in multiple tracks, reasonably popular, Active primarily post-2010, multiple official (uploaded by artist, label, hip hop magazine, etc.) videos available on YouTube
 - Selected for geographic diversity within US

Methods – Artist selection

- Selected videos by or featuring these artists to collect sizable subcorpora of YouTube comments
 - Criteria: >100k views, song features code-switching, Latinx identity forward
- Collected comments (using Python script / Google's API) from specified videos until I had >2500 per artist corpus and at least 2 videos from each artist

The Artists – Bodega Bamz

- Bodega Bamz
(Nathaniel DeLa Rosa)
- from Spanish Harlem, of Puerto Rican and Dominican heritage.
- Co-founder of ‘Tanboys’
movement/brand



“Tanboys is a movement of Latinos who are proud and powerful individuals who want to put our culture back on the forefront of this music. It’s a worldwide thing, it’s not a rap group. Any Latino who’s a proud and powerful individual, any Latino who ain’t a hater, any Latino who’s about uplifting his culture is a Tanboy. You’re born with that, it’s in your DNA.” (Ahmed 2013, interview on *Complex.com*)

Bodega Bamz

“Don Francisco” video

<https://youtu.be/GaMwGQFDHPA?t=30s>

Sample lyrics:

“Yeah, my skin tan, my fade low

My eyes green for those *pesos*

Eat *chicharrón sin huesos*

I’m Pitbull and I’m Tego

With the hoes like *Don Francisco*

I’m Lavoe and I’m Big Pun

All in a pot, little *Sazón*

Mix it all around, and *Bodega’s* what you got”

Bodega Bamz

- Bodega Bamz rarely raps in Spanish for lengthy stretches—most codeswitches are single word/phrase, straightforwardly emblematic, and tend toward Latinx cultural references (*chancletas* in one song, *plátanos* in another, and *chicharrón sin hueso* in ‘Don Francisco’).

- *Video corpus for comments:*

Bodega Bamz – Don Francisco

Bodega Bamz – Don Francisco remix feat. French Montana

Bodega Bamz – P.A.P.I.

Bodega Bamz – El-Rey

No Panty (Bodega Bamz, Joell Ortiz, Nitty Scott) – Hola

Joell Ortiz ft. !llmind, Emilio Rojas, Bodega Bamz, Chris

Rivers – Latino Pt. 2

(3244 comments collected)

The Artists - Snow tha Product

- Snow tha Product (Claudia Feliciano)
- from San Jose, CA, moved to San Diego and later Dallas/Ft.Worth
- 2nd gen. Mexican-American, parents from Michoacán and Zacatecas

(Snow Tha Product 2016, popsugar.com article)

“[...] whenever artists like myself try to break away from that, ‘don’t look at me as a female rapper, don’t look at me as a Latino rapper, just look at me as a rapper,’ our own community starts to be like, ‘Why aren’t you repping?’ so you can’t win either way. [...] No, it’s going to take away my Mexican card, but when you’re kind of assessing the whole situation like, ‘I’m still here, I’m still on, *si habla Espanol pero*,’ with no help at all, I’m just going to have to take it a little broader and represent everyone, you know.” (Ontoneyo 2016, thesource.com interview)



Snow tha Product

“AyAyAy!” video

https://youtu.be/8_MOdouFugw?t=1m3s

Sample lyrics:

Yo, look, my mama kicked me out, cause I was partyin' like
YOLO

Woke me up for church, I was hungover, she's like, "oh, no"

Gave my ass a bucket, *trapeador*, and got the *Cloro*

Pitbull jacked my mom, since I was a baby she yellin 'Como!?'

She like: *¿Cómo que no? Ponte a trapear*

Todo lo que compro te doy, ponte a limpiar

[continues in Spanish]

Snow tha Product

- Snow tha Product releases a number of songs in near-monolingual English, but also raps in Spanish (usually in the chorus or a separate verse) on other songs ('Hola', 'I Don't Wanna Leave')
- Due to the popularity of Snow's videos and large number of comments, the Snow tha Product corpus consists of comments from two videos.
- Video Corpus for comments:
Snow tha Product - "AyAyAy!"
Snow tha Product - "I Don't Wanna Leave Remix"
(7190 comments collected)

The Artists: Chingo Bling

- Chingo Bling (Pedro Herrera)
- Comedian, rapper, actor from Houston, TX
- 2nd gen. Mexican-American, parents from Tamaulipas



“It's almost like a Chicano film or clothing line, record label, sometimes it has a connotation as something dirty, tainted, not mainstream. A bastardized version of something else. It's not really Hip Hop. [...]”

“Other things that I wanted to express, that I had interest in, styles of clothing or music, different types of flow, subject matter, I would throw it all in this Chingo Bling umbrella, and it would get confusing to everybody. 'Well Chingo changed.' No, not really, it's just you're hearing another part of me. That's why people trip out if I'm wearing a different kind of hat. 'Wait a minute, you're supposed to wear the cowboy hat 24/7.' That just comes with the territory. [...]” (Dante 2013, latinrapper.com interview)

Chingo Bling

“Cerveza” Video

<https://youtu.be/fl9z2thdW3c?t=10s>

- Sample lyrics:

America be lookin like a party *y no invitan*

So we crash the party and be like ‘we brought *fajitas*’

And I beat it beat it up to *Sonora Dinamita*

Throw a party on the *torta* be like ‘girl I brought *fajitas*’

[...]

We pull up at the party hoppin out we brought *mariachis*

But she ain’t even hungry cause I just fed her kolache

Don’t lie, I look *firme*, *por favor* believe me

Your *prima* and your sister said they wanna double team me

Chingo Bling

- Chingo Bling tends toward exaggeration and extremes in terms of language mixing, blending this with the lyrical wordplay valued in hip hop. He releases a large number of parody videos of popular songs, but some original works as well. The corpus of videos concentrates on these.

- Video corpus for comments:

Chingo Bling – Brown & Proud

Chingo Bling – Cerveza (2 official videos)

Chingo Bling – Puro Pinche Pari (2 official videos)

Chingo Bling ft. Kap G & Dat Boi T – PE\$O\$

*Baby Bash ft. Low G, Chingo Bling, Juan Gotti – El Pinche
(2900 comments collected)*

Methods – comparison wordlists

- used wordlists from FU Berlin's Corpora from the Web (COW) to classify words in comments
 - COW has some pros:
 - Multiple languages available (major european langs)
 - Spanish corpus collected from multiple Spanish-speaking countries' TLDs (not just .es)
 - Corpora were crawled from the Web, so we don't see the restrictive forms we'd find from a spellchecker wordlist, etc. More messy, more diversity
 - And cons:
 - These are automatically collected, not hand-sorted by language
 - Many common English/Spanish words appear in both wordlists (particularly the most common words in each language)

Methods – comparison wordlists

- Took top 50k frequent wordforms (not lemmas) from COW corpora for English and Spanish
- Removed non-alphabetic characters and added duplicates without diacritics for Spanish words with diacritics (áéíóúñü)
- Intersected the wordlists and removed words that appeared in both as well as non-alphabetic characters, resulting in:
 - English wordlist (33204 wordforms)
 - Spanish wordlist (41038 wordforms)
 - Intersection wordlist (5037 wordforms)
 - The English Web apparently has some Spanish on it, and the Spanish Web apparently has some English on it (not particularly surprising)
 - The intersection wordlist functions as ‘stopwords’ – very common words ignored in corpus analyses

Methods – dumb classification

- Wrote a Python classifier to go through comments and compare with these wordlists, yielding for each comment (and each corpus) a count and ratio of:
 - Known English wordforms
 - Known Spanish wordforms
 - ‘Ambiguous’ wordforms (from the intersection list)
 - Unknowns (typos and the interesting stuff)
- The most frequent words ended up in the ‘ambiguous’ category. That’s OK, because while this is a really bad classifier for individual words, it’s an OK classifier for texts.

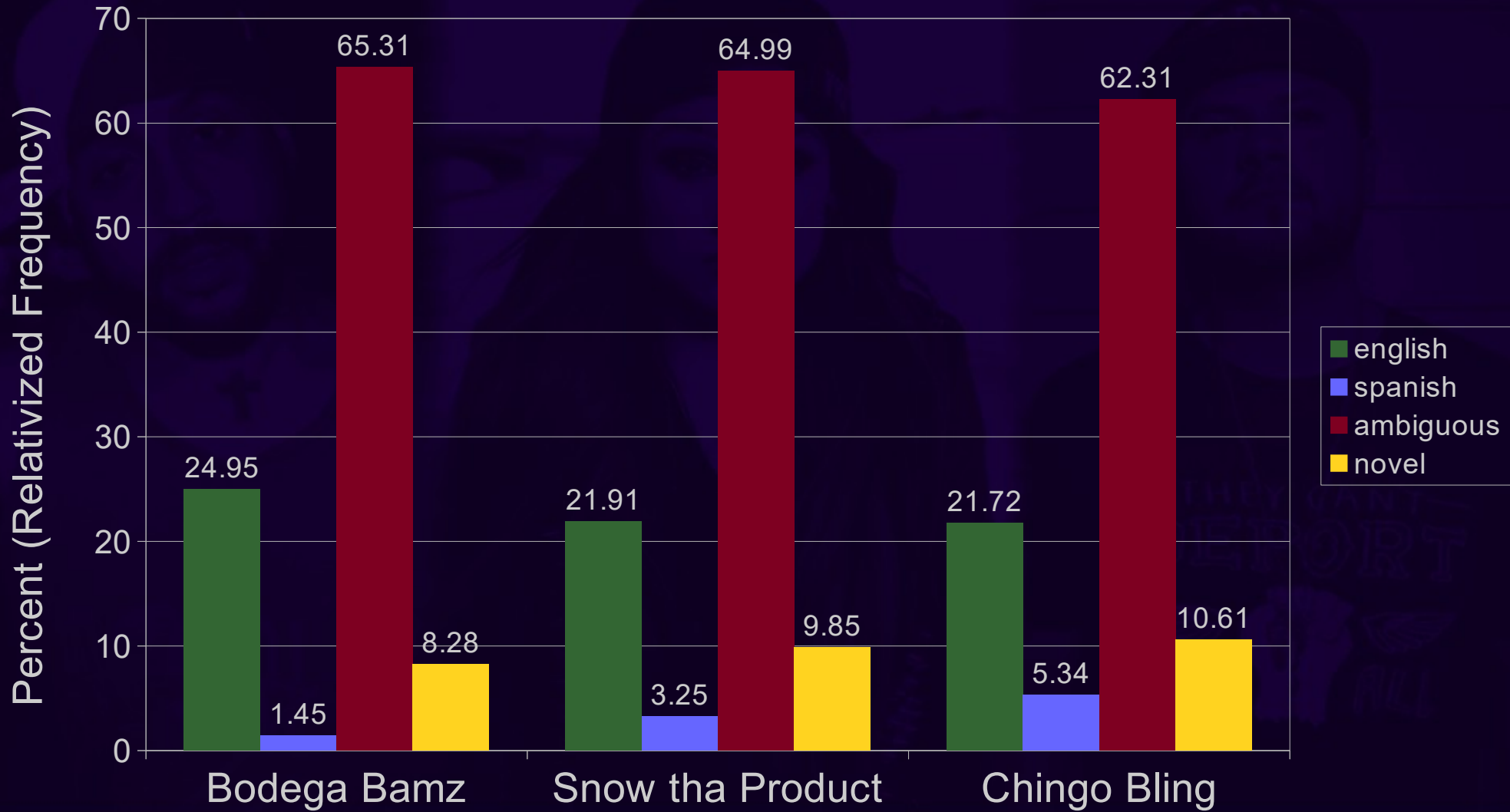
The Corpus

- 3 corpora, one for comments on each artist's videos
- Removed non-alphabetic characters
 - Anything Unicode counts as alphabetic is kept (so diacritics are unaltered)
 - Loses punctuation, emoji, hashtags, symbol spellings with \$ for s, etc.--but I have the original comments
- Ignored comments with no alphabetic content

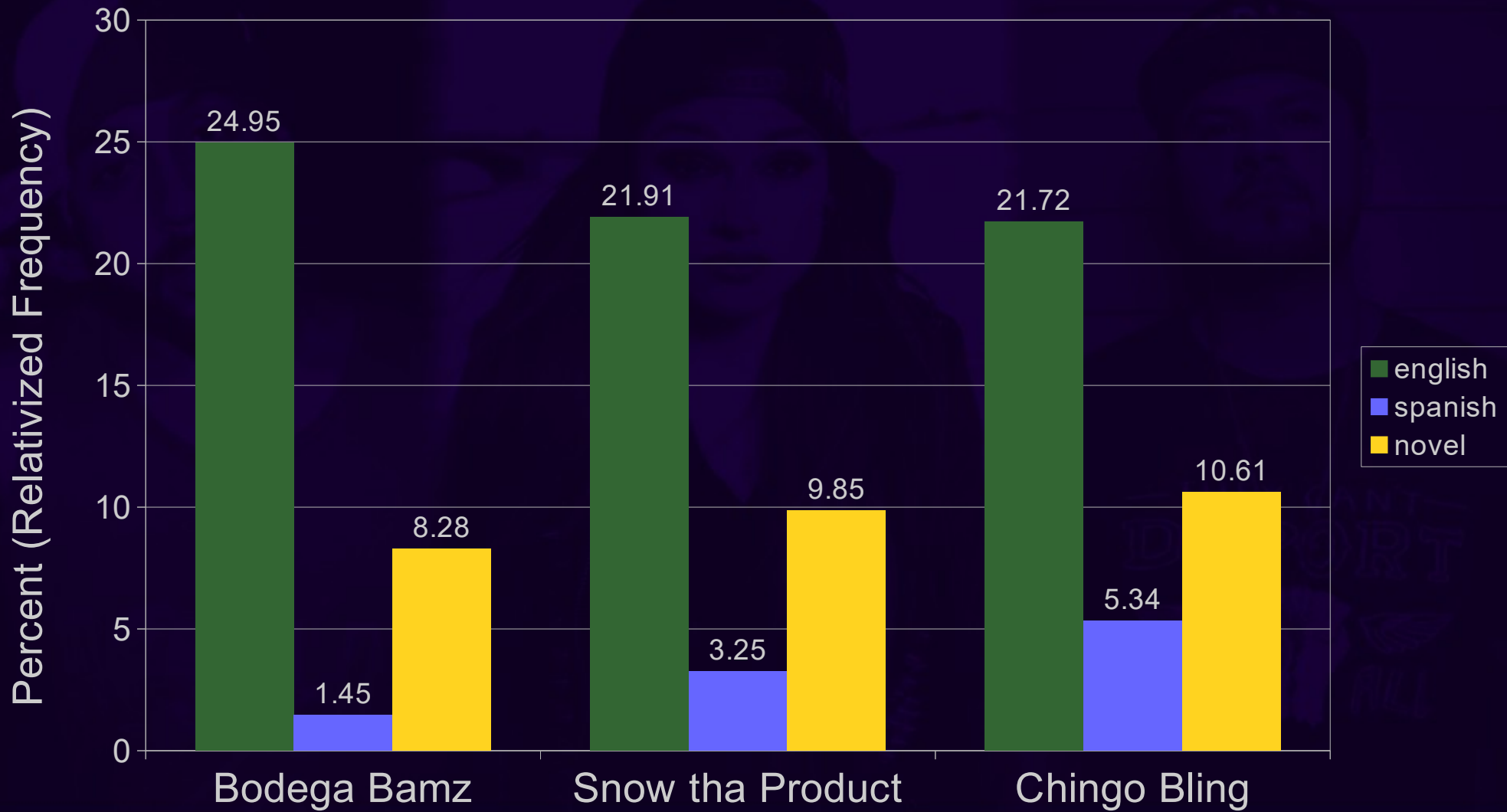
The Corpora

- Bodega Bamz corpus (after cleaning):
 - 3,577 comments, 50,857 words
- Snow tha Product corpus:
 - 6,954 comments, 80,209 words
- Chingo Bling corpus:
 - 2,848 comments, 35,086 words
- Not entirely even, but all of decent size

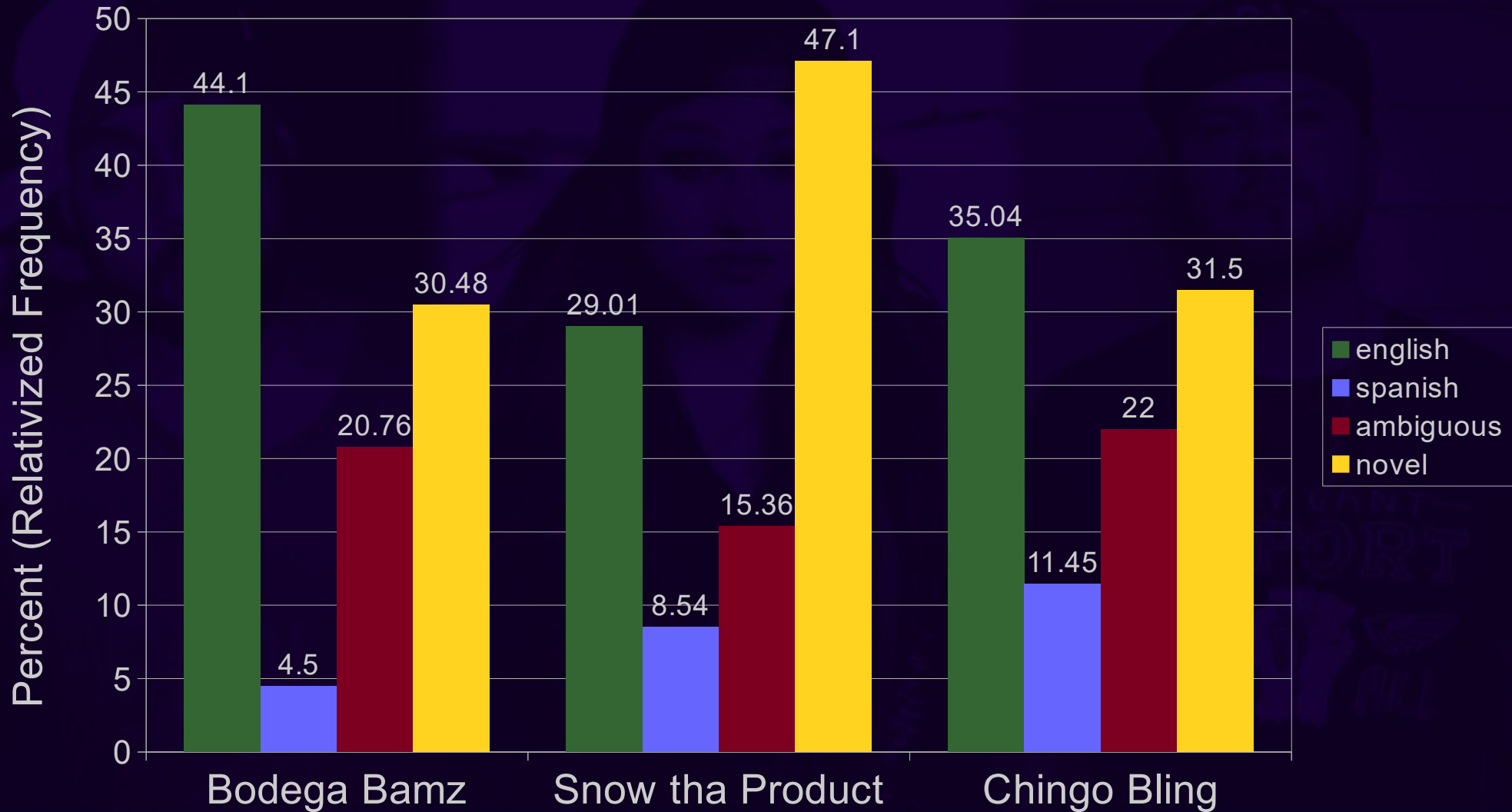
The overall picture – token classification



The overall picture – token classification



The overall picture – wordform classification



Overall picture: commentary

- So: The 'ambiguous' words, those found in both COW wordlists, are high-frequency, but aren't as diverse as the novel words
- The English-to-Spanish ratio (keeping in mind that these are lower frequency, more standard words for each language) is different in each case:
 - English is dominant in all cases, but the comments on Bodega Bamz' are much more imbalanced.
 - Bodega Bamz: ~17:1 Spanish:English
 - Snow tha Product: ~7:1 Spanish:English
 - Chingo Bling: ~4:1 Spanish:English
- This doesn't account for the language of ambiguous or novel wordforms per the classifier: it could be the case that Bamz commenters are using less recognizable Spanish forms (but this is unlikely looking at comments)

Quantitative Results

- Fan communities' language breakdowns (though a crude approximation) reflects the degree to which each artist incorporates Spanish in their lyrics.
- Bamz → smallest number (types and tokens) of recognized Spanish forms in comments
- Snow → reasonably large jump upward, reflecting a larger proportion of comments with recognized Spanish content.
- Chingo → smaller increase from Snow corpus, highest proportion of Spanish comments

Quantitative Results

- This is a fairly crude classifier, and the corpora are not entirely normalized in ways that would be useful, so:
 - No p-values today (sorry)
 - Lots and lots of noise: commenter names (when tagging each other in threads), lyrics, artist names, emoji, imperfect word chunking due to punctuation issues
- So what is this sort of thing good for?
 - Discovering ‘interesting’ comments for qualitative consideration in a large dataset
- Sorting the classified corpus for comments with highest token or ratio values for Spanish and Novel/Unknown categories sifts the most interesting (to me) stuff to the top

Comments

Comments from Bodega Bamz Corpus:

- *chicharron si eres latino entenderas papi mexicano puertorriqueño colombiano venezolano chileno peruano paraguayo* we all the only family go make dollars on usa nothing more *perro*
- good video *rey ya falta muy poco para poder verte y escucharte por acá en México*
- i fucks wit you bamz you my n---- holding it down *pa los latinos* respect from *el catracho del cinco zero cuatro nuevo orleans representando*
- sorry joell but you re saying *ay dos mio* and the text is saying *ay dios mio*
- king of manhattan wearing *chancletas* tan flag *que bonita bandera* thats that latin shit shouts from brasil tan nation

Comments

Comments from Bodega Bamz Corpus:

- yyyyyoooo bodega talking that talk son yo bamz we need u in houston soon my dude gully greasy grimy *t que bonita bandera manito* all day any latin flag
- dropping off some caribbean love *pana ya tu sabes cuba*
- this shit is *muy fuego*
- *respeto desde argentina rey*
- *mi gente* bringing nothing but that *fuego*
- bodega keeps bringing that *fuego*
- *bendiciones* from spain big up bodega bamz

Comments

Comments from Bodega Bamz Corpus:

- *boriqua asta la muerte que lo sepa*
- i could relate to all this guys tracks bro sheesh c me paran lo pelo latinostandup
- this is straight *fuego* boy
- the brazil watching you
- greetings from mexico good song

Comments - Commentary

- Most of the comments I've selected to share are those which have multilingual content or Spanish content—keep in mind that there's less of that overall in the Bamz corpus.
- The comments from the Bodega Bamz corpus reflect a few patterns:
- Shoutouts from various countries/locales, including Mexico, Spain, Brazil, a Honduran from New Orleans, etc.
- Lots of emblematic uses of words like *fuego*, which are more borrowing-like (but still salient as Spanish in origin) in hip hop discourse
- Some metalinguistic commentary on Joell Ortiz' pronunciation (and a longer comment excluded here which comments on Bamz' 'fake' 'country' pronunciation).

Comments - Commentary

- Expressions of pan-Latino and pan-Caribbean pride suggest fan uptake of Bamz' claimed and performed identities
- A few orthographic colloquialisms/linguistic identity-markers: *boriqua* (vs. *boricua*) *pa* (cf. *para*) *asta* (cf. *Hasta*)
- We'll see these patterns in the next two subcorpora, but more.

Comments – Snow tha Product

- *en valencia españa se te respeta mucho a toda la gente que junto a ti hace que esto llegue a nosotros muchas gracias un abrazo y seguimos a la escucha deaz*
- *mexicanos keep it g saludos desde méxico puro fuego*
- *me gustan tus canisiones yo ablo español y ingles y soy mexicana*
- this lady here got this lyrics heavy i fucking sware she got skills she just keeps bringing it to the fucking table so underrated but fucking dope *al cien si no eres hombre no vas a ganar* ha ha ha lol hell yea mad respect to all my fellow independent women tho

Comments – Snow tha Product

- i need this song on my spotify rly *increíble la canción ya tienes nuevo seguidor*
- *y yo no entiendo lo que dice en ingles pero es una maravilla escucharla besos desde argentina*
- its kinda mixed up the lyrics are over the top *pero esta perrona y el coro mas chingon todavia al [emoji]*
- i see tha progression u r making with your word play *especialmente en español sige dando les en la madre nieve h town love*
- i just shit my pants *mija eres la mas chingona te chingas asta los vatos tambien* you get down homegirl your shit is fire
- snow killed it as usual *que chingón lo hace en español también*

Comments – Snow tha Product

- *la mejor saludos desde colombia*
- *puro fire con esta mujer*
- *saludos desde españa big up*
- *i m in love with her she is the best saludos desde monterrey mx*
- *the best thing about this girl is that she kill it in english y tambien rompe en español dominican republic love what you do hiphop life*
- *i don t think i ve ever clicked so fast snow you killed that mucho amor de nuevo mexico*

Comments – Commentary

- The Snow tha Product corpus has more frequent and more extensive codeswitching
- Snow’s popularity has brought her to the attention of hip hop fans worldwide—a lot of *saludos desde* comments from Spanish-speaking countries and the rest of the world.
- Sorting by high ‘novel/unknown’ factor shows a fair number of ‘third-language’ comments entirely in German, Turkish, Arabic, and other languages
- Here we see a single-word switch to ‘fire’ (cf. *fuego*) with matrix lg. Spanish

Comments – Chingo Bling

- for reals *puro sentimiento de orgullo mi llena* when i hear this *casi hasta lágrimas de tanto felicidad y orgullo* even more with this video right on chingo *no hay mejor orgullo que de ser mexicano hacia arriba y hacia adelante porque no pueden construir las paredes suficientemente altos levante sus puños marrones* im brown and im proud
- im mexican *y vive en los estados unidos y de seguro ni habla español completo y ni habla ingles completo* u know what i mean ese
- *que viva chingo bling como siempre representando la raza mexicana por eso tú eres el chingón y puro pinche sas*
- hell yea chingo always putting it down my nikka chingo keep up da good work h town north lindale i remember once wen i was like i saw yu at *la tia pancha* well now is *cumbres aver cuando nos echamos unas frias*

Comments – Chingo Bling

- fricken *güera* ass mexican born raised in south florida but boyyy itell you i am never goinq to forget where icame from my pops is *tejano y mi jefa* is from *guanajuato mx ptm mi corazon es mas puto mexicano que el nopal* itself thanx chingo for makinq songs like these so iwill never forget why iam so proud of my culture im brown im proud
- this song actually goes hard chingo and i m not even a huge fan of rap hiphop *todos los k dicen k esta cancion esta wack metanse un pepino por el* you know what lol
- say chingo ths the type of sht tht we need *ese mucho amor desde dallas tx* keep at it **homito**
- loving it u need to do a whole cd of cumbia rap *saludos desde minnesota homie*
- *pinche* chingo funny *vato* keep them coming *arriba tejas*

Comments – Chingo Bling

- i want to push this *rola* on the radio *musica culturana* net will post this song over and over brown and proud and im still waiting for my free cancan **tchirt puto**
- rap game *cantinflas pinche chingo* got skillz love the mash up of styles yooooooo when we gown see a chingo bling riff raff colabo
- chingo bling bro you need to do **choutouts** to your fav youtube fans one vid *por favor*
- *apenas salo este bideo* i ve been **bumbin** this *rolla* for a min *tienes todo tu pinche* alter egos doin cameos
- hold up great song *pero donde está el* screwed and chopped **bersion**

Comments – Commentary

- In the comments on Chingo Bling videos we see an intensified linguistic creativity/hybridity.
- Orthographic representations of Chicano pronunciation: *choutout* (x2), *bideo*, *bersion* (x2), *bumbin* (?)
- Interesting codeswitches: *por el you know what, mas ... que el nopal itself*
- Chingo Bling's identity as a comedian who uses exaggerated, stereotypical pronunciations, extensive codeswitching, and *Caló* seems to predispose fans / commenters to engage in similar linguistic practices
- Chingo Bling's videos also seem to inspire commenters to go beyond 'shoutouts' and discuss their own identities and linguistic autobiographies

Conclusions

- This preliminary quantitative-qualitative examination shows that artists' linguistic behaviors are indeed reflected in the proportion and type of language production by fans in comments
- Could be a chicken/egg situation—since the artists often come from the communities with these language practices—they are to some extent innovators, but have a larger role in mediating and circulating language practices that are pre-existing
- Regional differences are suggested from the data—a more internally diverse Spanish-speaking population in the Northeast could have an effect

Future Directions

- Different lists of es/en wordforms & improved corpus normalization
- Advanced classification: use different methods to triangulate classification (looking at strings of characters, words or using real NLP evaluation methods, etc.)
- More videos from associated artists (network expansion to affiliates)
- Seriously consider how a translanguaging approach would have ramifications for this work
- More categorization/analysis of nonstandard orthographic forms in comments
- Mini-analysis on who is shouting out whom from where

Future Directions

- Different lists of es/en wordforms & improved corpus normalization
- Advanced classification: use different methods to triangulate classification (looking at strings of characters, words or using real NLP evaluation methods, etc.)
- More videos from associated artists (network expansion to affiliates)
- Seriously consider how a translanguaging approach would have ramifications for this work
- More categorization/analysis of nonstandard orthographic forms in comments
- Mini-analysis on who is shouting out whom from where

References

- Ahmed, Insanul. 2013. “Who is Bodega Bamz?” <http://www.complex.com/music/2013/05/who-is-bodega-bamz/>
- Ontoneyo. 2016. “Snow tha Product: ‘No dude wants to get killed on their record’” The Source. <http://thesource.com/2016/06/07/interview-snow-tha-product-no-dude-wants-to-get-killed-on-their-record/>
- Snow tha Product. 2016. “Rapper Snow tha Product: ‘Immigrants don’t feel safe to have a voice, but their children can stand up for them.’” PopSugar Latina. <https://www.popsugar.com/latina/Rapper-Snow-Tha-Product-Her-Mexican-Background-42236842>
- Schäfer, Roland. 2015. “Processing and Querying Large Web Corpora with the COW14 Architecture.” In Proceedings of Challenges in the Management of Large Corpora (CMLC-3) (IDS publication server). 28–34.
- Schäfer, Roland & Felix Bildhauer. 2012. “Building Large Corpora from the Web Using a New Efficient Tool Chain.” In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12). 486–493.
- McFarland, Pancho. 2008. Chicano rap: Gender and violence in the postindustrial barrio Austin: university of Texas Press.
- McFarland, Pancho. 2013. The Chican@ hip hop nation: politics of a new millennial mestizaje. East Lansing: Michigan State University Press.
- Rivera, Raquel Z. 2003. New York Ricans from the hip hop zone. Palgrave MacMillan.
- Flores, Juan. 2000. From bomba to hip hop: Puerto Rican culture and Latino identity. New York Columbia University Press.
- Iorio, Josh. 2009. “Effects of Audience on Orthographic Variation” Studies in the Linguistic Sciences: Illinois Working Papers 2009: 127-140. <https://hdl.handle.net/2142/14815>

Thanks!

Contact: mgarley@york.cuny.edu

