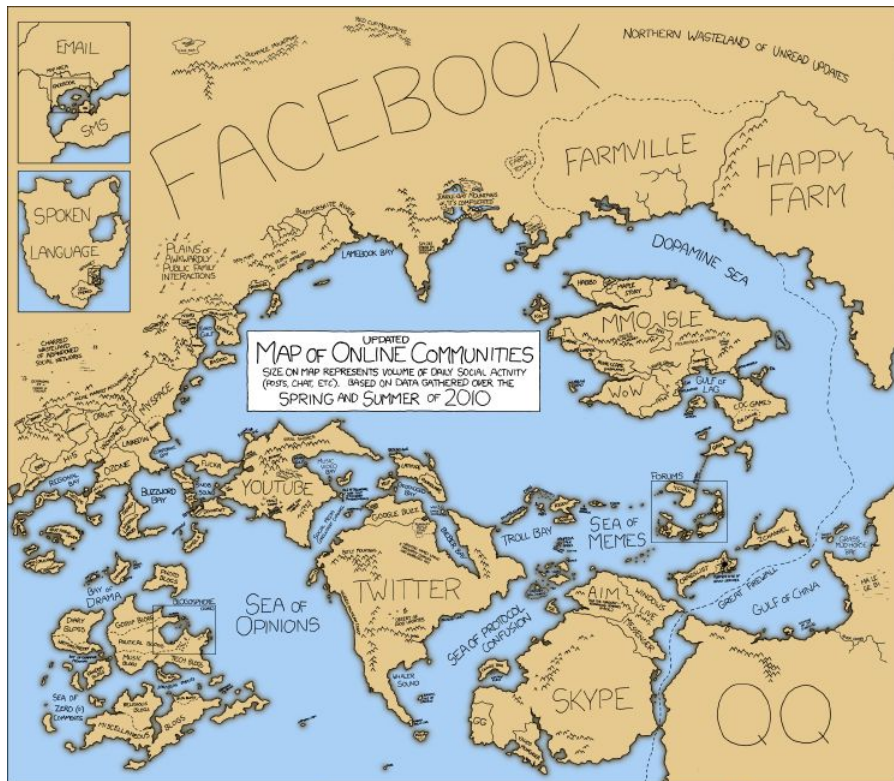


NLP for Social Media

Zach Wood-Doughty

zach@cs.jhu.edu



Outline

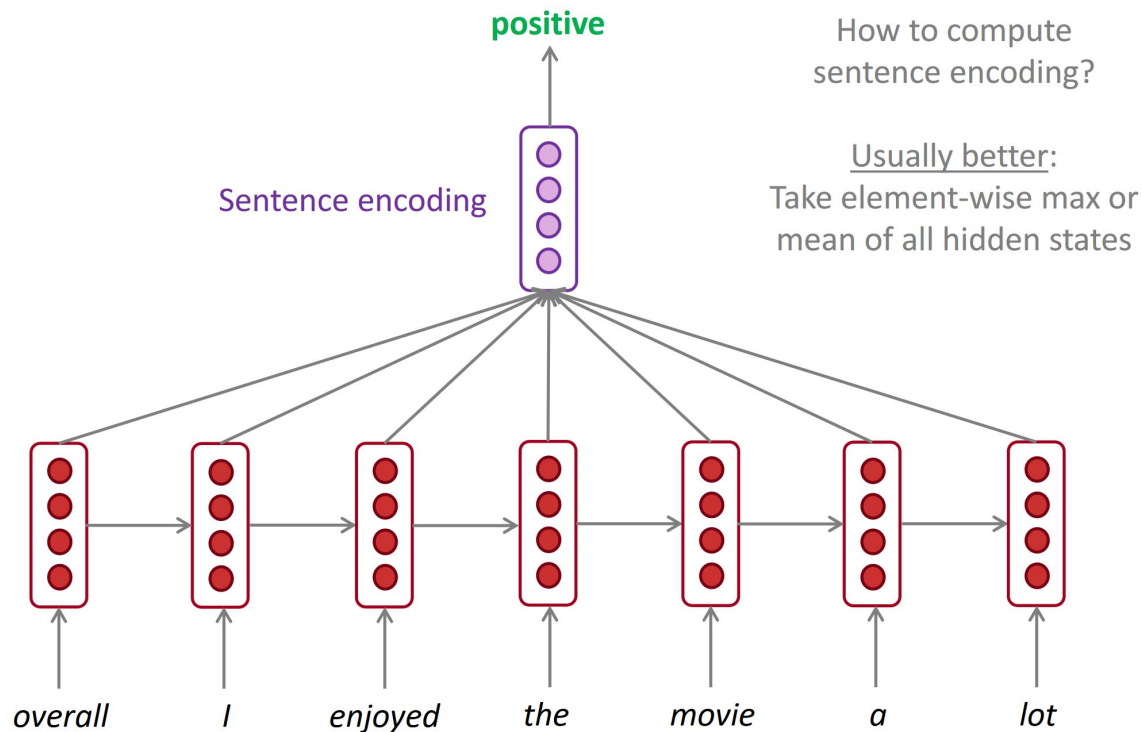
1. Challenges: What language do we speak online?
 - a. Code-switching and language identification
 - b. Informal language, capitalization and punctuation, memes
2. Opportunities: What language do we speak online?
 - a. How do we talk in a controlled setting?
 - b. How does language evolve, generally or during specific phenomena?

Outline

3. Challenge: Can't ignore speaker/audience context
 - a. Is your language offensive, or are you joking?

4. Opportunity: How do we prevent abuse/toxicity on Twitter?
 - a. Relative status and demographic attributes are essential context

Recap: Text classification



What language do we speak online?

Code-switching

Al principio del semestre el profesor de Bio nos dijo que para su clase ni de pedo podríamos tener las siguientes cosas al mismo tiempo:

1. Good grades
2. Social life
3. Sleep

Dije ok imma prove him wrong butttt now he's proving himself right 😞

Code-switching

At the beginning of the semester, my Bio professor told me it was impossible to have all three:

Al principio del semestre el profesor de Bio nos dijo que para su clase ni de pedo podríamos tener las siguientes cosas al mismo tiempo:

1. Good grades
2. Social life
3. Sleep

Dije ok imma prove him wrong butttt now he's proving himself right 😞

Informal language

I. D.o.n.t even wanna try and clean these rn I feel so disrespected



mstrinah

Replying to [@RepSwalwell](#) and [@realDonaldTrump](#)

mark about to get his Denice you were at my wedding moment form Rep Swalwell

2:03 PM · Mar 27, 2019 · [Twitter Web App](#)

12 Likes

can there be vaccine when health care is completely dead?How?!?Wey Buhari self na UK him dey run to

5:22 AM - 13 Oct 2017

Memes



sexy spooky siren 🎃

my
resume

handshake meme



not knowing how to format this

5:08 PM · Oct 12, 2019 · [Twitter for iPhone](#)



Kim Sellina

you: the creative process is a beautiful thing
me, a creative:



7:27 PM · May 16, 2018 · [Twitter Web App](#)

6.6K Retweets 24.8K Likes

Challenges: Pre-processing



sexy spooky siren 🎃

@ [redacted]

my
resume

handshake meme



not knowing how to format this

5:08 PM · Oct 12, 2019 · [Twitter for iPhone](#)



b-boy bouiebaisse ✓

@jbouie

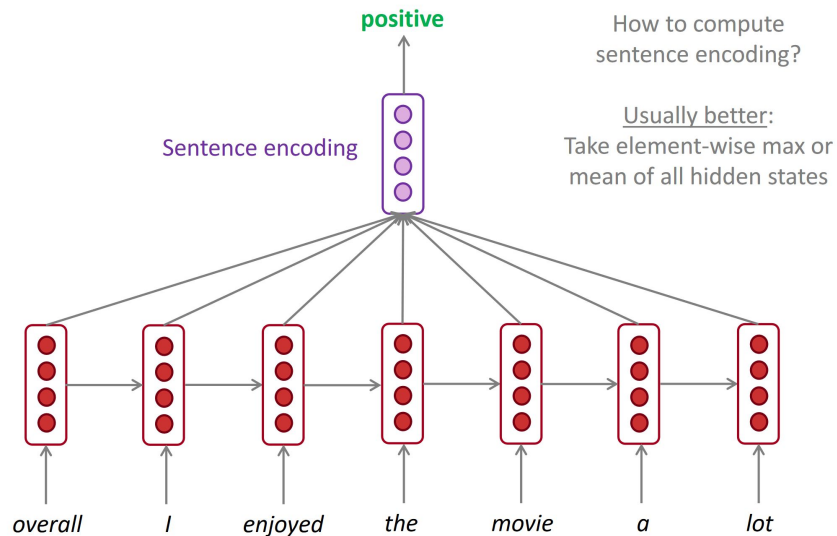
Replying to [@pkapauline1](#)

We aRe a rEpUbLic nOt a dEmOcRaCy

4:46 PM · Feb 28, 2019 · [Twitter for iPhone](#)

5 Retweets 175 Likes

Challenges: vocabulary size

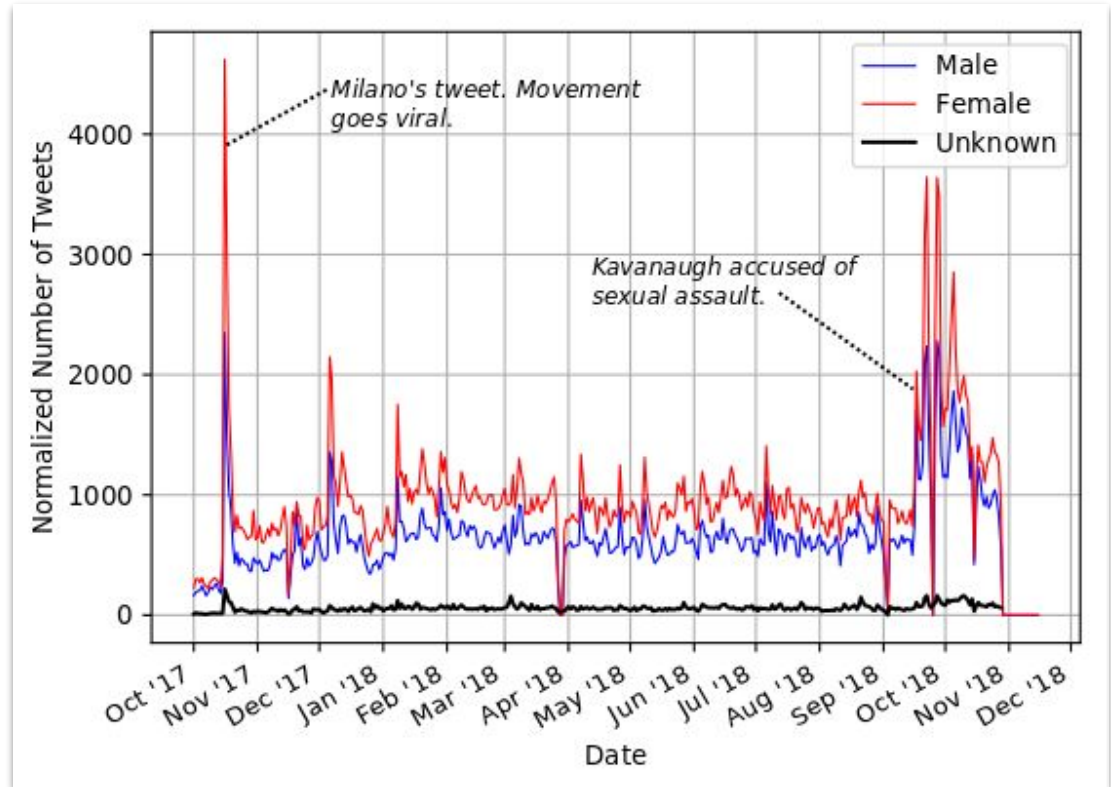


I. D.o.n.t even wanna try and clean these rn I feel so disrespected

A screenshot of a tweet from UMBC Athletics (@UMBCAthletics). The tweet text reads: "PUT SOME RESPECK ON IT! WE HAVE DEFEATED NO.1 OVERALL SEED VIRGINIA 74-54". The tweet was posted at 8:34 PM on March 16, 2018, from Charlotte, NC. It has 62,485 retweets and 208,807 likes. The interface shows a "Follow" button and a row of profile pictures of users who interacted with the tweet. At the bottom, there are icons for replies (1.8K), retweets (62K), and likes (209K).

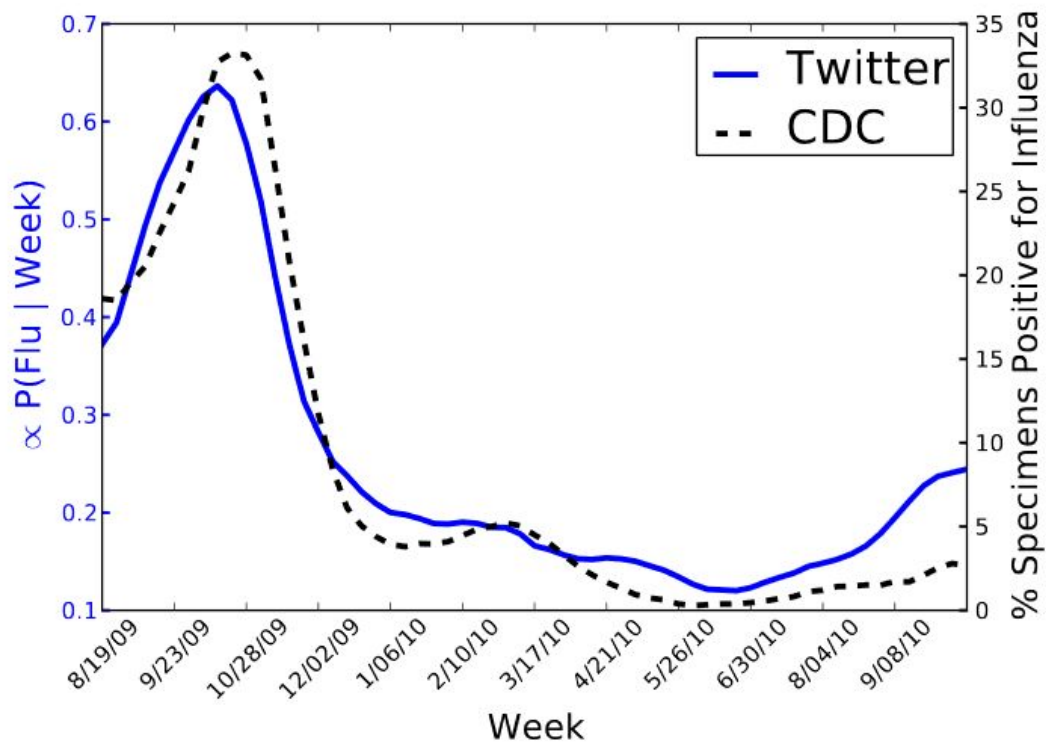
Opportunities: study language in a microcosm

#MeToo tweets over time



Opportunities: study language in a microcosm

Figure 2: Influenza rate from August 2009 to October 2010 as measured by CDC FluView (measured as percentage of specimen's that tested positive for influenza) and ATAM+'s "flu" ailment (the normalized probability of the "flu" ailment given the week): correlation coefficient of 0.958.



Opportunities: study language in a microcosm



William McNabb

@

Replying to [@JasonIsbell](#)

Legit question for rural Americans - How do I kill the 30-50 feral hogs that run into my yard within 3-5 mins while my small kids play?

12:01 PM · Aug 4, 2019 · [Twitter for iPhone](#)

Opportunities: st



Lauren (checkmark coming 11-17-19)

You still wake up sometimes, don't you Clarice? You wake up in the dark and hear the screaming of the 30-50 feral hogs

2:07 PM · Aug 5, 2019 · [Twitter for](#)



William McNabb

Legit question for rural Amer
50 feral hogs that run into m



summertime vibes dante



my milkshake brings 30-50 feral hogs to the yard

2:46 PM · Aug 5, 2019 · [TweetDeck](#)



"On the Internet, nobody knows you're 30-50 feral hogs"

Can't ignore speaker/audience context

Challenge: audience context



Ellen L Weintraub 

@EllenLWeintraub



I would not have thought that I needed to say this.



CHAIR ELLEN L. WEINTRAUB
FEDERAL ELECTION COMMISSION
WASHINGTON, D.C. 20463

Statement Regarding Illegal Contributions From Foreign Governments

June 13, 2019

Let me make something 100% clear to the American public and anyone running for public office: It is illegal for any person to solicit, accept, or receive anything of value from a foreign national in connection with a U.S. election.¹ This is not a novel concept. Electoral intervention from foreign governments has been considered unacceptable since the beginnings of our nation. Our Founding Fathers sounded the alarm about “foreign Interference, Intrigue, and Influence.” They knew that when foreign governments seek to influence American politics, it is always to advance their own interests, not America’s. Anyone who solicits or accepts foreign assistance risks being on the wrong end of a federal investigation. Any political campaign that receives an offer of a prohibited donation from a foreign source should report that offer to the Federal Bureau of Investigation.

Challenge: speaker context



b-boy bouiebaisse 

@jbouie

Replying to [@pkapauline1](#)

We aRe a rEpUbLic nOt a dEmOcRaCy

4:46 PM · Feb 28, 2019 · [Twitter for iPhone](#)

5 Retweets **175** Likes

Challenge: conversations



b-boy bouiebaisse ✓ @jbouie · Feb 28

the electoral college is bad, the framers knew it was bad, and circumventing it should be a top tier concern for anyone who cares about democracy



Opinion | The Electoral College Is the Greatest Threat to Our Democracy
It has not stood the test of time.

[nytimes.com](https://www.nytimes.com)

128

968

3.4K



pkapauline @ [redacted] · Feb 28

we are a republic, not a democracy

22



b-boy bouiebaisse ✓

@jbouie

Replying to @ [redacted]

We aRe a rEpUbLic nOt a dEmOcRaCy

Challenge: interpersonal relationships

Denise McAllister ✓ @McAllisterDen · Mar 25

Can someone explain to me the purpose of The View? It seems to me to be a roundtable of delusional mental midgets ricocheting ignorance and lack of emotional regulation. ilovemyfreedom.org/meghan-mccain-... via @ILMForG



Meghan McCain Says She's Having "Existential Crisis"; Admits Media Wil...
It appears as though Meghan McCain is starting to doubt the authenticity of her role as a conservative talk show host on ABC's The View. She an...
ilovemyfreedom.org

1.7K 2K 9.5K

Meghan McCain ✓ @MeghanMcCain

Replying to @McAllisterDen and @ILMForG

you were at my wedding Denise....

10:08 PM · Mar 25, 2019 · [Twitter Web Client](#)

Challenge: toxic or abusive language

Replying to @cf [redacted]

Lol. You suck. 😂😂😂


5:33 PM · Oct 14, 2019 · [Twitter for iPhone](#)

Replying to @D [redacted] and @t [redacted]

oh cheer up !!!

1:13 AM · Oct 13, 2019 · [Twitter Web App](#)

Opportunity: prevent abuse on Twitter



AMNESTY
INTERNATIONAL

TOXIC TWITTER - A TOXIC
PLACE FOR WOMEN

Chapter 1

RESEARCH

The image shows a person's profile and hand typing on a laptop keyboard, serving as a background for the text.

◆ 98% similar to comments people said were "toxic"

You're a stupid idiot!

Conclusion

1. What language do we speak online?
 - a. Harder to build models to learn from informal language
 - b. But we can study language in a relatively controlled setting, and we can explore social movements
2. Speaker/audience context matters!
 - a. Meaning changes depending on this context
 - b. Understanding this context can let us improve these platforms



Sign Bunny

@bunnywithsign

FREE FLU
SHOTS
TODAY
BLC 4-7pm



2:05 PM · Sep 17, 2014 · [Twitter Web Client](#)

Thanks!

zach@cs.jhu.edu