

CS 2731 Introduction to Natural Language Processing

Session 27: Computational social science, digital humanities

Michael Miller Yoder

April 17, 2024



University of
Pittsburgh

School of Computing and Information

Course logistics

- Last regular class session!
- No class next Mon Apr 22
- Final project presentations next **Wed Apr 24**
 - Instructions on [project webpage](#) and in the [Google Slide deck](#)
 - Slides link also posted as an announcement on Canvas
- Project report is **due next Thu Apr 25 at midnight**

Course logistics

- Feel free to join Michael at a talk at CMU by [Yulia Tsvetkov](#) from the University of Washington on **Fri Apr 19, 2-3:20pm**
- Email Michael if you're interested (completely optional)
- More info:
<https://lti.cmu.edu/misc-pages/lti-colloquium.html>
- Talk title: “LLMs under the microscope: illuminating the blind spots and improving the reliability of language models”



Final project report rubric

Rubric category	Points
Clear motivation for the work is provided	5
Research questions and/or task definition is clear	10
Sufficient grounding in relevant related literature	10
Applicable dataset/s are chosen	5
Methods are relevant. For new approach contributions, multiple methods are compared. For dataset contributions, annotation methodology is explained	15
Results are provided. For new approach contributions, results from multiple methods (at least one baseline) are presented. For dataset contributions, this may be a single set of results from a simple classifier, or other results if discussed with the instructor	20
Discussion is provided of the results and/or the potential uses or contributions of any new datasets contributed	10
Limitations of your approach or dataset are sufficiently discussed	5
Ethical issues that may be raised by your system or dataset are sufficiently discussed	5
Project content total	85
Meets all formatting requirements. Is maximum 8 pages, not including references or group member task breakdown	10
Writing is clear	15
Writing total	25
Group member had a sufficient amount of workload in the project	15
Task and roles assigned to this group member were completed sufficiently	15
Individual contribution total	30
Grand total	140

Core tasks and applications of NLP

CORE TASKS

text
classification

representation
learning

language
models

conditional
language
models

sequence
labeling

syntactic
parsing

MODULE 2

MODULE 3

MODULE 4

MODULE 5

APPLICATIONS

machine
translation

speech recognition
& synthesis

chatbots

computational
social science

MODULE 6

Overview of today's class session

- Language in social context
- Computational social science
- Digital humanities
- Time to complete OMETs
- Project time
 - Michael available to answer questions


Language is embedded in social context

What types of social situations do you encounter language in?

What types of social contexts?



English chatting very easy













Alan Black
has your back.

DWI
Criminal Defense
Family Law
Personal Injury

A professional headshot of Alan Black, a middle-aged man with a mustache, wearing a dark suit, light blue shirt, and yellow tie. To the right of the photo is a white background with the text "Alan Black has your back." in black. Below this is a red rounded rectangle containing the text "DWI Criminal Defense Family Law Personal Injury" in white.

What types of social contexts?

Euro quals	2:45 PM ET ESPN3  BEL  CYP	2:45 PM ET ESPN3  NED  EST	2:45 PM ET ESPN2/ESPN3  WAL  HUN	2:45 PM ET ESPN3  GER  NIR	2:45 PM ET ESPN3  POL  SVN
------------	--	--	--	--	--

On-Line Homework Instructions for Physics 1250-1251

Homework will be submitted and graded via the online software package WebAssign.

ACCESSING WEBASSIGN:







Open Internet Explorer or Netscape Navigator or Mozilla Firefox (Some other browsers may have difficulty), and go to the WebAssign login page (<https://www.webassign.net/osu/student.html>). (The WebAssign login page at <https://www.webassign.net/login.html> will get you to the site above as well, but the OSU login site should be your primary site.)


What types of social contexts?




What types of social contexts?

What's happening?


      [Tweet](#)


 **Odd Pittsburgh** @OddPittsburgh · 59m
[#Pittsburgh](#) in 1930





City of Pittsburgh

Trends for you

Trending in United States 
#DevinNunesIsAnIdiot
53.9K Tweets

Trending in United States 
#AdviceForBoomers
4,684 Tweets

Trending in United States 
Vindman
Trending with: Lt Col Vindman, Colonel Vindman, Col Vindman

Trending in United States 
#2009v2019
3,035 Tweets

[Show more](#)

What types of social contexts?



World

U.S.

Politics

N.Y.

Business

Opinion

Tech

Science

Health

Sports

Arts

Books

Style

How Not to Plot Secret Foreign Policy: On a Cellphone and WhatsApp

U.S. officials expressed wonderment that Rudy Giuliani ran an “irregular channel” of Ukraine diplomacy over open cell lines and apps penetrated by the Russians.

2h ago [494 comments](#)



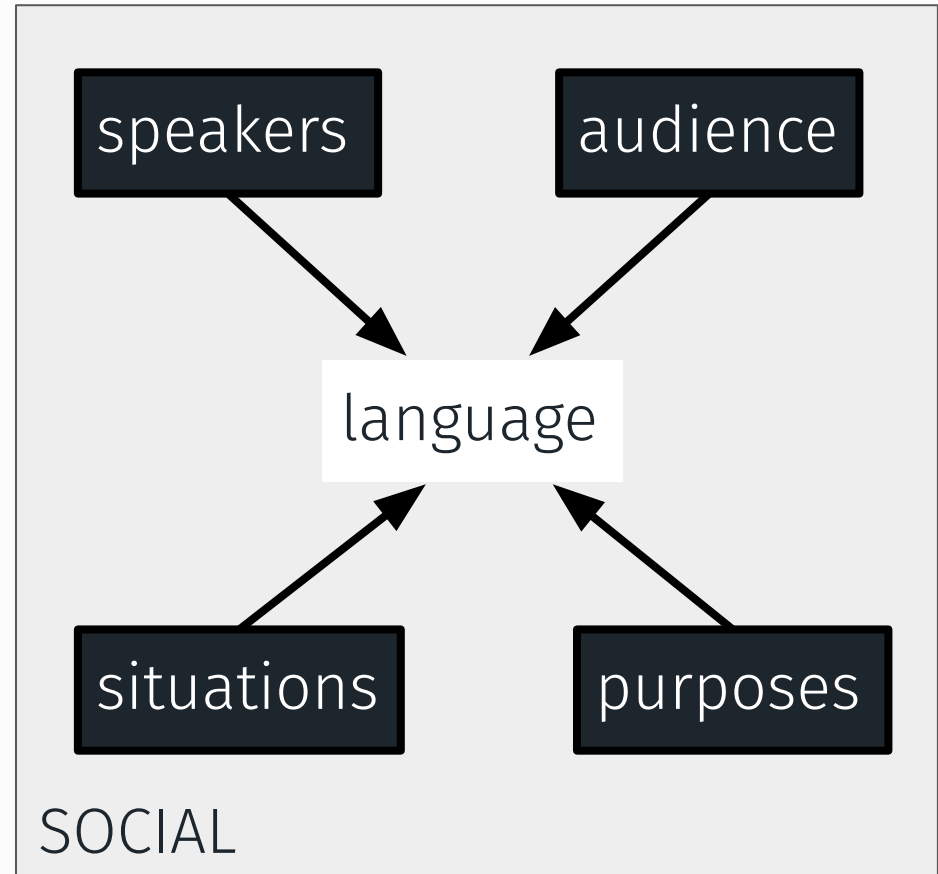
Rudolph W. Giuliani, President Trump's personal lawyer, makes a living selling cybersecurity advice.
Doug Mills/The New York Times

Who is Kurt Volker, President Trump's former special envoy to Ukraine?

28m ago

Tim Morrison, a hawkish aide loyal to Mr. Trump, will also testify this afternoon.

42m ago

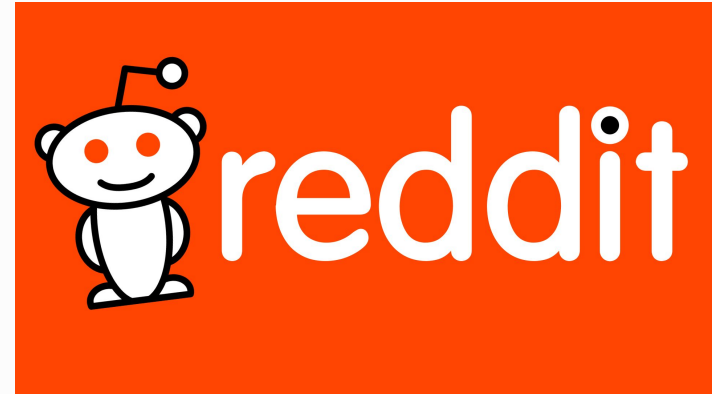


NLP + social science: applications

hate speech detection

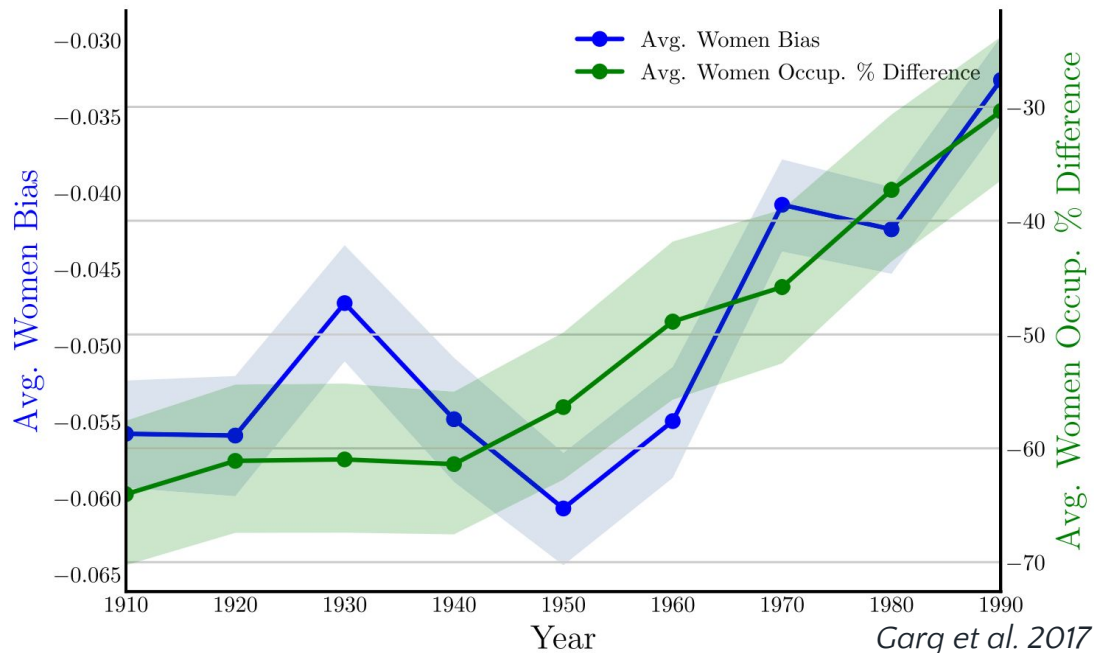


community norms



NLP + social science: applications

fairness and bias

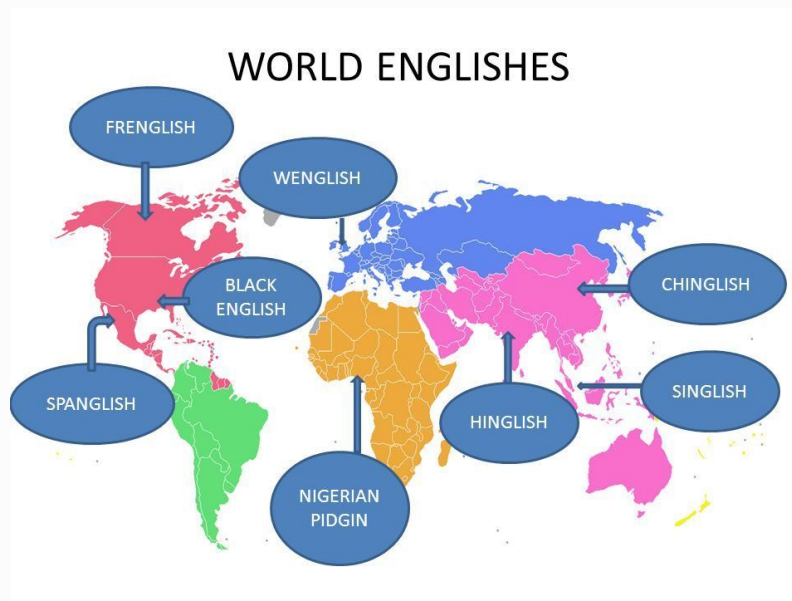


media framing



<https://criticalmediareview.wordpress.com/2015/10/19/what-is-media-framing/>

dialectal NLP tools



Computational social science

Computational social science

- Investigating (modeling, analyzing) social phenomena with computational tools [Cioffi-Revilla 2017]
- CSS goal: find out something about **people** (social science)
- NLP goal: build computational tools that can process or produce language
 - Social NLP: build tools that address language in social context
 - Hate speech detection, sarcasm and irony, dialects and language variation

Computational social science: methods and data

- Observational studies, not lab or survey studies



large datasets of social interaction

Computational social science example

Example: How fast does fake news spread? [Vosoughi et al. 2018]

 y $=$ $f(x)$ 

spread through a network



true/fake news

network analysis

NLP/text mining

Computational social science example

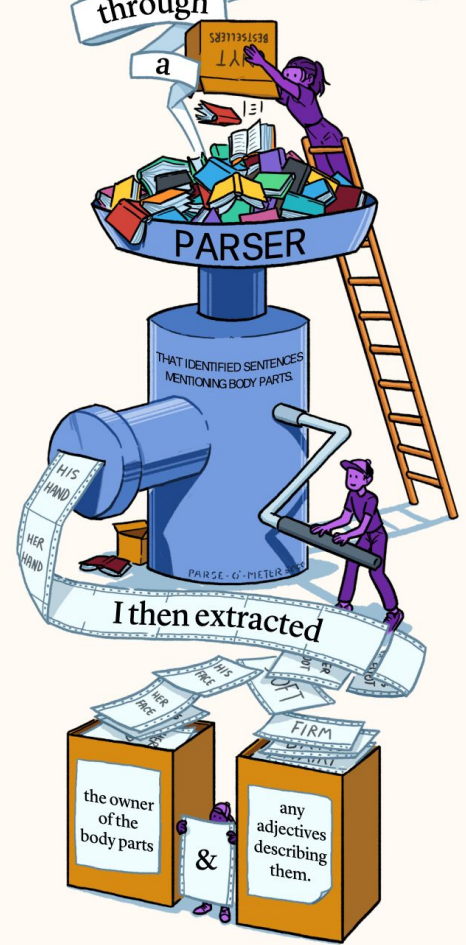
Example: Do police officers speak more respectfully to white drivers than Black drivers in traffic stops? [Voigt et al. 2017]



Digital humanities

Digital humanities

- Analyzing works from the humanities with digital methods
- Computational literary analysis
- Example: Genre prediction
 - What novels don't fit genres?
 - How have genres changed over time? [Underwood 2016]
- NLP/text mining
 - Patterns in language data



Wrapping up

- Language is embedded in social context
- Computational social science studies people and societies with computational models of observational data
 - Often uses NLP for analyzing text
- Digital humanities uses NLP methods to study literary works and other humanities artifacts

Please fill out OMETs

- CS 2731:
https://go.blueja.io/brQC_NPDBkGfnMCdj63bMg
- ISSP 2230:
https://go.blueja.io/-bGaP5Q2AUGzeQ_3tzbnqw
- All types of feedback are welcome (critical and positive)
- Completely anonymous
 - I won't be able to see them until after grades are in
- Please be specific in the feedback
- Closes this **Sun Apr 21**



See you next Wednesday

Thanks for a great semester!