Natural Language Processing

Info 159/259

Lecture 24: Social NLP (Nov. 16, 2017)

David Bamman, UC Berkeley

great


We are met on a great battle-field of that war.
λx.λy.meet(x,y)
λx.battlefield(x)
λx.war(x)

We are met on a great battle-field of that war.

Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal. Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. We are met on a great battle-field of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this. But, in a larger sense, we can not dedicate -- we can not consecrate -- we can not hallow -- this ground. The brave men, living and dead, who struggled here, have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember what we say here, but it can never forget what they did here. It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us -- that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion -- that we here highly resolve that these dead shall not have died in vain -- that this nation, under God, shall have a new birth of freedom -- and that government of the people, by the people, for the people, shall not perish from the earth.

Social NLP

- Social NLP covers a range of applications that analyze how language interacts with people in social settings.

- We leave behavioral traces in our interactions with others.

    - Tweets
    - Books
    - Emails
    - Audio transcripts

“Raw” data

- Social NLP often makes claims about the world using textual data.
- Data is not self-evident, neutral or objective
- Data is collected, stored, processed, mined, interpreted; each stage requires our participation.
- What is the process by which the data you have got to you?

Gitelman and Jackson (2013)

Twitter

TWEETS 542 | FOLLOWING 455 | FOLLOWERS 990 | LIKES 162 | LISTS 2

David Bamman
@dbamman

Assistant Professor, School of Information, UC Berkeley. Natural language processing, machine learning, computational social science, digital humanities.

Berkeley, CA

people.ischool.berkeley.edu/~dbamman/

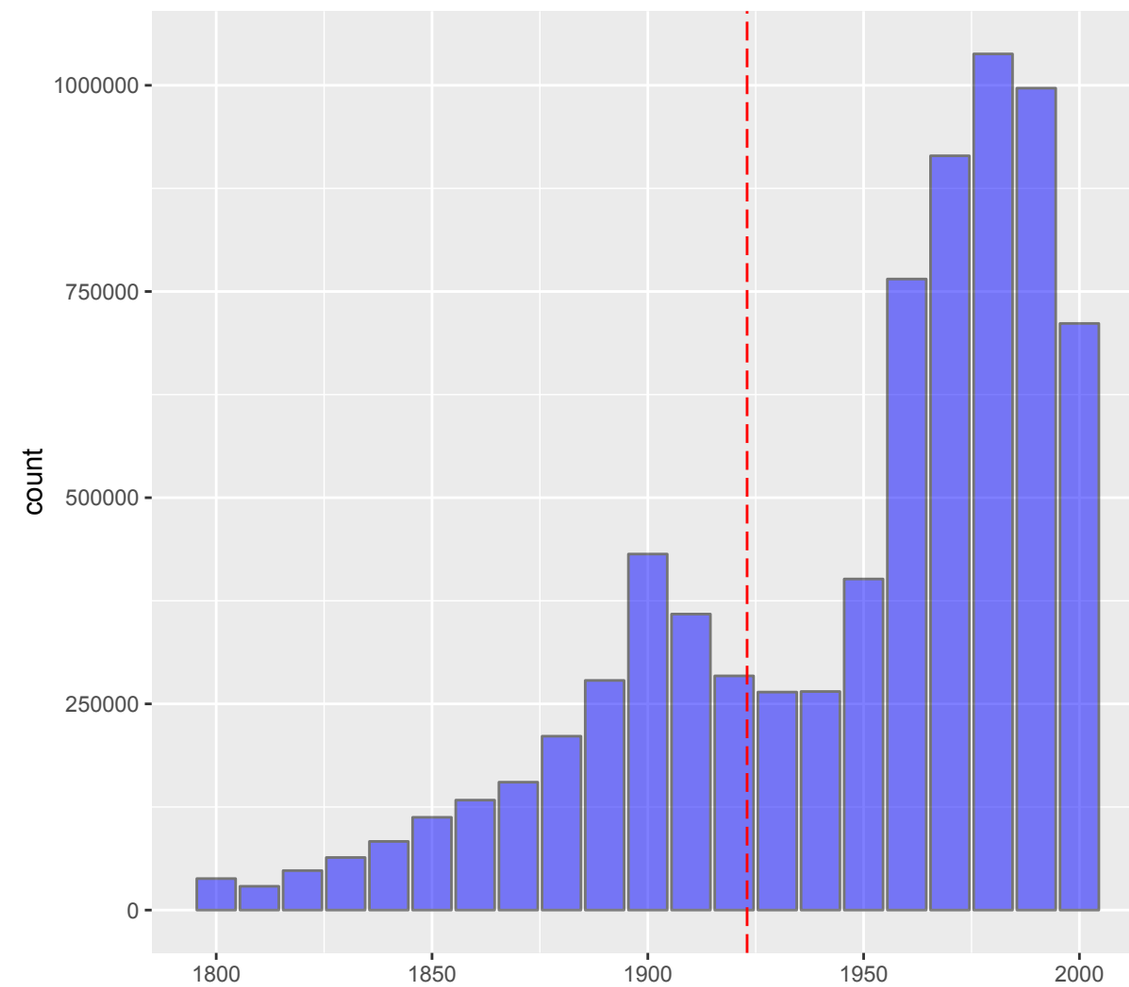
Joined October 2009

Tweets Tweets & replies Media

David Bamman @dbamman · Sep 23

Rounding out a quick NY trip for @NYUDataScience with a talk here today

Data

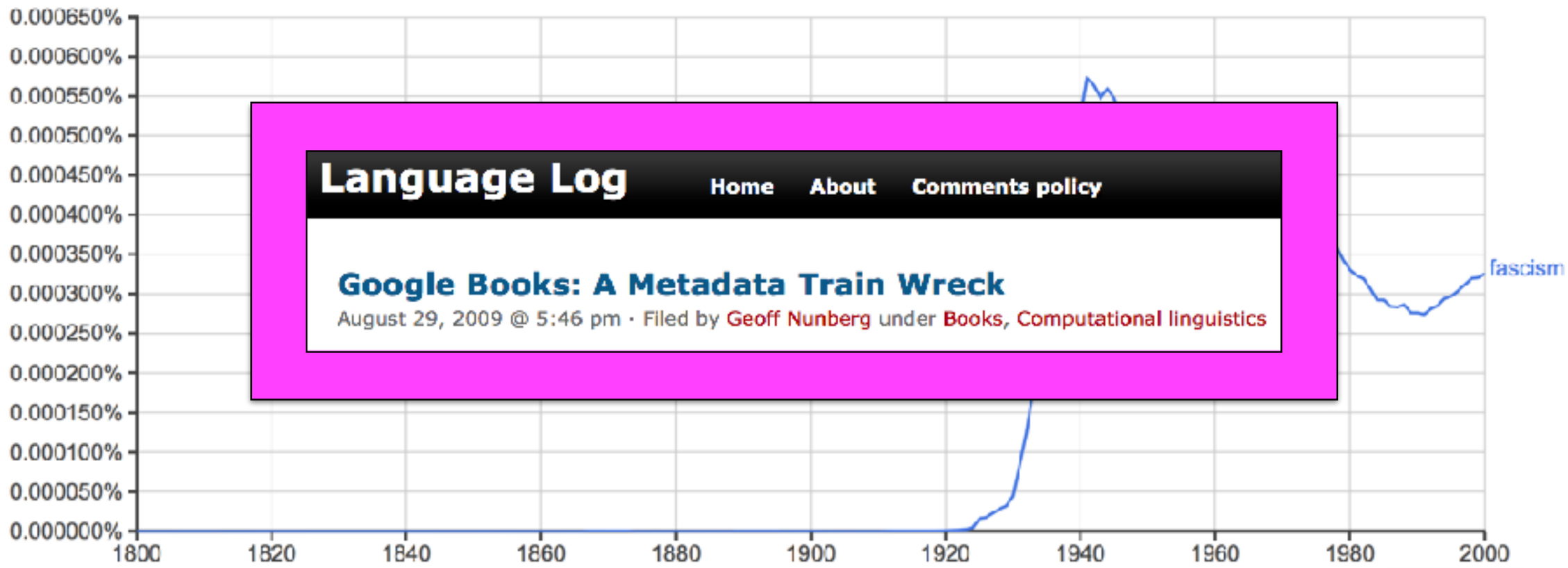


HathiTrust: 14.8M books; 5.7M in the public domain

Affordances

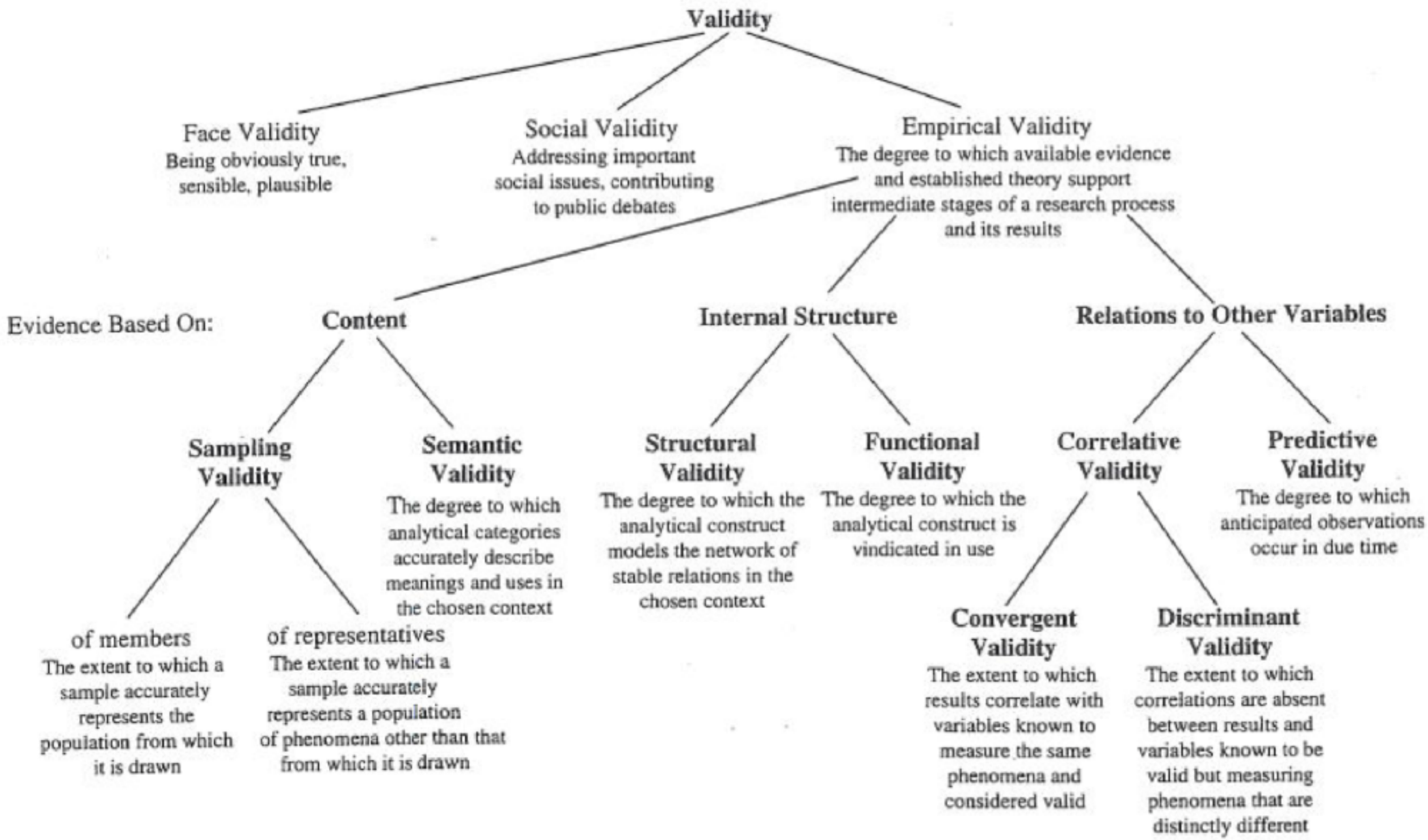
Large-scale analysis of culture

(*as recorded in published books, with university library accession policies)



Data Collection

- Data → Research Question
  - “Opportunistic data”
  - Research questions are shaped by what data you can find
- Research Question → Data
  - Research is driven by questions, find data to support answering it.



Krippendorff (2004)

Social NLP

- What are the research questions that we can ask when applying NLP to text to answer social and cultural questions?
- How do we answer those research questions using methods we've learned about?
    - data
    - algorithms
    - evaluation

Social NLP

- Manifestations of power in text
- Inferring social networks from text
- Measuring respect
- Explaining trolling behavior
- Analyzing claims about genre
- Asking for a favor

Power

How is power manifested in language?

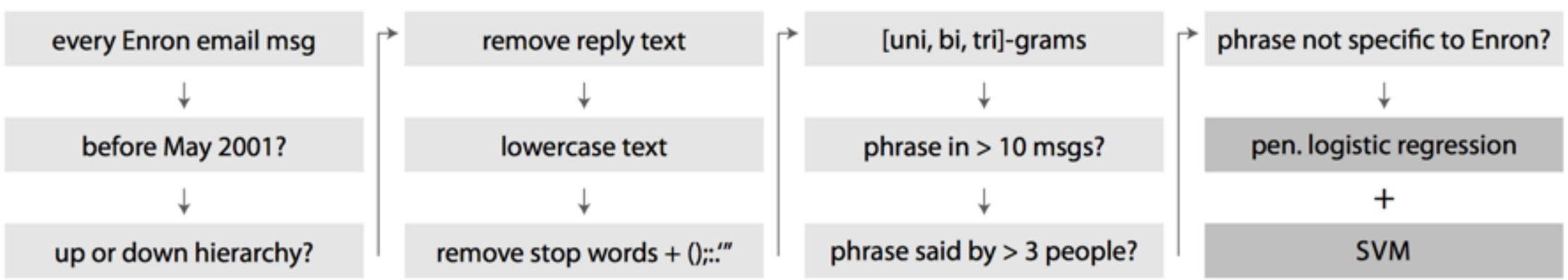
Power

- Text data: Enron emails
- Response: Enron org chart — for all pairs of entities in email (sender/recipients), who is higher on the org chart?

Gilbert 2012 (“Phrases that signal workplace hierarchy”)

Power

- Bag of words representation of text + binary classification.



Gilbert 2012 ("Phrases that signal workplace hierarchy")

these predict message going up

not going up

↑ phrases	β	↑ phrases	β	↔↓ phrases	β	↔↓ phrases	β
the ability to	6.76	attach	6.72	have you been	-8.46	to manage the	-6.66
I took	6.57	that we might	6.54	you gave	-6.64	let's discuss	-5.72
are available	6.52	the calendar	6.06	we are in	-5.44	publicly	-5.24
kitchen	5.72	can you get	5.72	title	-5.05	promotion	-5.02
thought you would	5.65	driving	5.61	need in	-4.80	good one	-4.62
, I'll be	5.51	thoughts on	5.51	opened	-4.57	determine the	-4.47
looks fine	5.50	shit	5.45	initiatives	-4.38	is difficult	-4.36
voicemail	5.43	we can talk	5.41	. I would	-4.34	man	-4.26
tremendous	5.27	it does	5.21	we will probably	-4.12	number we	-4.11
will you	5.17	involving	5.15	any comments	-4.06	contact you	-4.05
left a	5.07	the report	5.04	you said	-3.99	the problem is	-3.97
I put	4.90	please change	4.88	I left	-3.88	you did	-3.78
you ever	4.80	issues I	4.76	can you help	-3.68	cool	-3.54
I'll give	4.69	is really	4.65	send this	-3.47	your attention	-3.44
okay ,	4.60	your review	4.56	whether we	-3.44	to think	-3.44
to send it	4.48	europe	4.45	the trade	-3.40	addition to the	-3.30
communications	4.38	weekend .	4.35	and I thought	-3.28	great thanks	-3.24
a message	4.35	have our	4.33	should include	-3.19	selected	-3.16
one I	4.28	interviews	4.28	please send	-3.14	ext	-3.13
can I get	4.28	you mean	4.26	existing	-3.06	and let me	-3.05
worksheet	4.15	haven't been	4.10	mondays	-3.02	security	-3.01
liked	4.07	me . 1	4.07	presentation on	-2.95	got the	-2.94
I gave you	3.95	tiger	3.94	let's talk	-2.94	get your	-2.88
credit will	3.88	change in	3.88	the items	-2.78	this week and	-2.77
you make	3.86	item	3.84	i hope you	-2.77	team that	-2.75
together and	3.82	a decision	3.82	did it	-2.75	a deal	-2.71
have presented	3.78	a discussion	3.74	test	-2.69	yours .	-2.68
think about	3.71	sounds good	3.65	be sure	-2.65	briefing	-2.60

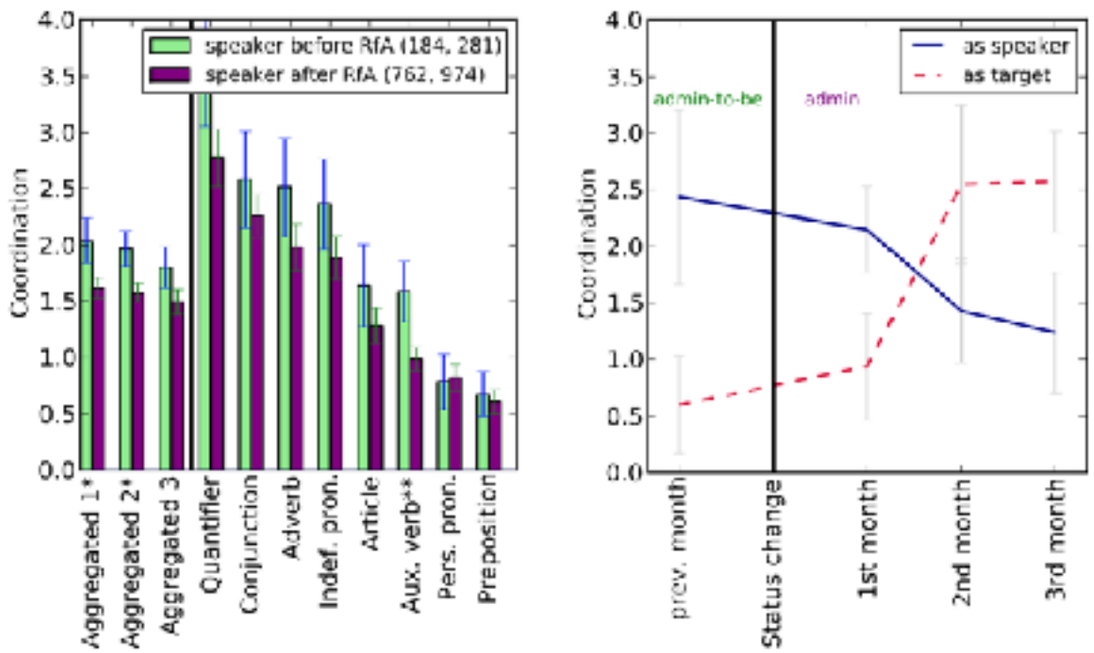
Power

- Text data: Wikipedia discussions, SCOTUS arguments
- Response: Wikipedia admins/non-admins; SCOTUS justices/laywers

Danescu-Niculescu-Mizil et al. 2012 (“Echoes of Power”)

Power

- LIWC representation of text + measurements of accommodation (adapting your speech to the language of your interlocutor)



(a) Supporting $\mathcal{P}'_{speaker}$ (b) Timed effect of **status change** ($\mathcal{P}$)

Networks

Can we learn a social network from mentions in text?

Networks

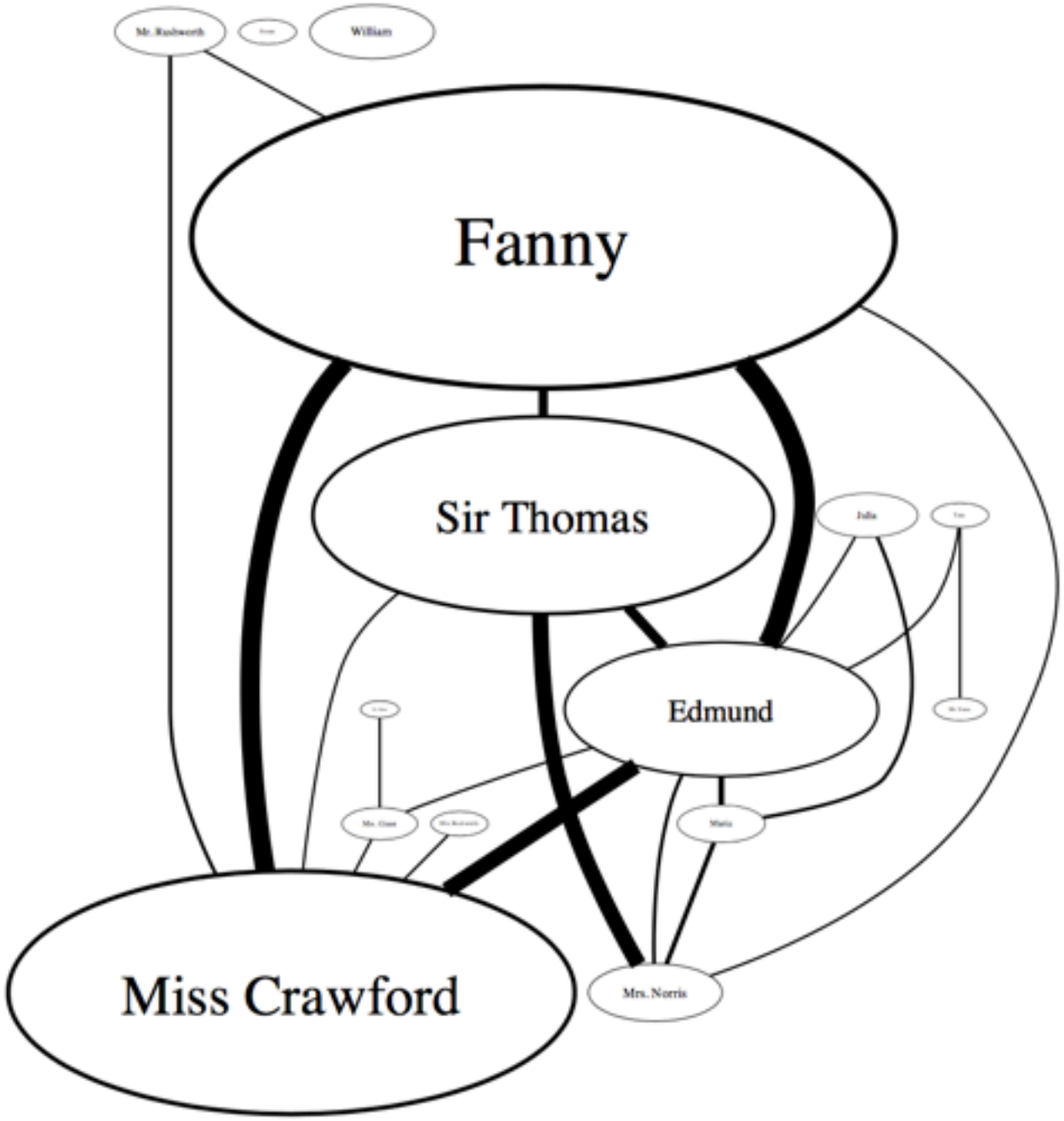
- Data: 60 novels from Project Gutenberg
- Conversational network:
  - Characters are in the same place at the same time
  - Characters take turns speaking
  - The characters are mutually aware of each other and each character's speech is mutually intended for the other to hear.

Elson et al. (2010), "Extracting Social Networks from Literary Fiction"

Networks

- Alias clustering: Tom, Tom Sawyer, Mr. Saywer = TOM SAWYER)
- Quoted speech attribution ("Yes," said TOM SAWYER)
- Network construction
  - Divide book into 10-paragraph sections, count number of sections with two characters
  - Count occurrences of one character mentioning another in dialogue

Networks



Respect

- Data: transcripts of 981 OPD traffic stops (everyday interactions)
- Response: race

Voigt et al. 2017, “Language from police body camera footage shows racial disparities in officer respect”

Respect

- Present one dialogue turn (police/driver) to be rated by people for respect (4-point Likert scale). High IAA.
- Build a predictive model mapping text to respect.

Voigt et al. 2017, “Language from police body camera footage shows racial disparities in officer respect”

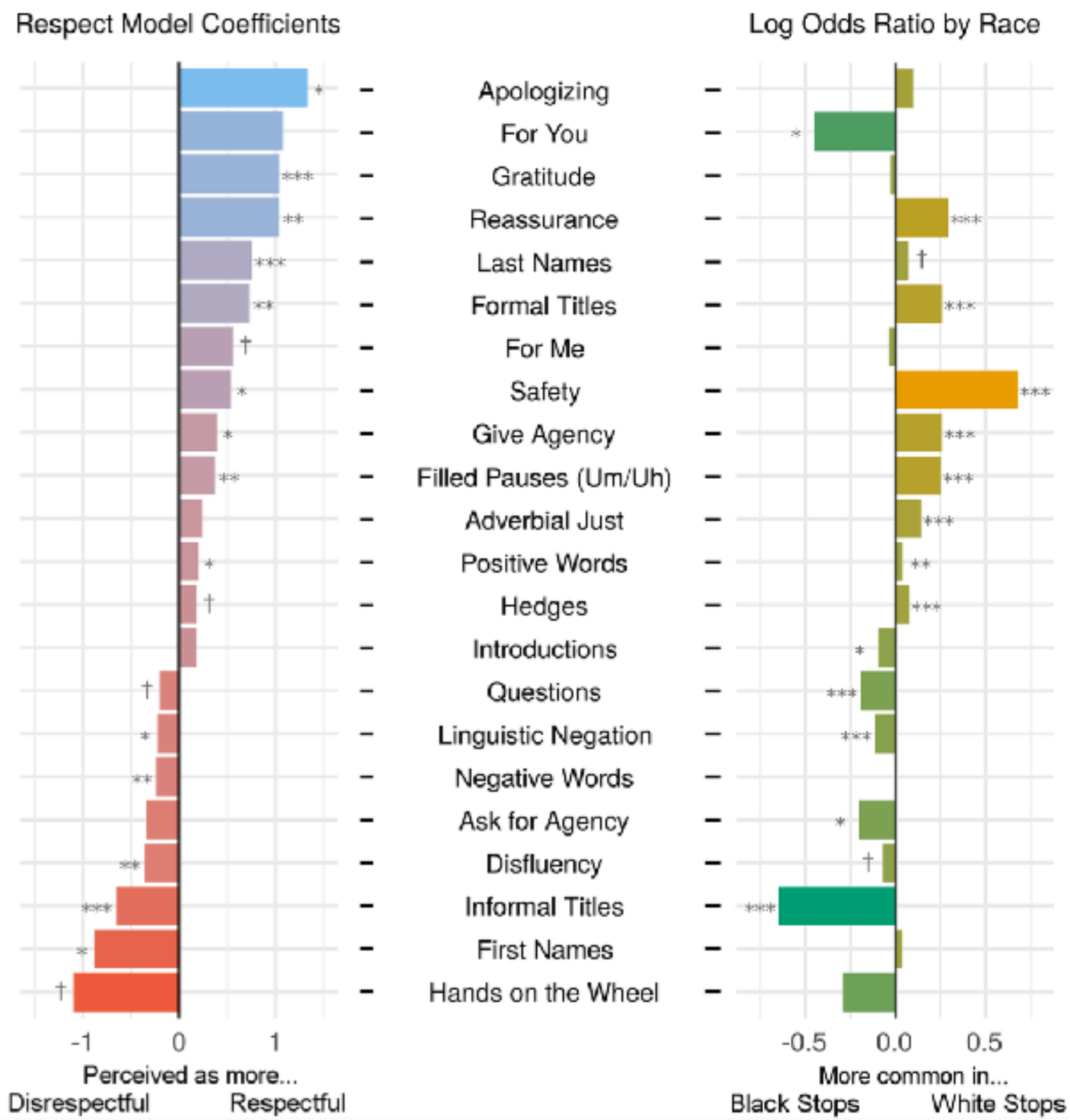
Feature Name	Implementation
Adverbial "Just"	"Just" occurs in a dependency arc as the head of an advmod relation
Apologizing	Lexicon: "sorry", "oops", "woops", "excuse me", "forgive me", "apologies", "apologize", "my bad", "my fault"
Ask for Agency	Lexicon: "do me a favor", "let me", "allow me", "can i", "should i", "may i", "might i", "could i"
Bald Command	The first word in a sentence is a bare verb with part-of-speech tag VB ("look", "give", "wait" etc.) but is not one of "be", "do", "have", "thank", "please", "hang".
Colloquialism	Regular expression capturing "y'all", "ain't" and words ending in "in'" such as "walkin'", "talkin'", etc., as marked by transcribers
Conditional	Lexicon: "if"
Disfluency	Word fragment ("Well I thi-") as indicated by transcribers
Filled Pauses	Lexicon: "um", "uh"
First Names	Top 1000 most common first names from the 1990 US Census, where first letter is capitalized in transcript
Formal Titles	Lexicon: "sir", "ma'am", "maam", "mister", "mr*", "ms*", "madam", "miss", "gentleman", "lady"
For Me	Lexicon: "for me"
For You	Lexicon: "for you"
Give Agency	Lexicon: "let you", "allow you", "you can", "you may", "you could"
Gratitude	Lexicon: "thank", "thanks", "appreciate"
Goodbye	Lexicon: "goodbye", "bye", "see you later"
Hands on the Wheel	Regular expression capturing cases like "keep your hands on the wheel" and "leave your hands where I can see them": "hands? ([+,?!:;]+)?(wheel|see)"

Voigt et al. 2017, “Language from police body camera footage shows racial disparities in officer respect”

Hedges	All words in the "Tentat" LIWC lexicon
Impersonal Pronoun	All words in the "Impron" LIWC lexicon
Informal Titles	Lexicon: "dude*", "bro*", "boss", "bud", "buddy", "champ", "man", "guy*", "guy", "brotha", "sista", "son", "sonny", "chief"
Introductions	Regular expression capturing cases like "I'm Officer |name| from the OPD" and "How's it going?": "((i|my name).+officer | officer.+(oakland|opd))|((hi|hello|hey|good afternoon|good morning|good evening|how are you doing|how 's it going))"
Last Names	Top 5000 most common last names from the 1990 US Census, where first letter is capitalized in transcript
Linguistic Negation	All words in the "Negate" LIWC lexicon
Negative Words	All words in the "Negativ" category in the Harvard General Inquierer, matching on word lemmas
Positive Words	All words in the "Positiv" category in the Harvard General Inquierer, matching on word lemmas
Please	Lexicon: "please"
Questions	Occurrence of a question mark
Reassurance	Lexicon: "'s okay", "n't worry", "no big deal", "no problem", "no worries", "'s fine", "you 're good", "is fine", "is okay"
Safety	Regular expression for all words beginning with the prefix "safe", such as "safe", "safety", "safely"
Swear Words	All words in the "Swear" LIWC lexicon
Tag Question	Regular expression capturing cases like "..., right?" and "..., don't you?": ", (((all right|right|okay|yeah|please|you know)(sir| ma'am| miss| son)?)|((are|is|do|can|have|will|won't) (n't)?(i|me|she|us|we|you|he|they|them))) [?]"
The Reason for the Stop	Lexicon: "reason", "stop* you", "pull* you", "why i", "why we", "explain", "so you understand"
Time Minimizing	Regular expression capturing cases like "in a minute" and "let's get this done quick": "(a|one|a few) (minute|min|second|sec|moment)s?|this[-,?!]+quick|right back"

Respect

Example	Respect Score
First Name, Ask For Agency, Questions [name], can I see that driver's license again? It- it's showing suspended. Is that- that's you? Disfluency, Negative Word, Disfluency	-1.07
Informal Title, Ask For Agency, Adverbial "Just" All right, my man. Do me a favor. Just keep your hands on the steering wheel real quick. "Hands On The Wheel"	-0.51
Apology, Introduction, Last Name Sorry to stop you. My name's Officer [name] with the Police Department.	0.84
Formal Title, Safety, Please There you go, ma'am. Drive safe, please.	1.21
Adverbial "Just", Filled Pause, Reassurance It just says that, uh, you've fixed it. No problem. Thank you very much, sir. Gratitude, Formal Title	2.07


Respect Model Coefficients
Log Odds Ratio by Race
Apologizing
For You
Gratitude
Reassurance
Last Names
Formal Titles
For Me
Safety
Give Agency
Filled Pauses (Um/Uh)
Adverbial Just
Positive Words
Hedges
Introductions
Questions
Linguistic Negation
Negative Words
Ask for Agency
Disfluency
Informal Titles
First Names
Hands on the Wheel
-1
0
1
Perceived as more...
Disrespectful
Respectful
-0.5
0.0
0.5
More common in...
Black Stops
White Stops

Respect

- Higher respect to white drivers, older drives, when a citation is issued.

- Lower respect when a search is conducted.

Genre

- Text: Books labeled {detective, gothic, science fiction} from bibliographies + random books
- Question: Are genres defined by shorter time periods more coherent as a category than those with longer lifetimes (e.g., detective fiction)?

Underwood 2016, “The Life Cycles of Genres”

Genre

- Bag of words representation of book
- Operationalize as a prediction task: coherence = high cross-validated accuracy.

Underwood 2016, “The Life Cycles of Genres”

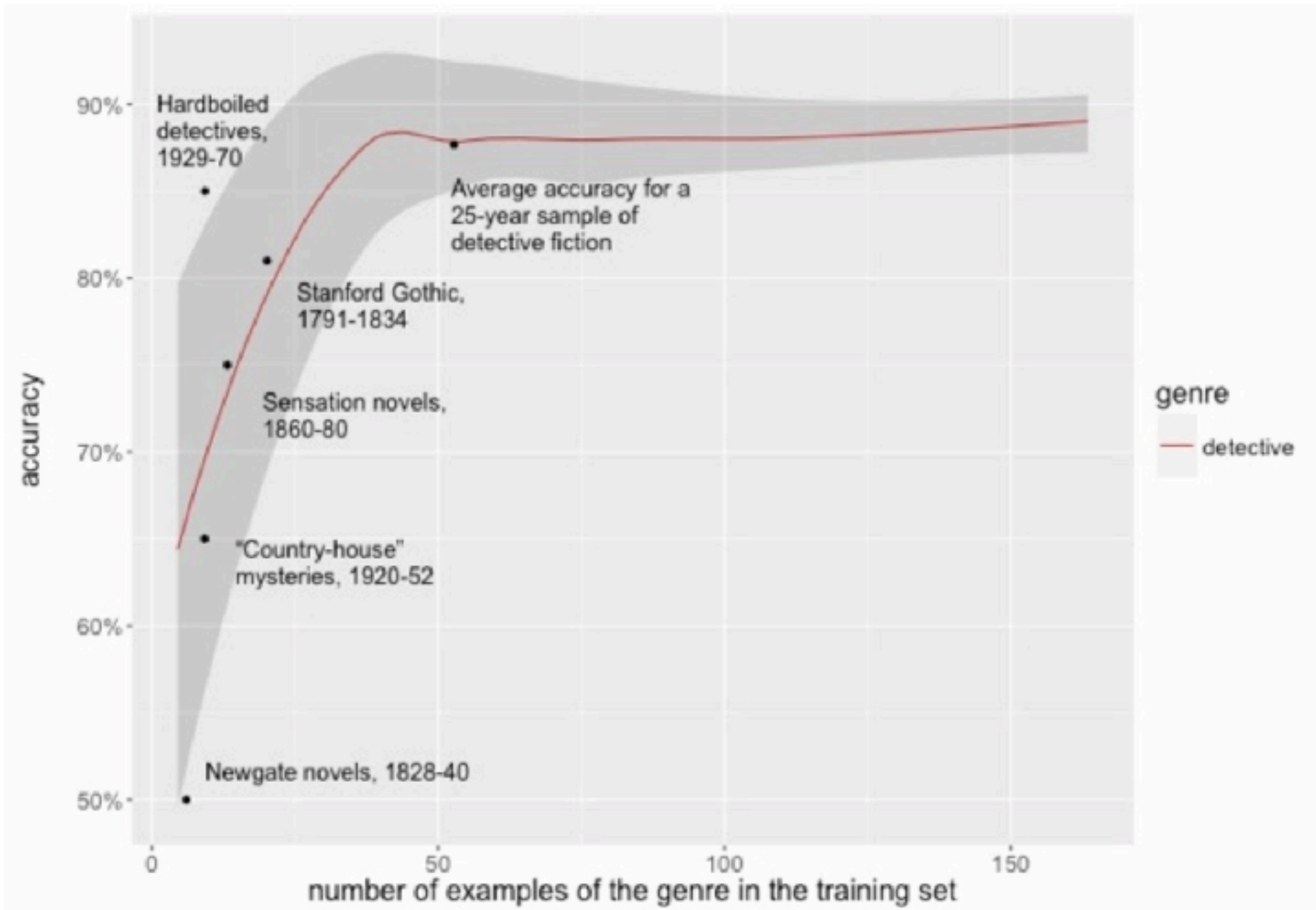


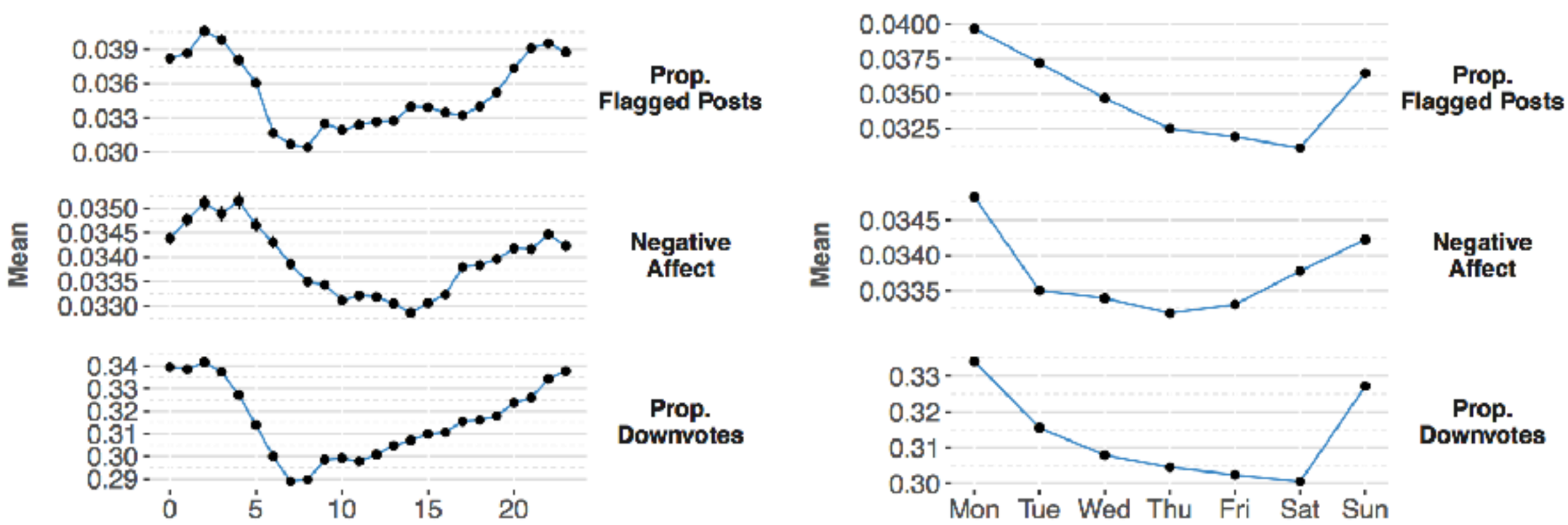
Figure 5. Several genres of roughly generational size, plotted relative to a curve that indicates the range of accuracy for a random sample of detective fiction drawn from 1829-1989. The shaded ribbon is a predictive band that covers 90% of models for a random sample of detective fiction.

Trolling

- Data: comments on CNN
- Response: comment was flagged for removal or not.
- Question: does mood or discussion context make people troll?

Cheng et al. (2017), "Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions"

Trolling



(a) Time of Day (EDT)

(b) Day of Week

Cheng et al. (2017), "Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions"

- Data: Random acts of pizza (subreddit)
- Response: Is a request successful in getting a pizza?

Althoff et al. (2014), "How to Ask for a Favor: A Case Study on the Success of Altruistic Requests

Favors

- Representation of text:
    - Temporal features of post/author
    - Politeness
    - Evidentiality (image link in text)
    - Reciprocity (“pay it forward”, “return the favor”)
    - Sentiment
    - Length
    - Status (reddit karma points)
    - “Money”, “Job” “Student”, “Family”, “Craving” topics

Coefficient	Estimate	SE
Intercept	-2.02^{***}	0.14
Community Age (Decile)	-0.13^{***}	0.01
First Half of Month (Binary)	0.22^{**}	0.08
Gratitude (Binary)	0.27^{**}	0.08
Including Image (Binary)	0.81^{***}	0.17
Reciprocity (Binary)	0.32^{**}	0.10
Strong Positive Sentiment (Binary)	0.14	0.08
Strong Negative Sentiment (Binary)	-0.07	0.08
Length (in 100 Words)	0.30^{***}	0.05
Karma (Decile)	0.13^{***}	0.02
Posted in RAOP before (Binary)	1.34^{***}	0.16
Narrative Craving (Binary)	-0.34^{***}	0.09
Narrative Family (Binary)	0.22^{*}	0.09
Narrative Job (Binary)	0.26^{**}	0.09
Narrative Money (Binary)	0.19^{**}	0.08
Narrative Student (Binary)	0.09	0.09

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Features	ROC AUC
Random Baseline	0.500
Unigram Baseline	0.621^{***}
Bigram Baseline	0.618^{***}
Trigram Baseline	0.618^{***}
Text Features	0.625^{***}
Social Features	0.576^{***}
Temporal Features	0.579^{***}
Temporal + Social	0.638^{***}
Temporal + Social + Text	$\mathbf{0.669}^{***}$
Temporal + Social + Text + Unigram	$\mathbf{0.672}^{***}$

Social NLP

- Many different methods from NLP
    - Text-based classification, regression, sequence labeling (e.g., NER)
- Representation is important:
    - Bag of words
    - Features derived from parts of speech, syntax
- When testing theories, interpretable models are important.

Social NLP

Text also provides a lens into exploratory analysis of social and cultural phenomena