Just in Time

Three Case Studies in Diachronic Text Analysis

Jacob Eisenstein

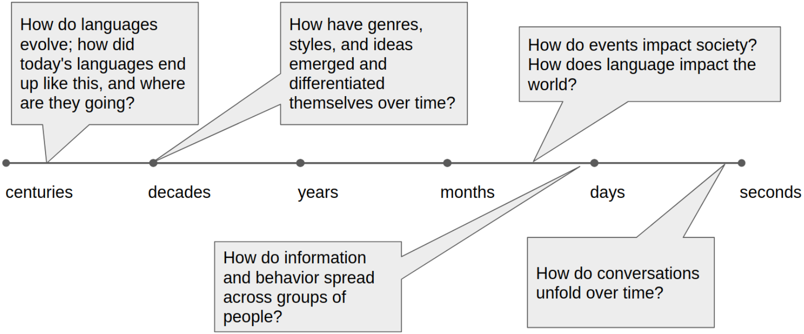
Diachronic text and computational social science

Text-over-time is a powerful tool for computational social science and digital humanities:

- **Text** allows us to operationalize a wide range of constructs using primary sources. Language and written artifacts are also objects of study in their own right.
- **Time** links text to real-world events, and can shed light on causation, influence, and diffusion.

The increasing availability of large-scale diachronic corpora offers a range of exciting new opportunities.

Some questions about language change



Sources of information about language change

- **Lexicons**: lists of word types or cognate sets
- **Simulation**: simple models and their ability to explain observed phenomena
- **Apparent time**: differences between individuals by age
- **Corpora**: time-stamped ("diachronic") texts

Sources of information about language change

- **Lexicons**: lists of word types or cognate sets
- **Simulation**: simple models and their ability to explain observed phenomena
- **Apparent time**: differences between individuals by age
- **Corpora**: time-stamped ("diachronic") texts

	Known Modern Languages			
Gloss[†]	Fijian	Pazeh	Melanau	Inabaknon
star	kalokalo[§]	mintol	biten	bitu'on
to hold	taura	maːraʔ	magem	kumkom
house	vale	xumaʔ	lebuʔ	ruma
bird	manumanu	aiam	manuk	manok
to cut, hack	tata	taːtatak	tutek	hadhad
at	e	-[¶]	gaʔ	-
what?	cava	ʔaxai	uaʔ inew	ay
this	oqo	ʔimini	itew	yayto
wind	cagi	varə	paŋay	bariyo

Figure: Bouchard-Côté et al. 2013

Sources of information about language change

- **Lexicons**: lists of word types or cognate sets
- **Simulation**: simple models and their ability to explain observed phenomena
- **Apparent time**: differences between individuals by age
- **Corpora**: time-stamped ("diachronic") texts

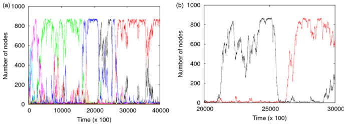


Figure: Fagyal et al. 2010

Sources of information about language change

- **Lexicons**: lists of word types or cognate sets
- **Simulation**: simple models and their ability to explain observed phenomena
- **Apparent time**: differences between individuals by age
- **Corpora**: time-stamped ("diachronic") texts

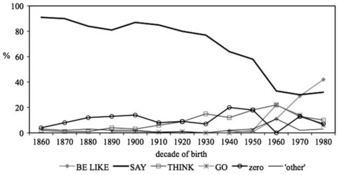


Figure: D'Arcy 2012

Sources of information about language change

- **Lexicons**: lists of word types or cognate sets
- **Simulation**: simple models and their ability to explain observed phenomena
- **Apparent time**: differences between individuals by age
- **Corpora**: time-stamped ("diachronic") texts

Corpus linguistics: the dinosaur analogy[1]

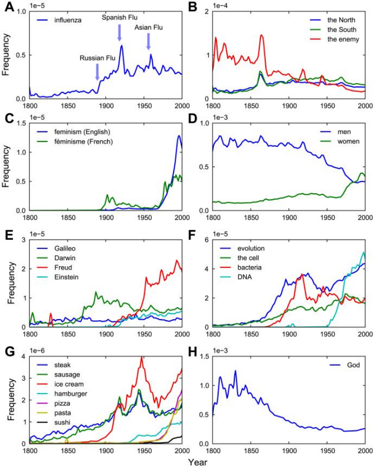
Culturnomic results are a new type of evidence in the humanities. As with fossils of ancient creatures, the challenge of culturnomics lies in the interpretation of this evidence... These, together with the billions of other trajectories that accompany them, will furnish a great cache of bones from which to reconstruct the skeleton of a new science.

[1]Michel et al. 2011.

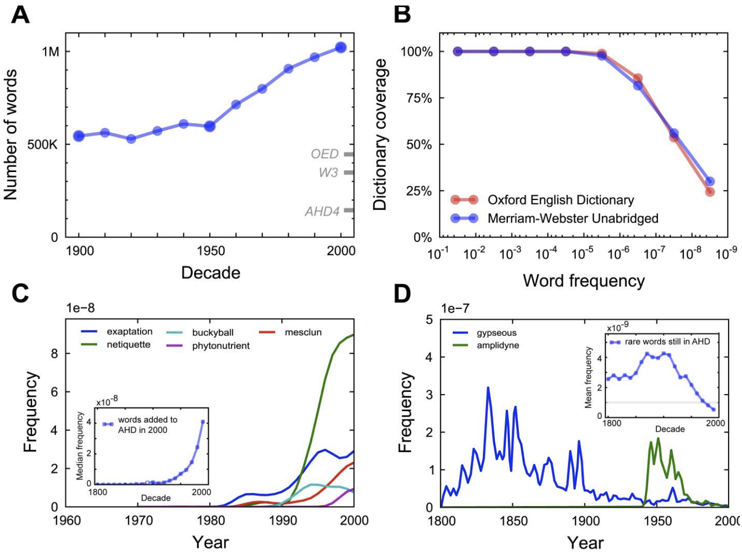
Big data and word frequency

Culturnomics: Quantitative analysis of culture using millions of digitized books (Michel et al. 2011)

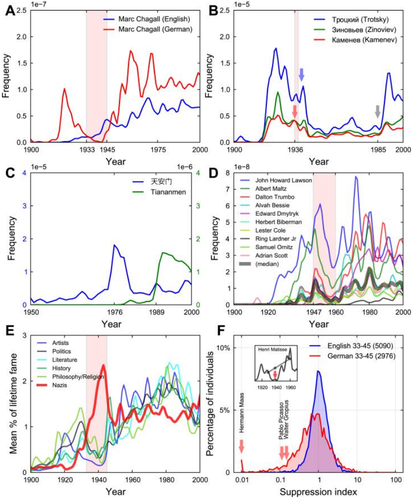
- 5M digitized books: "4% of all books ever published"
- 500B word tokens, 361B in English



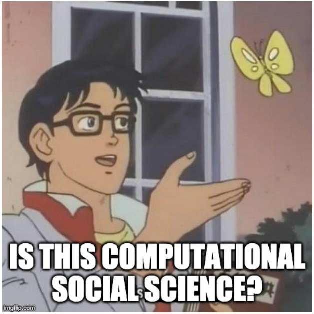
Culturnomical lexicography



Culturnomical political science



Some questions



Seems pretty easy! But research on frequency trajectories invites a number of questions:

- Are word frequencies independent random variables?
- What is the correspondence between words and meanings?
- When is temporal association likely to indicate causation?

Outline

- Motivations and perspectives on diachronic text
- Case studies
    1. Lexical innovations in linguistic context
    2. Semantic progressiveness and citation influence
    3. Cause and effect in online hate speech
- Concluding thoughts on CSS

Outline

- Motivations and perspectives on diachronic text
- Case studies
  1. **Lexical innovations in linguistic context**
  2. Semantic progressiveness and citation influence
  3. Cause and effect in online hate speech
- Concluding thoughts on CSS

Making fetch happen

The influence of social and linguistic context on the rise and fall of lexical innovations[2]

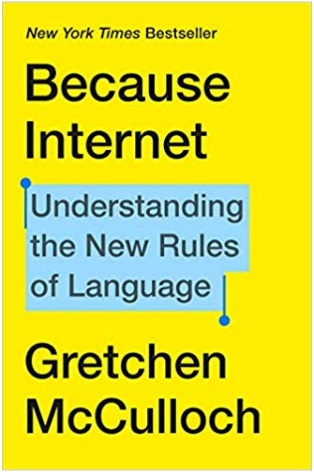
Stop trying to make "fetch" happen! It's not going to happen!

Regina George, Mean Girls (2005)

[2] Ian Stewart and Jacob Eisenstein (2018). "Making "fetch" happen: The influence of social and linguistic context on the success of lexical innovations". In: *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*.

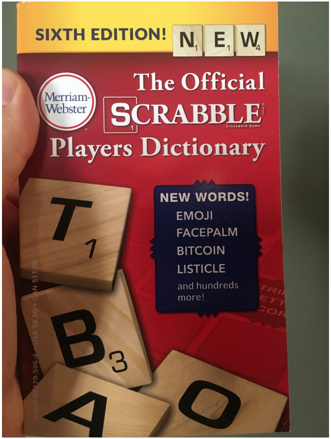
Background

- In any living language, new words constantly enter and exit the lexicon[3]
- It's not clear whether computer-mediated communication accelerates this process, but it definitely makes it more visible.



[3] Pierrehumbert 2010.

Some recent changes



Rates of change

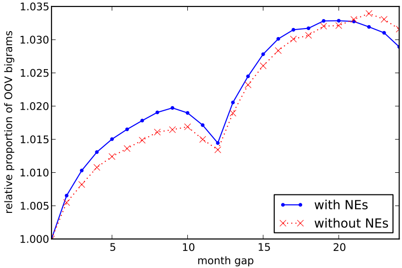
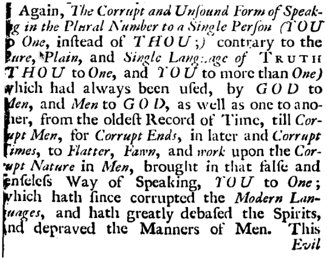


Figure: Out-of-vocabulary bigrams on Twitter, in comparison with reference lexicon from first month[5]

[5]Eisenstein 2013.

Strong feelings about lexical change

The Corrupt and Unfound Form of Speaking in the Plural Number to a Single Person (YOU to One, instead of THOU;) contrary to the Pure, Plain and Single Language of TRUTH *THOU to One, and YOU to more than One) which had always been used, by GOD to Men, and Men to GOD, as well as one to another, from the oldest Record of Time, till Corrupt Men, for Corrupt Ends, in later and Corrupt Times, to Flatter, Fawn, and work upon the Corrupt Nature in Men, brought in that false and senseless Way of Speaking, YOU to One ...*


Again, *The Corrupt and Unfound Form of Speak-*
g in the Plural Number to a Single Person (YOU
o One, instead of *THOU*;) contrary to the
ure, *Plain*, and *Single Lang..age of* TRUTH
THOU to *One*, and *YOU* to more than *One)*
which had always been ufed, by *GOD* to
Men, and *Men* to *GOD*, as well as one to ano-
her, from the oldeft Record of Time, till *Cor-*
upt Men, for *Corrupt Ends*, in later and *Corrupt*
Times, to *Flatter*, *Fawn*, and *work* upon the *Cor-*
upt Nature in *Men*, brought in that falfe and
enfelefs Way of Speaking, *YOU* to *One*;
which hath fince corrupted the *Modern Lan-*
uages, and hath greatly debafed the Spirits,
nd depraved the Manners of Men. This
Evil

https://languagelog.ldc.upenn.edu/nll/?p=26554

Picking winners and losers

What factors predict whether an innovative slang term will succeed or fail?

- Prior work has focused largely on **social factors**: who are the early adopters, how is their social network organized, and how influential are they?[6]
- This work considers **linguistic factors**: how does the innovation fit into the existing linguistic system?

[6] Altmann, Pierrehumbert, and Motter 2011; Garley and Hockenmaier 2012.

Social dissemination

Altmann, Pierrehumbert, and Motter (2011): successful innovations disseminate widely across social contexts.

- For example, it is better to have three adopters in three cities than in one city.

---

[7] Garley and Hockenmaier 2012.

[8] Altmann, Pierrehumbert, and Motter 2011.

Social dissemination

Altmann, Pierrehumbert, and Motter (2011): successful innovations disseminate widely across social contexts.

- For example, it is better to have three adopters in three cities than in one city.

Quantifying dissemination:

$$D = \log \frac{\text{count-of-contexts}}{E[\text{count-of-contexts} \mid \text{total-counts}]} \tag{1}$$

- one context = one user[7]
- one context = one newsgroup[8]

[7] Garley and Hockenmaier 2012.

[8] Altmann, Pierrehumbert, and Motter 2011.

Linguistic dissemination

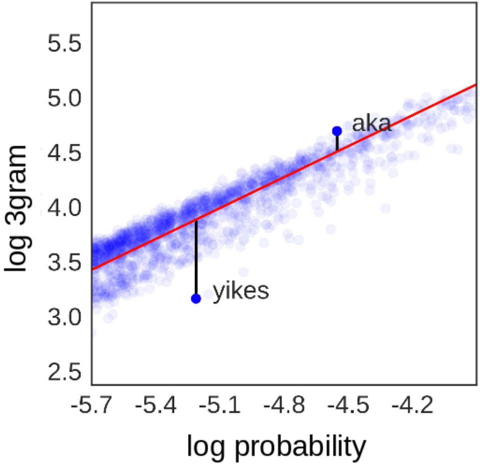
- This and other prior work treats language no differently from hashtags[9] or hyperlinks.[10] But language is different, because innovations must interact with the rest of the linguistic system.
- Our hypothesis is that linguistically versatile innovations tend to succeed. We define **linguistic dissemination**: one context = one trigram.

$$D^{(\ell)} = \log \frac{\text{count-of-trigrams}}{E[\text{count-of-trigrams} \mid \text{total-counts}]} \tag{2}$$

[9] Romero, Meeder, and Kleinberg 2011.

[10] Bakshy et al. 2012.

Linguistic dissemination



Data

- 1.6B public Reddit posts and comments from 2013-2016
  - Filtered known bots and spammers[11]
  - English-language subreddits only
- Vocabulary methodology: automatically search, manually filter.[12]
  1. Automatically identify words with a period of consistent growth.
  2. Manually filter out proper nouns and standard English ($\kappa = .79$).

[12] Tan and Lee 2015.

[12] Eisenstein, O'Connor, et al. 2014.

Analyses

1. Does (linguistic/social) dissemination **cause** word frequency to increase?
2. Can dissemination help to **predict**
    - which words will increase in frequency?
    - how long each innovation will survive?

Causal analysis

Potential outcomes perspective: "if this individual had/hadn't been treated, what would have been the outcome?"

- **Treatment**: amount of dissemination;
- **Outcome**: whether word increases in frequency after 12 months;
- **Covariates**: everything else we know about each word.

Propensity score matching is a well-known approach to this problem,[13] but extra care is required when the treatment is continuous.

[13] Rosenbaum and Rubin 1983.

Average dose-response function[14]

1. Fit a model of the treatment from the covariates,

$$Z_i \mid X_i \sim N(\beta \cdot x_i, \sigma_Z^2). \tag{3}$$

The generalized propensity score R_i is the conditional likelihood $P(z_i \mid x_i)$.

2. Regress the outcome against the treatment and the generalized propensity score,

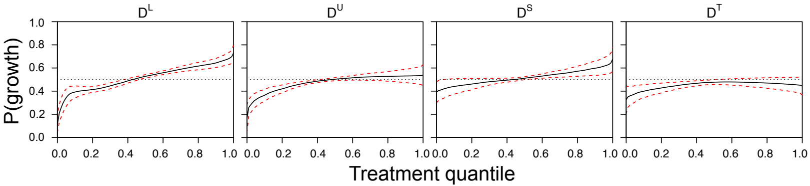
$$\hat{Y}_i = \sigma(\hat{\alpha}_0 + \hat{\alpha}_1 Z_i + \hat{\alpha}_2 R_i). \tag{4}$$

3. At each treatment quantile, s_z, compute the average predicted outcome for each instance,

$$\hat{\mu}(s_z) = \frac{1}{|s_z|} \sum_{i:z_i \in s_z} \hat{Y}_i. \tag{5}$$

[14] Hirano and Imbens 2004.

Average dose-response results



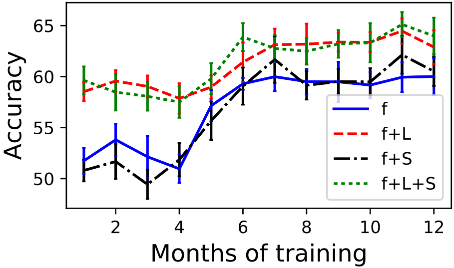
- Linguistic dissemination (D^L) steadily increases the probability that an innovation will be adopted (left).
- Of the three social dissemination indicators, only subreddit dissemination (D^S) makes a significant impact on adoption.

Predicting word success

Given t months of training data, can we predict whether a word will continue to increase in frequency?

Predicting word success

Given t months of training data, can we predict whether a word will continue to increase in frequency?



- f: frequency
- L: linguistic dissemination
- S: social dissemination

Predicting word survival

Can we predict when innovations will start to lose popularity?

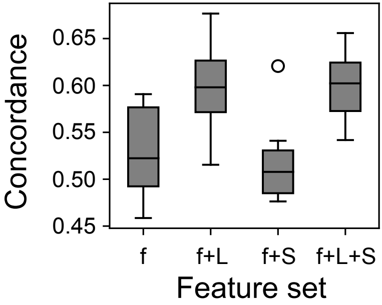
- Cox proportional hazards model,

$$\lambda_i(t) = \lambda_0(t) \exp(\beta \cdot \mathbf{x}_i), \tag{6}$$

  where
  - $\lambda_i(t)$ is the hazard of "death" at time t;
  - $\mathbf{x}_i$ is a vector of predictors;
  - β is a vector of weights.
- Must adjust for right-censored data, since not all innovations decline during our sample.

Predicting word survival

- Of all the dissemination statistics, only linguistic dissemination is a statistically significant predictor of survival.
- Including linguistic dissemination significantly increases predictive accuracy (as measured by concordance).



Summary of this part

- Successful innovations disseminate into a diverse set of phrases, rather than a few popular fixed expressions. Innovations that do not are unlikely to succeed.

- Linguistic innovations can help to measure social phenomena, but they are different from other types of innovations, like hashtags and hyperlinks.[15]

  Open question: we've argued that systematicity shapes lexical change. How should this perspective be incorporated into models of diffusion and influence?

[15] Rotabi, Danescu-Niculescu-Mizil, and Kleinberg 2017.

Outline

- Motivations and perspectives on diachronic text
- Case studies
    1. **Lexical innovations in linguistic context**
    2. Semantic progressiveness and citation influence
    3. Cause and effect in online hate speech
- Concluding thoughts on CSS

Outline

- Motivations and perspectives on diachronic text
- Case studies
  1. Lexical innovations in linguistic context
  2. **Semantic progressiveness and citation influence**
  3. Cause and effect in online hate speech
- Concluding thoughts on CSS

Follow the leader

Documents on the Leading Edge of Semantic Change Get More Citations[16]

[16] Sandeep Soni, Kristina Lerman, and Jacob Eisenstein (2019). "Follow the Leader: Documents on the Leading Edge of Semantic Change Get More Citations". In: *arXiv preprint arXiv:1909.04189*.

Follow the leader?

- Languages change by assigning new meanings to existing signs.[17]
- Recent work on **diachronic word embeddings** can capture such changes.[18]
- Can we identify **documents** that lead semantic changes? Are these documents especially influential?

[17] Traugott and Dasher 2001.

[18] Kulkarni et al. 2015; W. Hamilton, Leskovec, and Jurafsky 2016; Rosenfeld and Erk 2018.

Diachronic word embeddings

- Word embeddings are vector representations of word meaning, estimated from large unlabeled corpora. $\mathbf{v}_i \approx \mathbf{v}_j$ implies i and j are semantically similar.[19]
- Which words changed their meanings? Naive approach is to compare embeddings over time, $||\mathbf{v}_i^{t+1} - \mathbf{v}_i^t||$.

  This doesn't work because words are not distributed evenly through embedding space.

---

[19] Mikolov et al. 2013.

[20] W. L. Hamilton, Leskovec, and Jurafsky 2016.

Diachronic word embeddings

- Word embeddings are vector representations of word meaning, estimated from large unlabeled corpora. $\mathbf{v}_i \approx \mathbf{v}_j$ implies i and j are semantically similar.[19]
- Which words changed their meanings? Naive approach is to compare embeddings over time, $||\mathbf{v}_i^{t+1} - \mathbf{v}_i^{t}||$.

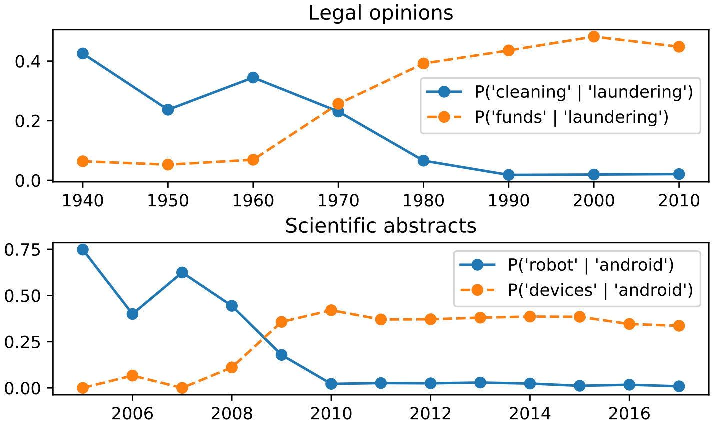
  This doesn't work because words are not distributed evenly through embedding space.

- Neighbor-based approach:[20]
  1. Let $\mathcal{N}_w^{(t)}$ be the near-neighbors of word w at time t.
  2. A word undergoes semantic change when $|\mathcal{N}_w^{(t)} \cap \mathcal{N}_w^{(t+1)}|$ is small.

[19] Mikolov et al. 2013.

[20] W. L. Hamilton, Leskovec, and Jurafsky 2016.

Examples


Legal opinions
0.4
0.2
0.0
1940
1950
1960
1970
1980
1990
2000
2010
P('cleaning' | 'laundering')
P('funds' | 'laundering')
Scientific abstracts
0.75
0.50
0.25
0.00
2006
2008
2010
2012
2014
2016
P('robot' | 'android')
P('devices' | 'android')

Identifying progressive usages

- Is a given usage the "old" or "new" meaning?
- The skipgram word embedding model computes the probability of the context around each word,

$$\log P(w_{i+k} \mid w_i) = \mathbf{v}_{w_{i+k}} \cdot \mathbf{u}_{w_i} - \log \sum_{w'} \exp \mathbf{v}_{w'} \cdot \mathbf{u}_{w_i}. \quad (7)$$

- The "progressiveness" of a usage is the log-odds ratio,

$$r_i \triangleq \sum_k \log \frac{P^{(\text{new})}(w_{i+k} \mid w_i)}{P^{(\text{old})}(w_{i+k} \mid w_i)}. \quad (8)$$

The progressiveness of a document with respect to word i is the sum of this statistic across usages.

Examples

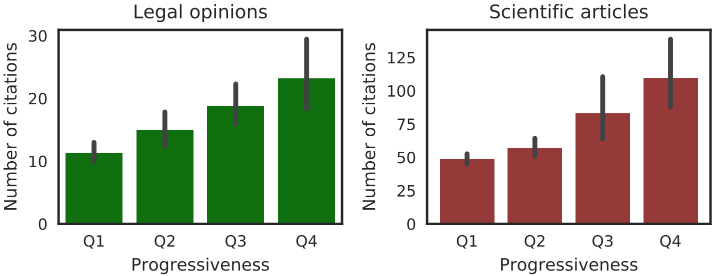
Corpus	Innovation	Leading document
Legal	laundering	United States v. Talmadge G. Rauhoff (7th Cir. 1975)
	asylum	Bertrand v. Sava (S.D.N.Y. 1982)
	fertilization	Planned Parenthood vs Casey (505 U.S. 833)
Science	ux	Hassenzahl and Tractinsky (2006)
	surf	Bay et al (2008)
	android	Shabtai et al (2010)

Examples

Corpus	Innovation	Leading document
Legal	laundering	United States v. Talmadge G. Rauhoff (7th Cir. 1975)
	asylum	Bertrand v. Sava (S.D.N.Y. 1982)
	fertilization	Planned Parenthood vs Casey (505 U.S. 833)
Science	ux	Hassenzahl and Tractinsky (2006)
	surf	Bay et al (2008)
	android	Shabtai et al (2010)

- ...two-week gestational increments from **fertilization** to full term ...
- ...$15,000 as part of the **'laundering'** process ...
- ...first step in the successful **laundering** of the funds...

Do semantic leaders get more citations?



Multivariate Poisson regression

- **Structural controls**: outgoing citations; number of authors; age; length
- **Text based controls**:
  - expected citations from bag-of-words regression[21]
  - number of semantically innovative terms used

[21] Yogatama et al. 2011.

Multivariate Poisson regression

- **Structural controls**: outgoing citations; number of authors; age; length
- **Text based controls**:
  - expected citations from bag-of-words regression[21]
  - number of semantically innovative terms used
- **Results**: the top quartile (Q4) of semantically innovative documents gets more citations.
  - Scientific articles: 2x as many citations as Q1.
  - Legal opinions: 60% more citations than Q1.

[21] Yogatama et al. 2011.

Summary

- It's not just what you say, it's what you mean.
- What causes what?
  - Addressing a novel and high-profile topic (e.g., money laundering) causes a document to be cited.
  - Highly-cited documents are likely to have their terminology adopted (e.g., LDA)
  - Being an early adopter of an innovative meaning is symptomatic of conceptual innovation?

Outline

- Motivations and perspectives on diachronic text
- Case studies
  1. Lexical innovations in linguistic context
  2. **Semantic progressiveness and citation influence**
  3. Cause and effect in online hate speech
- Concluding thoughts on CSS

Outline

- Motivations and perspectives on diachronic text
- Case studies
  1. Lexical innovations in linguistic context
  2. Semantic progressiveness and citation influence
  3. **Cause and effect in online hate speech**
- Concluding thoughts on CSS

You can't stay here

The Effectiveness of Reddit's 2015 Ban Through the Lens of Hate Speech[22]

[22] Eshwar Chandrasekharan et al. (2018). "You Can't Stay Here: The Effectiveness of Reddit's 2015 Ban Through the Lens of Hate Speech". In: *Proceedings of Computer-Supported Cooperative Work (CSCW)*.

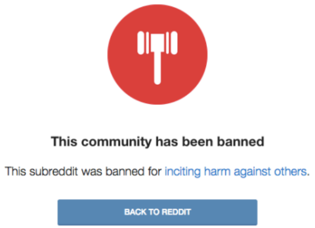
Hate speech on Reddit

What happens when forums for hate speech are shut down?

- Do participants export hate speech elsewhere?
- Or does the elimination of the "echo chamber" reduce hate speech overall?

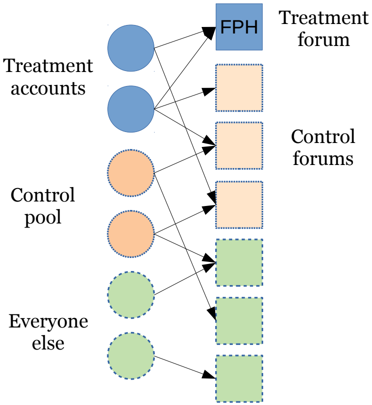
A natural experiment

- In 2015, Reddit closed several forums for violations of its anti-harassment policy.
- This enables a **natural experiment** on the effectiveness of this intervention.



Causal inference design

- **Treatment group**: user accounts that post in the forums that were banned
- **Control forums**: other forums where the treatment group posts
- **Control pool**: other accounts who post in the control forums
- **Control group**: user accounts selected by Mahalanobis Distance Matching in the control pool



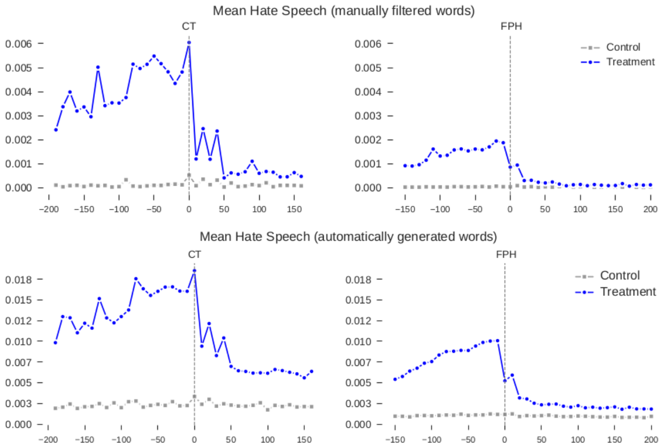
Measuring hate speech

1. Identify words that are unusually frequent in each forum, using SAGE.[23].
2. Examine the top 100, manually remove words that are not intrinsically linked to hate speech (EU Court of Human Rights definition)
   - the forum itself: fph, ct
   - the act of posting offensive content: shitposting, shitlord
   - words often used in non-hate speech contexts: IQ, welfare, cellulite

   High interrater agreement, $\kappa \approx .88$

[23] Eisenstein, Ahmed, and Xing 2011.

Causal effect on hate speech



Aftermath


47.0k
Reddit's bans of r/coontown and r/fatpeoplehate worked--many accounts of frequent posters on those subs were abandoned, and those who stayed reduced their use of hate speech
comp.social.gatech.edu
5 months ago by asbruckman
Professor | Interactive Computing
x2
6649 comments share save hide report

Aftermath

[–] **Hey-Grandan2** 349 points 5 days ago

What excacly qualifies for hate speech?

permalink embed save report give gold reply

[–] **eegilbert** Author of Article 652 points 5 days ago

One of the authors here. There was an unsupervised computational process used, documented on pages 6 and 7, and then a supervised human annotation step. Both lexicons are used throughout the rest of work.

permalink embed save parent report give gold reply

[+] *Comment removed 5 days ago* (58 children)*

[–] **Laminar_flo** 92 points 5 days ago

Ok, adding to that, how did you ensure that the manual filtering process was ideological neutral and not just a reflection of the political sensitivities of the person filtering?

permalink save parent report give gold reply

[–] **qwenjwenfljnanq** 11 points 5 days ago

But then how did you differentiate between hate speech and people talking *about* hate speech?

permalink save parent report give gold reply

[–] **Mode1961** -14 points 5 days ago

> number of words that indicate hate speech

Who choose those words.

permalink save parent report give gold reply

Aftermath

U.S.

Reddit Bans Nazi Groups and Others in Crackdown on Violent Content

By **CHRISTINE HAUSER** OCT. 26, 2017

Steve Huffman, a co-founder and chief executive of Reddit, in 2016. The company has started to implement a new policy to remove content that glorifies and incites violence from its site. David Paul Morris/Bloomberg

RELATED COVERAGE

ON TECHNOLOGY
How Hate Groups Forced Online Platforms to Reveal Their True Nature AUG. 21, 2017

Opinion | Op-Ed Contributor
My Time Undercover With the Alt-Right
SEPT. 27, 2017

THE SHIFT
This Was the Alt-Right's Favorite Chat App. Then Came Charlottesville. AUG. 15, 2017

THE SHIFT
Reddit Limits Noxious Content by Giving Trolls Fewer Places to Gather SEPT. 25, 2017

(Why) did it work for Reddit?

- Reddit's federated structure delegates norm enforcement to moderators.

  It would be hard for Facebook and Twitter to target hate speech *communities* in the same way

- Some users went to alternative sites like Voat.

  Still a win for Reddit?

- Our algorithms detect only specific subsets of hate speech.

  Did hate speech shift to a form that is harder to detect?

Some conclusions

Diachronic text is a powerful tool that makes it possible to address big questions:

- Why do some innovations succeed and others fail?
- What makes some documents influential?
- Can we do anything about online hate speech?

Other modalities offer similar opportunities!

- Space[24]
- Social networks[25]

---

[24] Nguyen and Eisenstein 2017; Pavalanathan and Eisenstein 2015.

[25] Krishnan and Eisenstein 2015.

Outline

- Motivations and perspectives on diachronic text
- Case studies
  1. Lexical innovations in linguistic context
  2. Semantic progressiveness and citation influence
  3. Cause and effect in online hate speech
- **Concluding thoughts on CSS**

Prediction or explanation?[27]

- We are not interested in super accurate predictions from a black box model. This requires a mindset shift for computer scientists.
- In two of the three projects, we took explicit steps to make valid causal inferences from observational data.[26]
- In all three projects, predictive accuracy is a critical sanity check.

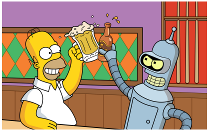
[26] Rosenbaum 2017.

[27] Breiman 2001.

Automatic or manual coding?

In each case, automation was used to identify cases of interest at scale:

1. **Lexical innovations**: consistently increasing frequency.
2. **Semantic changes**: substantial changes in embeddings over time.
3. **Hate speech**: high frequency in forums labeled as sites for hate speech.



But in cases 1 & 3, a second level of manual filtering was applied to increase the precision of the independent variable.

[27] For another approach to semi-automated coding, see Voigt et al. 2017.

Is this computer science or social science?

Collaborations are key to asking substantive questions that matter, and to understanding what is methodologically possible.

- Speed bumps: publication, funding, advising practices; Microsoft Word
- Major challenge: which axis of generalization to emphasize?

What's next

- **Ian Stewart** (postdoc candidate!): how does form of reference vary with information status over the course of a news event?
- **Sandeep Soni** (postdoc candidate soon): who were the leaders and followers of semantic changes in 19th century abolitionist newspapers?
- **Eshwar Chandrasekharan** (faculty candidate!): how can we use AI to make Reddit moderation easier?

Thank you!

Thank you!

- **Collaborators**: Eshwar Chandrasekharan, Lauren F. Klein, Eric Gilbert, Adam Glynn, Kristina Lerman, Umashanthi Pavalanathan, Sandeep Soni, Ian Stewart, Xiaochuang Han.
- **Sponsors**: National Endowment for the Humanities, National Science Foundation, National Institutes for Health, Air Force Office of Scientific Research, Google.

References I

Altmann, Eduardo G., Janet B. Pierrehumbert, and Adilson E. Motter (2011). "Niche as a determinant of word fate in online groups". In: *PloS one* 6.5, e19009.

Bakshy, Eytan et al. (2012). "The role of social networks in information diffusion". In: *Proceedings of the Conference on World-Wide Web (WWW)*, pp. 519–528.

Bouchard-Côté, Alexandre et al. (2013). "Automated reconstruction of ancient languages using probabilistic models of sound change". In: *Proceedings of the National Academy of Sciences*, p. 201204678.

Breiman, Leo (2001). "Statistical Modeling: the Two Cultures". In: *Statistical science* 16.3.

References II

Chandrasekharan, Eshwar et al. (2018). "You Can't Stay Here: The Effectiveness of Reddit's 2015 Ban Through the Lens of Hate Speech". In: *Proceedings of Computer-Supported Cooperative Work (CSCW)*.

D'Arcy, Alexandra (2012). "The diachrony of quotation: evidence from New Zealand English". In: *Language Variation and Change* 24.3, pp. 343–369.

Eisenstein, Jacob (2013). "What to do about bad language on the internet". In: *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pp. 359–369.

Eisenstein, Jacob, Amr Ahmed, and Eric P. Xing (2011). "Sparse Additive Generative Models of Text". In: *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 1041–1048.

References III

Eisenstein, Jacob, Brendan O'Connor, et al. (Nov. 2014). "Diffusion of Lexical Change in Social Media". In: *PLoS ONE* 9.

Fagyal, Zsuzsanna et al. (2010). "Centers and peripheries: Network roles in language change". In: *Lingua* 120.8, pp. 2061–2079.

Garley, Matt and Julia Hockenmaier (2012). "Beefmoves: dissemination, diversity, and dynamics of English borrowings in a German hip hop forum". In: *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 135–139.

References IV

Hamilton, William L, Jure Leskovec, and Dan Jurafsky (2016). "Cultural shift or linguistic drift? comparing two computational measures of semantic change". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*. Vol. 2016. NIH Public Access, p. 2116.

Hamilton, William, Jure Leskovec, and Dan Jurafsky (2016). "Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change". In: *Proceedings of the Association for Computational Linguistics (ACL)*.

Hirano, Keisuke and Guido W Imbens (2004). "The propensity score with continuous treatments". In: *Applied Bayesian modeling and causal inference from incomplete-data perspectives* 226164, pp. 73–84.

References V

Krishnan, Vinodh and Jacob Eisenstein (2015). "“You're Mr. Lebowski, I'm The Dude”: Inducing Address Term Formality in Signed Social Networks". In: *NAACL*.

Kulkarni, Vivek et al. (2015). "Statistically Significant Detection of Linguistic Change". In: *Proceedings of the Conference on World-Wide Web (WWW)*, pp. 625–635.

Michel, Jean-Baptiste et al. (2011). "Quantitative analysis of culture using millions of digitized books". In: *science* 331.6014, pp. 176–182.

Mikolov, Tomas et al. (2013). "Distributed representations of words and phrases and their compositionality". In: *Advances in Neural Information Processing Systems*, pp. 3111–3119.

Nguyen, Dong and Jacob Eisenstein (2017). "A Kernel Independence Test for Geographical Language Variation". In: *Computational Linguistics* 43 (3).

References VI

Pavalanathan, Umashanthi and Jacob Eisenstein (2015). "Confounds and Consequences in Geotagged Twitter Data". In: *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*.

Pierrehumbert, Janet B. (2010). "The dynamic lexicon". In: *Handbook of Laboratory Phonology*. Ed. by A. Cohn, M. Huffman, and C. Fougeron. Oxford University Press, pp. 173–183.

Romero, Daniel M., Brendan Meeder, and Jon Kleinberg (2011). "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter". In: *Proceedings of the Conference on World-Wide Web (WWW)*, pp. 695–704.

Rosenbaum, Paul R (2017). *Observation and experiment: an introduction to causal inference*. Harvard University Press.

References VII

Rosenbaum, Paul R and Donald B Rubin (1983). "The central role of the propensity score in observational studies for causal effects". In: *Biometrika* 70.1, pp. 41–55.

Rosenfeld, Alex and Katrin Erk (2018). "Deep Neural Models of Semantic Shift". In: *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pp. 474–484.

Rotabi, Rahmtin, Cristian Danescu-Niculescu-Mizil, and Jon Kleinberg (2017). "Competition and selection among conventions". In: *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, pp. 1361–1370.

References VIII

- Soni, Sandeep, Kristina Lerman, and Jacob Eisenstein (2019). "Follow the Leader: Documents on the Leading Edge of Semantic Change Get More Citations". In: *arXiv preprint arXiv:1909.04189*.
- Stewart, Ian and Jacob Eisenstein (2018). "Making "fetch" happen: The influence of social and linguistic context on the success of lexical innovations". In: *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*.
- Tan, Chenhao and Lillian Lee (2015). "All who wander: On the prevalence and characteristics of multi-community engagement". In: *Proceedings of the Conference on World-Wide Web (WWW)*, pp. 1056–1066.
- Traugott, Elizabeth Closs and Richard B Dasher (2001). *Regularity in semantic change*. Cambridge University Press.

References IX

Voigt, Rob et al. (2017). "Language from police body camera footage shows racial disparities in officer respect". In: *Proceedings of the National Academy of Sciences*, p. 201702413.

Yogatama, Dani et al. (2011). "Predicting a Scientific Community's Response to an Article". In: *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 594–604.