Measuring and Modeling Language Change

Jacob Eisenstein

Objectives

In this tutorial, you will learn about:

- the questions that can be asked of time-stamped text, across linguistics, social science, and the humanities;
- statistical methods for answering these questions, via **hands-on replications**;
- how to validate these methods;
- limitations of current NLP approaches, and prospects for doing better.

Outline

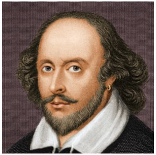
- Motivations and perspectives on language change
- Practical methods and case studies

  1. Word frequency
  2. Differences
  3. Word meaning
  4. Leaders and followers
  5. Cause and effect

- Next steps

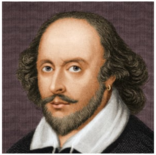
Change as a constant

Full fathom five thy father lies; Of his bones are coral made.

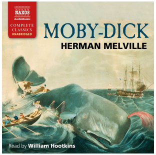


Change as a constant

Full fathom five thy father lies; Of his bones are coral made.

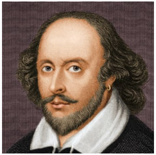


Aye, aye! it was that accursed white whale that razed me; made a poor pegging lubber of me for ever and a day!

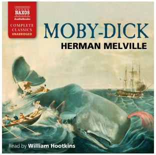


Change as a constant

Full fathom five thy father lies; Of his bones are coral made.

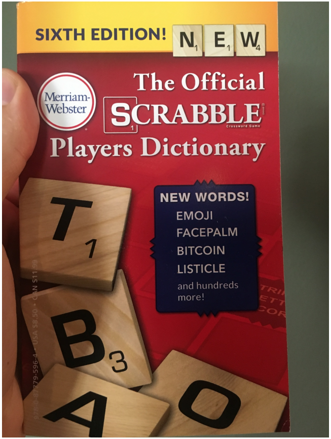


Aye, aye! it was that accursed white whale that razed me; made a poor pegging lubber of me for ever and a day!



Now if you'll excuse me, I'm going to go on an overnight drunk, and in 10 days I'm going to set out to find the shark that ate my friend and destroy it.

Short term change



Even shorter term change

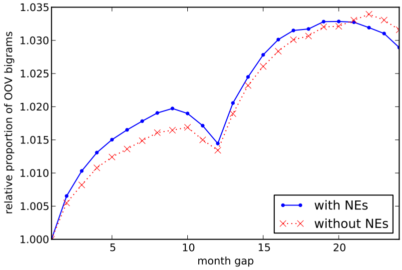
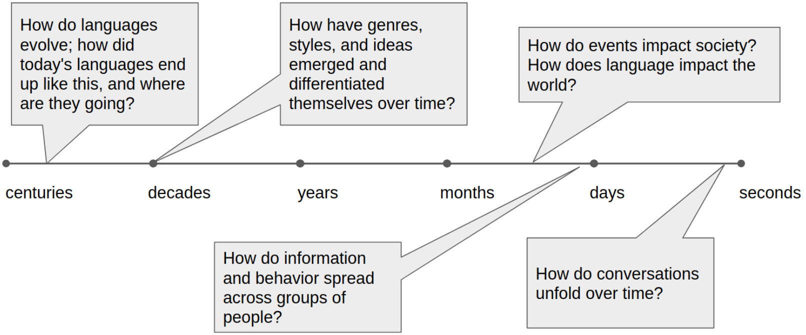


Figure: from Eisenstein (2013)

Some questions about language change



What changed?

Systemic change: what counts as grammatical / coherent / felicitous / appropriate in a language.

- full fathom five thy father lies
- historical linguistics and sociolinguistics

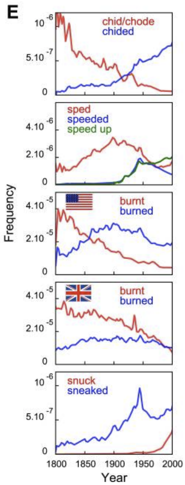


Figure: Michel et al. 2011

What changed?

Usage change: what people talk about.

- steam/electricity, horse/train
- digital humanities, computational social science, social computing

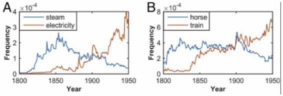


Figure: Lansdall-Welfare et al. 2017

Language change and sociolinguistics

Weinreich, Labov, and Herzog (1968) present five problems:

- **Constraints**: what changes are possible?
- **Transition**: how does a change propagate in a community of speakers?
- **Embedding**: what implications does a change have for the larger linguistic system?
- **Evaluation**: what is the social meaning of a particular change?
- **Actuation**: why this change, and why now?

Sources of information about language change

- **Corpora**: time-stamped ("diachronic") texts
- **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

Sources of information about language change

- **Corpora**: time-stamped ("diachronic") texts
- **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

	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

- **Corpora**: time-stamped ("diachronic") texts
- **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

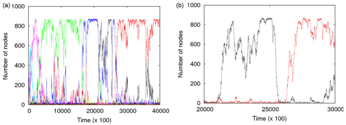


Figure: Fagyal et al. 2010

Sources of information about language change

- **Corpora**: time-stamped ("diachronic") texts
- **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

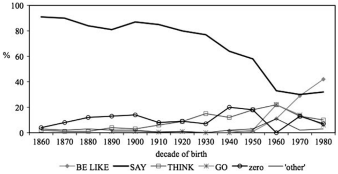


Figure: D'Arcy 2012

Language change beyond linguistics

- **Computational social science** (*text-as-data*)

  public attention, political communication, information diffusion, issue framing

- **Digital humanities** (*distant reading*)

  evolution of genres and writing styles, attitudes about race and gender, detection of influence

- **Human-computer interaction** (*social computing*)

  impact of social media on well-being, effects of policy changes on online communities, formation of social norms online

Secret agenda

Most of these fields are interested in **explanation**, not **prediction**.[1]

- Natural language processing can play an important role, by **operationalizing** variables of interest.
- But! Rigorously evaluating explanations is different and usually harder than evaluating predictions, and requires a different way of thinking.
- The tutorial includes several "case studies" to get you started.

[1] Shmueli 2010; Hofman, Sharma, and Watts 2017.

Outline

- Motivations and perspectives on language change
- Practical methods and case studies
  1. Word frequency
  2. Differences
  3. Word meaning
  4. Leaders and followers
  5. Cause and effect
- Next steps

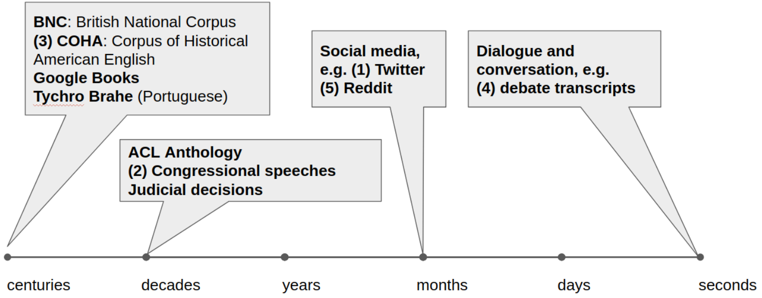
Outline

- Motivations and perspectives on language change
- Practical methods and case studies

1. Word frequency	Are you happier on the weekend?
2. Differences	Is polarization on the rise?
3. Word meaning	When did peers apply pressure?
4. Leaders and followers	Who is shaping the debate?
5. Cause and effect	Should we feed the trolls?

- Next steps

Case studies and data: a temporal view



Outline

- Motivations and perspectives on language change
- Practical methods and case studies

1. Word frequency	Are you happier on the weekend?
2. Differences	Is polarization on the rise?
3. Word meaning	When did peers apply pressure?
4. Leaders and followers	Who is shaping the debate?
5. Cause and effect	Should we feed the trolls?

- Next steps

Outline

- Motivations and perspectives on language change
- Practical methods and case studies

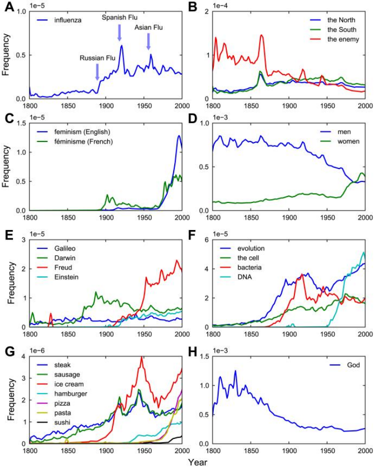
1. **Word frequency**	Are you happier on the weekend?
2. Differences	Is polarization on the rise?
3. Word meaning	When did peers apply pressure?
4. Leaders and followers	Who is shaping the debate?
5. Cause and effect	Should we feed the trolls?

- Next steps

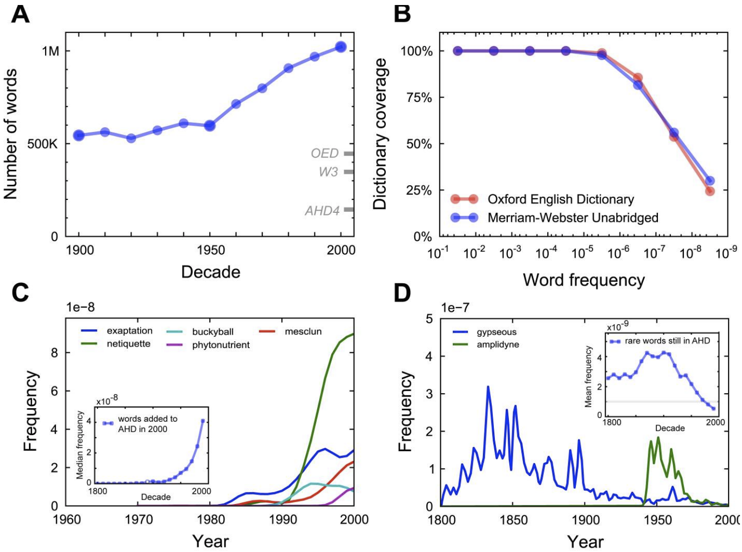
Big data and word frequency

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

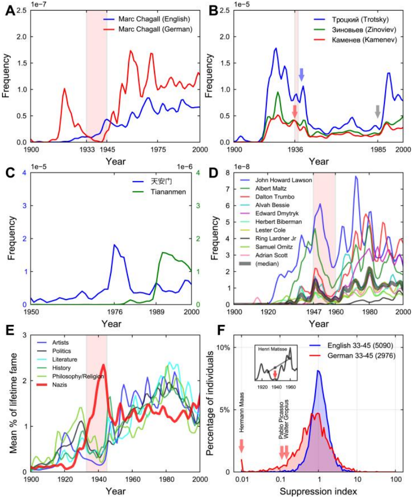
- 5M digitized books: "4% of all books ever published"
- 500B word tokens, 361B in English



Culturnomical lexicography



Culturnomical political science



Questions

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

- Which words? (what even is a word?)
- Which texts?
- What to count?
- Word frequencies and events have timestamps, but did one really cause the other?

Problemf recognifing characterf

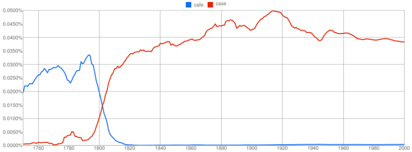


Figure: case vs cafe

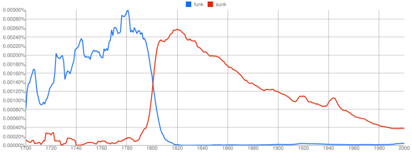
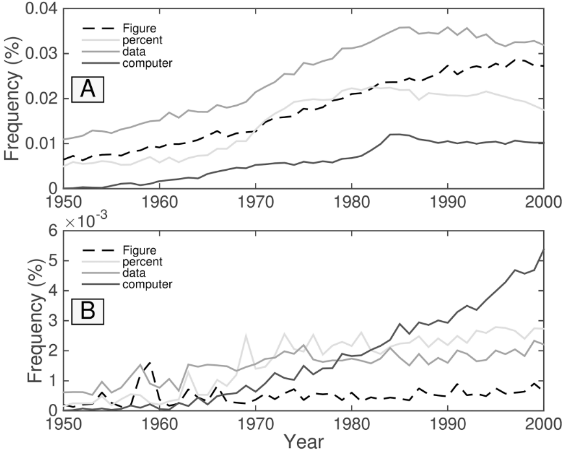


Figure: sunk vs funk

Liberman (2010) argues that OCR problems are a special case of a more general issue: Google Books' focus on strings.

Composition effects

Scientific articles occupy an increasing proportion of Google Books over time.[2]



[2]Pechenick, Danforth, and Dodds 2015.

Artisanal corpora

Corpus linguists have not produced datasets on the scale of Google books, but pay more attention to digitization and balance across genres and topics.

- Corpus of Historical American English (COHA; Davies 2012, 400M words)
- British National Corpus (BNC; Leech and Rayson 2014, 100M words)
- Global Web-based English Corpus (GlowBe; Davies and Fuchs 2015, 1.9B words)

Artisanal corpora

CHANGE TO VERTICAL CHART / CLICK TO SEE CONTEXT

SECTION	ALL	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
FREQ	586	0	1	1	1	0	28	1	14	13	17	22	43	46	50	20	21	24	91	97	96
WORDS (M)	405	1.2	6.9	13.8	16.0	16.5	17.1	18.6	20.3	20.6	22.1	22.7	25.7	24.6	24.3	24.5	24.0	23.8	25.3	27.9	29.6
PER MIL	1.45	0.00	0.14	0.07	0.06	0.00	1.64	0.05	0.69	0.63	0.77	0.97	1.68	1.87	2.05	0.81	0.88	1.01	3.59	3.47	3.25

SEE ALL YEARS AT ONCE

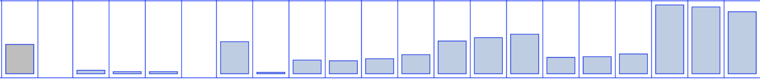


Figure: Frequency of funk in the Corpus of Historical American English (COHA)

A case study

Golder and Macy 2011: Diurnal and Seasonal Mood Varywith Work, Sleep, and Daylength Across Diverse Cultures

- Main question: how does mood shift across the day, week, and season?

  Prior work has “relied heavily on small homogeneous samples of American undergraduates who are not necessarily representative of the larger population”

- Text: positive and negative affect lexicons from LIWC[3]

---

[3] Tausczik and Pennebaker 2010.

Replication time!

Launch the jupyter notebook and follow along:

```
jupyter-notebook HappierOnTheWeekend.ipynb
```

Validation and robustness

- Did population differences drive the results?

  "We removed between-individual differences by mean-centering the measures of PA and NA at the individual level"

- Does LIWC measure affect?

  "For all emotional expression words, LIWC's sensitivity and specificity values were 0.88 and 0.97[4]"

- Differences are driven by cultural norms of self-expression?

  "Because these norms are unlikely to be universal, the robust patterns we observed across diverse cultures give us confidence that affective expression is a reliable indicator . . ."

[4] Bantum and Owen 2009.

Outline

- Motivations and perspectives on language change
- Practical methods and case studies

1. **Word frequency**	Are you happier on the weekend?
2. Differences	Is polarization on the rise?
3. Word meaning	When did peers apply pressure?
4. Leaders and followers	Who is shaping the debate?
5. Cause and effect	Should we feed the trolls?

- Next steps

Outline

- Motivations and perspectives on language change
- Practical methods and case studies

1.	Word frequency	Are you happier on the weekend?
2.	**Differences**	Is polarization on the rise?
3.	Word meaning	When did peers apply pressure?
4.	Leaders and followers	Who is shaping the debate?
5.	Cause and effect	Should we feed the trolls?

- Next steps

The importance of difference

Many questions about time are questions about *difference*:

- Is political polarization increasing?
- Is polarization becoming more about identity groups than policy ideas?[5]
- Are gender roles becoming less stereotypical in fiction?
- Are NAACL and EMNLP becoming more or less similar?

[5] Iyengar, Sood, and Lelkes 2012.

Quantifying differences

What to measure:

- word frequencies;[6]
- latent topics[7]
- classifier accuracy.[8]

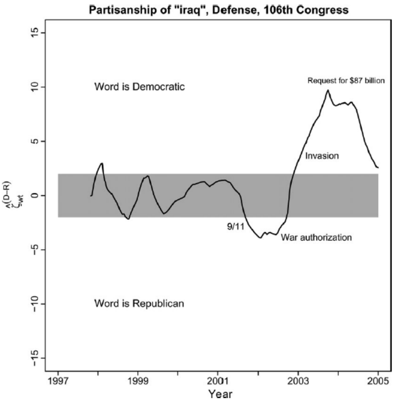
[6] Monroe, Colaresi, and Quinn 2008; Gentzkow, Shapiro, and Taddy 2016.

[7] Hall, Jurafsky, and Manning 2008; Tsur and Rappoport 2015; Barron et al. 2018.

[8] Peterson and Spirling 2018; Underwood, Bamman, and Lee 2018.

Frequencies: polarization of single words

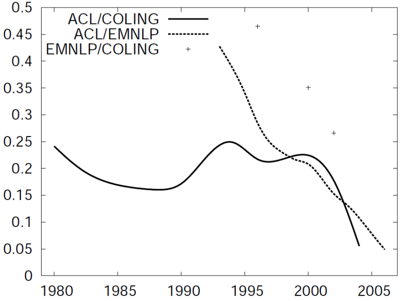
From Monroe, Colaresi, and Quinn (2008), “Fightin words”



We will replicate this figure!

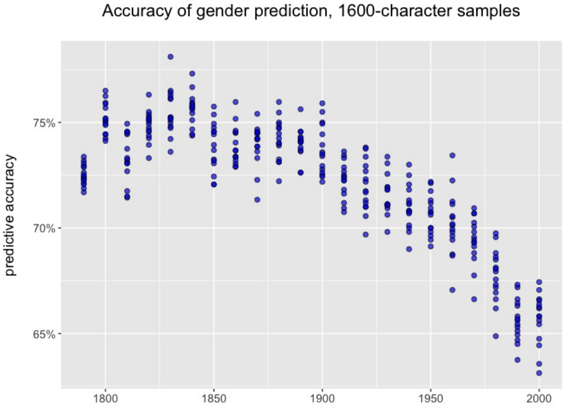
Topics: conferences over time

From Hall, Jurafsky, and Manning (2008)



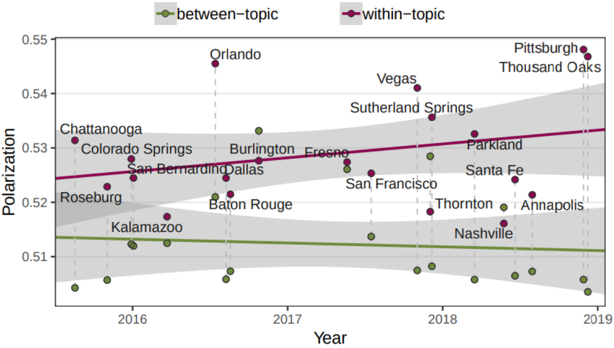
Classifiers: gender stability of fictional characters

From Underwood, Bamman, and Lee (2018)



Polarization between and within topics

Demszky et al. (2019) combine lexical and topic analysis to show how polarization about mass shootings is increasing.



Quantifying differences

What to measure:

- word frequencies;[6]
- latent topics[7]
- classifier accuracy.[8]

---

[6] Monroe, Colaresi, and Quinn 2008; Gentzkow, Shapiro, and Taddy 2016.

[7] Hall, Jurafsky, and Manning 2008; Tsur and Rappoport 2015; Barron et al. 2018.

[8] Peterson and Spirling 2018; Underwood, Bamman, and Lee 2018.

Quantifying differences

What to measure:

- word frequencies;[6]

  relies fairly directly on the original data, but word frequencies are high-variance and not really IID.

- latent topics[7]

- classifier accuracy.[8]

---

[6] Monroe, Colaresi, and Quinn 2008; Gentzkow, Shapiro, and Taddy 2016.

[7] Hall, Jurafsky, and Manning 2008; Tsur and Rappoport 2015; Barron et al. 2018.

[8] Peterson and Spirling 2018; Underwood, Bamman, and Lee 2018.

Quantifying differences

What to measure:

- word frequencies;[6]

  relies fairly directly on the original data, but word frequencies are high-variance and not really IID.

- latent topics[7]

  aggregates related words, but may depend on idiosyncrasies of preprocessing and hyperparameters

- classifier accuracy.[8]

---

[6] Monroe, Colaresi, and Quinn 2008; Gentzkow, Shapiro, and Taddy 2016.

[7] Hall, Jurafsky, and Manning 2008; Tsur and Rappoport 2015; Barron et al. 2018.

[8] Peterson and Spirling 2018; Underwood, Bamman, and Lee 2018.

Quantifying differences

What to measure:

- word frequencies;[6]

  relies fairly directly on the original data, but word frequencies are high-variance and not really IID.

- latent topics[7]

  aggregates related words, but may depend on idiosyncrasies of preprocessing and hyperparameters

- classifier accuracy.[8]

  powerful classifiers can relax IID assumption, but are computationally expensive and sensitive to dataset design

---

[6] Monroe, Colaresi, and Quinn 2008; Gentzkow, Shapiro, and Taddy 2016.

[7] Hall, Jurafsky, and Manning 2008; Tsur and Rappoport 2015; Barron et al. 2018.

[8] Peterson and Spirling 2018; Underwood, Bamman, and Lee 2018.

Aggregation and variance

Main challenge in measuring differences over time in word frequencies is high variance:

> *The number of phrases a speaker could choose is large relative to the total amount of speech we observe, so many phrases are said mostly by one party or the other purely by chance.*
>
> *Naive estimators interpret such differences as evidence of partisanship, leading to a bias we show can be many orders of magnitude larger than the true signal in the data.*[9]

Monroe, Colaresi, and Quinn (2008): use MAP estimate of word frequency rates rather than empirical frequencies.

[9] *Gentzkow, Shapiro, and Taddy 2016.*

Replication time!

Replication: measuring and aggregating word-level partisanship in *Fightin Words*.

```
jupyter-notebook SameDifference.ipynb
```

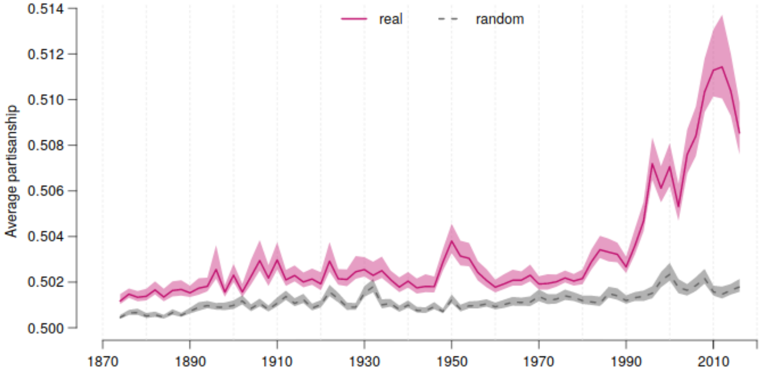
Partisanship by L1-regularized modeling

Gentzkow, Shapiro, and Taddy (2016):

- For each group g, estimate probabilities for each word j at time t,
$$q_{t,j}^{(g)} \propto \exp(\mu_{t,j} + \phi_{t,j}^{(g)}), \tag{1}$$
with $\phi_t^{(g)}$ optimized under L1 regularization.[10]
- They then compute partisanship as the expected per-word classification accuracy of a Naive Bayes classification model.

[10] Eisenstein, Ahmed, and Xing 2011; Taddy 2013.

Partisanship by L1: results[11]

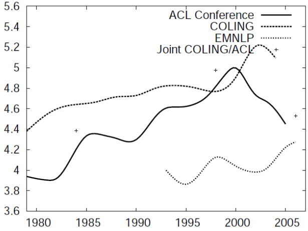


Confidence intervals require parametric bootstrap sampling, which we will discuss later.

[11] Gentzkow, Shapiro, and Taddy 2016.

Information theoretic measures of difference

Entropy: how concentrated is probability mass over words/topics, $H(p) = -\sum_j p_j \log p_j$.



In this figure,[12] p is defined over latent topics rather than words.[13]

[12] Hall, Jurafsky, and Manning 2008.

[13] See also Doyle and Frank (2015) on entropy of tweets over the course of a baseball game.

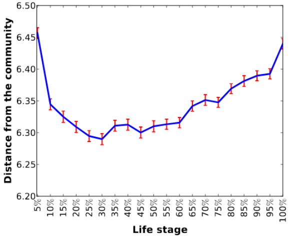
Individuals versus communities

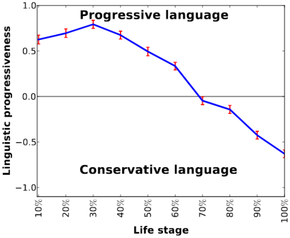
How similar is an individual's language to the community?

- Danescu-Niculescu-Mizil, West, et al. (2013) construct a "snapshot language model" for each month, and then compute the **cross-entropy**,

$$H(\boldsymbol{x}_{t,i}, \boldsymbol{q}_t) = -\frac{1}{N_{t,i}} \sum_j x_{t,i,j} \log q_{t,j}.$$

- *Progressiveness* is defined as the relative offset of the month that minimizes cross-entropy.





Information theoretic differences over time

- Kullback-Liebler (KL) divergence between p and q:

$$D_{\text{KL}}(p||q) = \sum_j p_j \log \frac{p_j}{q_j}, \tag{2}$$

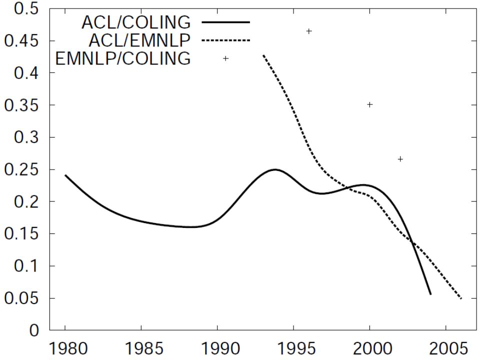
  also known as **relative entropy**.

- The Jensen-Shannon divergence is a symmetric measure based on KL-divergence,

$$\begin{aligned} D_{\text{JS}}(p||q) =& \frac{1}{2} D_{\text{KL}}(p||r) + \frac{1}{2} D_{\text{KL}}(q||r) & (3) \\ r =& \frac{1}{2} p + \frac{1}{2} q. & (4) \end{aligned}$$

Difference between conferences over time

Jensen-Shannon divegences between pairs of conferences, based on latent topic proportions:[14]



[14] Hall, Jurafsky, and Manning 2008.

The information theory of the French Revolution

Application to a dataset of 40K speeches from the *Archives Parliamentaires* (1787-1794):[15]

- "the celebrated radicals Robespierre and Pétion" diverged from the past and shaped the future;
- comments by technical specialists diverged from both past and future;
- the right wing was similar to both the past and future.

[15] Barron et al. 2018.

Outline

- Motivations and perspectives on language change
- Practical methods and case studies

1. Word frequency	Are you happier on the weekend?
2. **Differences**	Is polarization on the rise?
3. Word meaning	When did peers apply pressure?
4. Leaders and followers	Who is shaping the debate?
5. Cause and effect	Should we feed the trolls?

- Next steps

Outline

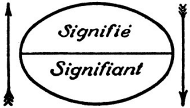
- Motivations and perspectives on language change
- Practical methods and case studies

1. Word frequency	Are you happier on the weekend?
2. Differences	Is polarization on the rise?
3. **Word meaning**	When did peers apply pressure?
4. Leaders and followers	Who is shaping the debate?
5. Cause and effect	Should we feed the trolls?

- Next steps

Change in the lexicon

Lexical innovation can happen on the level of new wordforms (signs) and new meanings (signifieds).



- Changes in a corpus may be driven by new real-world events and entities (e.g., email, tablet).
- Linguistic "fashions" involve new signs for existing meanings (lol).
- Existing signs can be repurposed to new meanings (hot, actually): "linguistic drifts" versus "cultural shifts"[16]

Changes can reverberate throughout the lexical system.[17]

[16] Traugott and Dasher 2001; Hamilton, Leskovec, and Jurafsky 2016a.

[17] Pierrehumbert 2010.

Meaning change in distributional semantics

Distributional representations of word meaning have a very long history in NLP:[18]

- latent semantic analysis
- Brown clusters
- word2vec

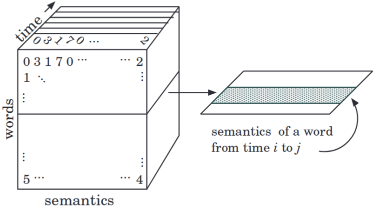


Figure: Jurgens and Stevens (2009)

Hypothesis: if the meaning of word w is summarized by a distributional representation $\boldsymbol{u}_w$, then changes in meaning should be reflected in $\boldsymbol{u}_w$.

[18]Deerwester et al. 1990; Brown et al. 1992; Mikolov, Sutskever, et al. 2013.

Word embeddings and lexical semantics

For our purposes a **word embedding** has two properties:

- it is a fixed-length vector that is computed from the set of contexts in which a word appears;
- if words i and j are semantically similar, then the vector correlation is large:[19]

$$\text{corr}_{i,j} = \frac{\boldsymbol{u}_i \cdot \boldsymbol{u}_j}{||\boldsymbol{u}_i|| \times ||\boldsymbol{u}_j||}. \tag{5}$$

As a corollary, a change in a word's embedding implies a change in meaning.

[19] This property seems to hold at a coarse level (Mikolov, Yih, and Zweig 2013) but important questions remain (Antoniak and Mimno 2018; Wendlandt, Kummerfeld, and Mihalcea 2018).

Anecdotal evidence

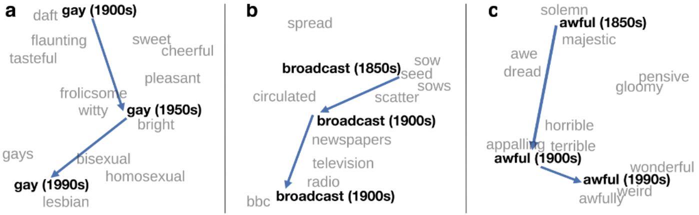
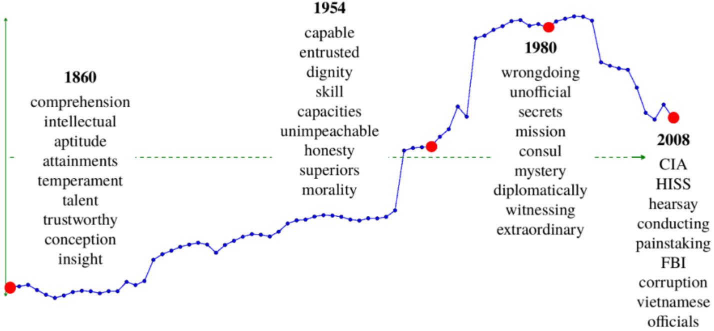


Figure: Figure from Hamilton, Leskovec, and Jurafsky (2016). For a closely related idea using raw word-context counts, see Gulordava and Baroni (2011); for an early clustering-based approach, see Wijaya and Yeniterzi (2011)

Anecdotal evidence



(b) INTELLIGENCE in U.S. Senate speeches (1858–2009)

Figure: Figure from Rudolph and Blei (2018)

Research timeline

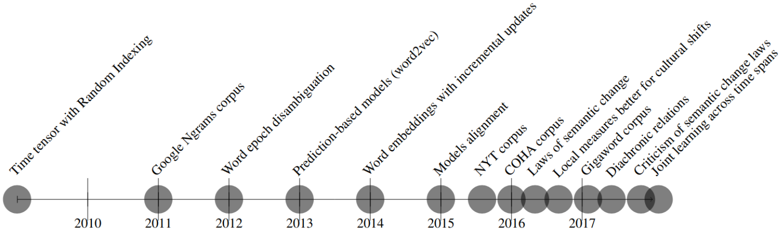


Figure: Kutuzov et al. 2018

Lots of work in this area!

- We will focus on a few relatively simple techniques.
- For much more, there are several recent surveys.[20].

[20]Kutuzov et al. 2018; Tahmasebi, Borin, and Jatowt 2018; Tang 2018.

Detecting changes in meaning

General recipe:

1. Estimate temporal word embeddings, $\boldsymbol{u}_i^{(t)}$.
2. Aggregate embeddings into a single score, $\{\boldsymbol{u}_i^{(t)}\}_{t=1}^{T} \mapsto \psi_i$.
3. Evaluate against a small list of known semantic changes in English.[21]

[21] Hamilton, Leskovec, and Jurafsky 2016b.

Detecting changes in meaning

General recipe:

1. Estimate temporal word embeddings, $\boldsymbol{u}_i^{(t)}$.
2. Aggregate embeddings into a single score, $\{\boldsymbol{u}_i^{(t)}\}_{t=1}^{T} \mapsto \psi_i$.
3. Evaluate against a small list of known semantic changes in English.[21]

[21] Hamilton, Leskovec, and Jurafsky 2016b.

Estimating temporal word embeddings

Most papers divide the corpus into "epochs", and estimate embeddings separately in each epoch.[22]

- The embeddings must be *aligned* to be comparable.
- The epochs should be similar in size and composition.

There are a few models that operate over the entire corpus:

- Bayesian priors:[23] $\boldsymbol{u}_i^{(t)} \sim f(\boldsymbol{u}_i^{(t-1)})$.
- Refinement by feedforward network:[24] $\boldsymbol{u}_i^{(t)} = g(t, \overline{\boldsymbol{u}}_i)$.

---

[22] e.g., Kulkarni et al. 2015; Hamilton, Leskovec, and Jurafsky 2016b.

[23] Bamler and Mandt 2017; Rudolph and Blei 2018.

[24] Rosenfeld and Erk 2018.

Procrustes alignment of word embeddings

- Cross-epoch embeddings may be superficially different but fundamentally the same – just permute the columns!
- To align word embeddings across epochs, we solve the orthogonal Procrustes problem,

$$\min_{\Omega^\top \Omega = \mathbb{I}} ||\Omega A - B||_F,$$

where $||M||_F$ is defined as $\sqrt{\sum_i \sum_j m_{i,j}^2}$.
- The solution is obtained by singular value decomposition,

$$\begin{aligned} U\Sigma V^\top =& \mathrm{SVD}(BA^\top) \\ \Omega =& UV^\top. \end{aligned}$$

Detecting changes in meaning

General recipe:

1. Estimate temporal word embeddings, $\boldsymbol{u}_i^{(t)}$.
2. Aggregate embeddings into a single score, $\{\boldsymbol{u}_i^{(t)}\}_{t=1}^{T} \mapsto \psi_i$.
3. Evaluate against a small list of known semantic changes in English.[21]

[21] Hamilton, Leskovec, and Jurafsky 2016b.

Detecting changes in meaning

General recipe:

1. Estimate temporal word embeddings, $\boldsymbol{u}_i^{(t)}$.
2. Aggregate embeddings into a single score, $\{\boldsymbol{u}_i^{(t)}\}_{t=1}^{T} \mapsto \psi_i$.
3. Evaluate against a small list of known semantic changes in English.[21]

[21] Hamilton, Leskovec, and Jurafsky 2016b.

Aggregating word embedding differences

Hamilton, Leskovec, and Jurafsky (2016) distinguish *global* and *local* metrics for embedding differences over time:

- **Global**: compare $\boldsymbol{u}_i^{(t)}$ and $\boldsymbol{u}_i^{(t')}$ (after projection), e.g.,

$$\psi_{t,i}^{(G)} = 1 - \cos(\boldsymbol{u}_i^{(t)}, \boldsymbol{u}_i^{(t+\delta)}) \tag{6}$$

- **Local**:[25]
  1. identify nearest neighbors of i at times t and t';
  2. compute similarity to each of these neighbors, arrange as vector $\boldsymbol{s}^{(t)}$;
  3. compute $\psi_{t,i}^{(L)} = 1 - \cos(\boldsymbol{s}^{(t)}, \boldsymbol{s}^{(t')})$.

---

[25] See Kulkarni et al. (2015) for an alternative neighbor-based approach.

Detecting changes in meaning

General recipe:

1. Estimate temporal word embeddings, $\boldsymbol{u}_i^{(t)}$.
2. Aggregate embeddings into a single score, $\{\boldsymbol{u}_i^{(t)}\}_{t=1}^{T} \mapsto \psi_i$.
3. Evaluate against a small list of known semantic changes in English.[21]

[21] Hamilton, Leskovec, and Jurafsky 2016b.

Detecting changes in meaning

General recipe:

1. Estimate temporal word embeddings, $\boldsymbol{u}_i^{(t)}$.
2. Aggregate embeddings into a single score, $\{\boldsymbol{u}_i^{(t)}\}_{t=1}^{T} \mapsto \psi_i$.
3. Evaluate against a small list of known semantic changes in English.[21]

[21] Hamilton, Leskovec, and Jurafsky 2016b.

Validation

Hamilton, Leskovec, and Jurafsky (2016): 28 word pairs, distinguish whether they are moving together or apart.

Word	Moving towards	Moving away	Shift start	Source
gay	homosexual, lesbian	happy, showy	ca 1920	(Kulkarni et al., 2014)
fatal	illness, lethal	fate, inevitable	<1800	(Jatowt and Duh, 2014)
awful	disgusting, mess	impressive, majestic	<1800	(Simpson et al., 1989)
nice	pleasant, lovely	refined, dainty	ca 1900	(Wijaya and Yeniterzi, 2011)
broadcast	transmit, radio	scatter, seed	ca 1920	(Jeffers and Lehiste, 1979)
monitor	display, screen	—	ca 1930	(Simpson et al., 1989)
record	tape, album	—	ca 1920	(Kulkarni et al., 2014)
guy	fellow, man	—	ca 1850	(Wijaya and Yeniterzi, 2011)
call	phone, message	—	ca 1890	(Simpson et al., 1989)

[26] Mihalcea and Nastase 2012; Popescu and Strapparava 2015.

Validation

Hamilton, Leskovec, and Jurafsky (2016): 28 word pairs, distinguish whether they are moving together or apart.

Word	Moving towards	Moving away	Shift start	Source
gay	homosexual, lesbian	happy, showy	ca 1920	(Kulkarni et al., 2014)
fatal	illness, lethal	fate, inevitable	<1800	(Jatowt and Duh, 2014)
awful	disgusting, mess	impressive, majestic	<1800	(Simpson et al., 1989)
nice	pleasant, lovely	refined, dainty	ca 1900	(Wijaya and Yeniterzi, 2011)
broadcast	transmit, radio	scatter, seed	ca 1920	(Jeffers and Lehiste, 1979)
monitor	display, screen	—	ca 1930	(Simpson et al., 1989)
record	tape, album	—	ca 1920	(Kulkarni et al., 2014)
guy	fellow, man	—	ca 1850	(Wijaya and Yeniterzi, 2011)
call	phone, message	—	ca 1890	(Simpson et al., 1989)

- Precision-based evaluations: are the changes that the model identifies meaningful?
- Predictive evaluations: given a passage, when was it written?[26]

[26] Mihalcea and Nastase 2012; Popescu and Strapparava 2015.

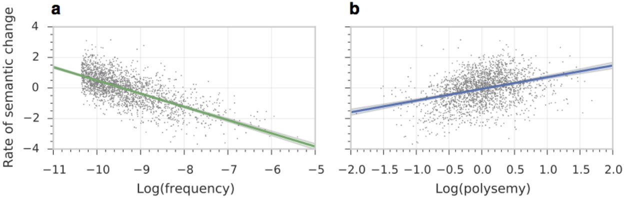
Replication time!

Replication: measuring and aggregating semantic change.[27]

```
jupyter-notebook DirtyLaundering.ipynb
```

[27] Data from Hamilton (2016), who is hereby nominated for the replicability hall of fame.

Laws of semantic change



Hamilton, Leskovec, and Jurafsky (2016):

- **Conformity**: frequently-used words are more stable.
- **Innovation**: polysemous words change more quickly.

Eger and Mehler (2016):

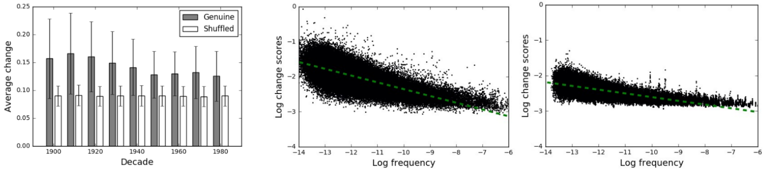
- **Linearity**: self-similarity decays linearly over time.

Robustness check

Dubossarsky, Weinshall, and Grossman (2017): randomly shuffle timestamps, see if these laws still hold.

Robustness check

Dubossarsky, Weinshall, and Grossman (2017): randomly shuffle timestamps, see if these laws still hold.



- Original data: $r = -0.748$, shuffled data: $r = -0.747$
- Similar findings for polysemy.
- These results are for SVD-based word embeddings, not tested for skipgrams.

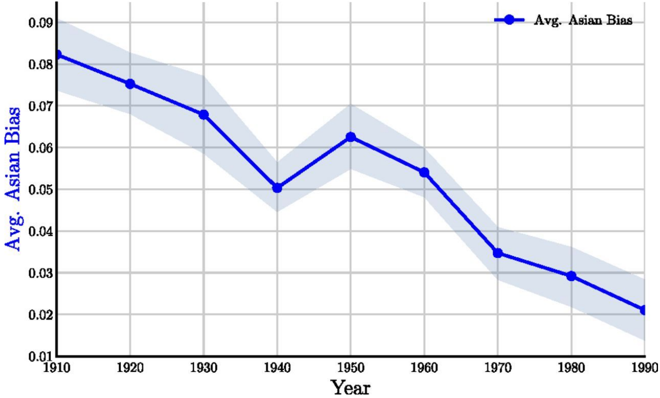
Diachronic word embeddings and bias

Garg et al. (2018): words whose embeddings are more similar to the embeddings of Asian-American last names than White-American last names, based on COHA embeddings.

1910	1950	1990
Irresponsible	Disorganized	Inhibited
Envious	Outrageous	Passive
Barbaric	Pompous	Dissolute
Aggressive	Unstable	Haughty
Transparent	Effeminate	Complacent
Monstrous	Unprincipled	Forceful
Hateful	Venomous	Fixed
Cruel	Disobedient	Active
Greedy	Predatory	Sensitive
Bizarre	Boisterous	Hearty

Diachronic word embeddings and bias

Garg et al. (2018): similarity between Asian-American name embeddings (COHA) and a lexicon of words characterizing "outsiders."



Outline

- Motivations and perspectives on language change
- Practical methods and case studies

1. Word frequency	Are you happier on the weekend?
2. Differences	Is polarization on the rise?
3. **Word meaning**	When did peers apply pressure?
4. Leaders and followers	Who is shaping the debate?
5. Cause and effect	Should we feed the trolls?

- Next steps

Outline

- Motivations and perspectives on language change
- Practical methods and case studies

1. Word frequency	Are you happier on the weekend?
2. Differences	Is polarization on the rise?
3. Word meaning	When did peers apply pressure?
4. **Leaders and followers**	Who is shaping the debate?
5. Cause and effect	Should we feed the trolls?

- Next steps

Language change and sociolinguistics

Weinreich, Labov, and Herzog (1968) present five problems:

- **Constraints**: what changes are possible?
- **Transition**: how does a change propagate in a community of speakers?
- **Embedding**: what implications does a change have for the larger linguistic system?
- **Evaluation**: what is the social meaning of a particular change?
- **Actuation**: why this change, and why now?

Language change and sociolinguistics

Weinreich, Labov, and Herzog (1968) present five problems:

- **Constraints**: what changes are possible?
- **Transition**: how does a change propagate in a community of speakers?
- **Embedding**: what implications does a change have for the larger linguistic system?
- **Evaluation**: what is the social meaning of a particular change?
- **Actuation**: why this change, and why now?

Who leads? Who follows? Who resists?

Language change on the social network: nodes

- Language change is transmitted across a social network of individuals.
- The likelihood of a speaker to adopt a change, and their success at propagating a change, depends on their position in the network.[28]

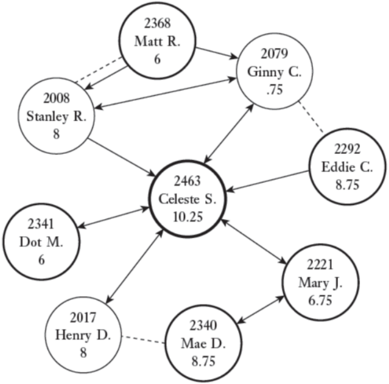


Figure: Labov (2001)

[28]Labov 2001; Milroy 1991.

Language change on the social network: dyads

- The transmission of an innovation between two individuals requires a communication channel and a social evaluation.
- Diachronic text data makes it possible to reconstruct the latent network of linguistic influence.

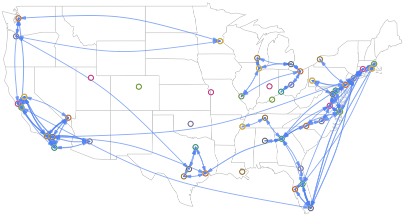


Figure: Eisenstein, O'Connor, et al. (2014)

Language change on the social network: attributes

- Social connections and status are shaped by attributes such as race, class, and gender.
- Diachronic language data can help to unpack the social meaning of such attribute through their influence on language change.

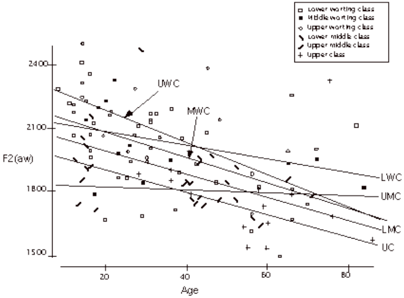


Figure: Labov (2001)

Leadership and influence across time scales

- **Decades**: systemic change in phonology and syntax[29]
- **Years**: usage change in subcommunities such as scientific research articles[30]
- **Months**: adoption of conventions and styles, e.g., in social media[31]
- **Days**: spread of news stories and memes[32]
- **Minutes**: linguistic accommodation[33] and topic control[34]

---

[29] Labov 2001.

[30] e.g., Gerow et al. 2018.

[31] e.g., Kooti et al. 2012; Eisenstein, O'Connor, et al. 2014; Goel et al. 2016; Del Tredici and Fernández 2018; Stewart and Eisenstein 2018.

[32] e.g., Leskovec, Backstrom, and Kleinberg 2009.

[33] Danescu-Niculescu-Mizil and Lee 2011.

[34] Nguyen et al. 2014.

Estimating influence from diachronic data

- Goal is to estimate parameters $\alpha_{i \to j}$ for all i, j in population $\mathcal{I}$.
- Given covariates $\{\boldsymbol{x}_i\}_{i \in \mathcal{I}}$, we may set $\alpha_{i \to j} = f(\boldsymbol{x}_i, \boldsymbol{x}_j)$.
- Observed data:
  - **Time series**: counts or proportions $\boldsymbol{y}_{t,i}$ at each $t = (1, 2, \ldots, T)$.
  - **Event cascades**: a set of events $\mathcal{E} = \{(s_n, t_n)\}$, where $s_n \in \mathcal{I}$ is the source and $t_n \in \mathbb{R}_+$ is the timestamp.
- Typically, we optimize $\{\alpha_{i \to j}\}$ to assign high likelihood to the observations in some probability model.

Estimating influence from diachronic data

- Goal is to estimate parameters $\alpha_{i \to j}$ for all i, j in population $\mathcal{I}$.
- Given covariates $\{\boldsymbol{x}_i\}_{i \in \mathcal{I}}$, we may set $\alpha_{i \to j} = f(\boldsymbol{x}_i, \boldsymbol{x}_j)$.
- Observed data:
  - **Time series**: counts or proportions $\boldsymbol{y}_{t,i}$ at each $t = (1, 2, \ldots, T)$.
  - **Event cascades**: a set of events $\mathcal{E} = \{(s_n, t_n)\}$, where $s_n \in \mathcal{I}$ is the source and $t_n \in \mathbb{R}_+$ is the timestamp.
- Typically, we optimize $\{\alpha_{i \to j}\}$ to assign high likelihood to the observations in some probability model.

Dynamical systems

A latent variable time series model:

$$\boldsymbol{z}_{t,i} \sim f\left(\sum_{j \in \mathcal{I}} \alpha_{j \to i} \boldsymbol{z}_{t-1,j}\right) \tag{7}$$

$$\boldsymbol{y}_{t,i} \sim g(\boldsymbol{z}_{t,i}). \tag{8}$$

Lots of interesting special cases:

- **Kalman filter/smoother**: f and g are affine+Gaussian
- **Generalized Kalman**: f and g are affine+other[35]
- **Dynamic topic model**: g is a topic model, f models changes in topic *proportions*[36] or *content*[37]

---

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

[36] Nguyen et al. 2014.

[37] Gerrish and Blei 2010; Gerow et al. 2018.

Dynamical systems

A latent variable time series model:

$$\boldsymbol{z}_{t,i} \sim f\left(\sum_{j \in \mathcal{I}} \alpha_{j \to i} \boldsymbol{z}_{t-1,j}\right) \tag{7}$$

$$\boldsymbol{y}_{t,i} \sim g(\boldsymbol{z}_{t,i}). \tag{8}$$

Lots of interesting special cases:

- **Kalman filter/smoother**: f and g are affine+Gaussian
- **Generalized Kalman**: f and g are affine+other[35]
- **Dynamic topic model**: g is a topic model, f models changes in topic *proportions*[36] or *content*[37]

---

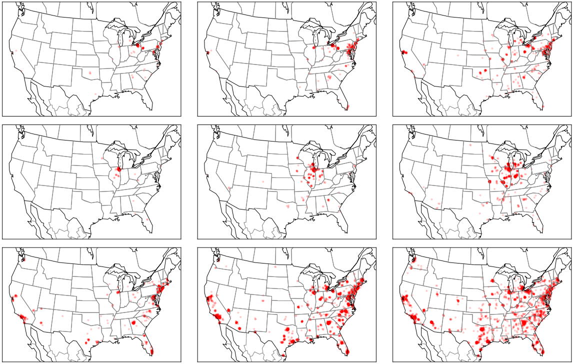
[35] Eisenstein, O'Connor, et al. 2014.

[36] Nguyen et al. 2014.

[37] Gerrish and Blei 2010; Gerow et al. 2018.

City-to-city influence from Twitter data[38]

Thousands of words have changing frequencies.



ctfu

lbvs

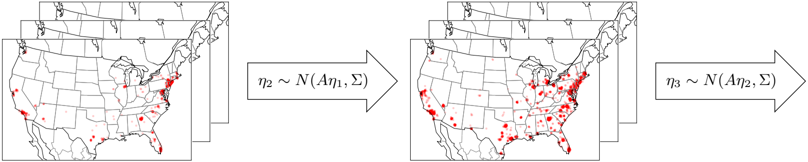
-_-

- Each spatiotemporal trajectory is idiosyncratic.
- What's the aggregate picture?

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

Language change as an autoregressive process

Raw data: counts for thousands of words, binned into 200 metro areas and 165 weeks.

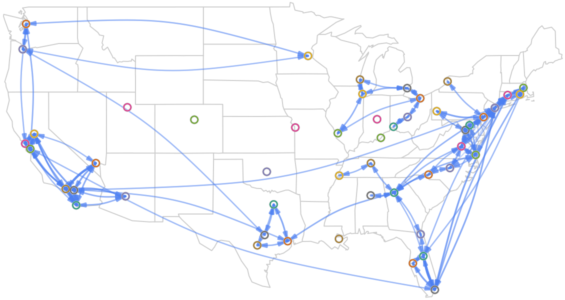


$$\begin{aligned} c_{\text{ctfu},1} &\sim \text{Binomial}(f(\eta_{\text{ctfu},1}), N_1) \\ c_{\text{hella},1} &\sim \text{Binomial}(f(\eta_{\text{hella},1}), N_1) \\ &\ldots \end{aligned}$$

$$\begin{aligned} c_{\text{ctfu},2} &\sim \text{Binomial}(f(\eta_{\text{ctfu},2}), N_2) \\ c_{\text{hella},2} &\sim \text{Binomial}(f(\eta_{\text{hella},2}), N_2) \\ &\ldots \end{aligned}$$

- The dynamics matrix A encodes city-to-city linguistic influence.
- It can be estimated by Monte Carlo Estimation Maximization

Aggregated city-to-city influence



Weights on demographic features can then be estimated by regressing the city-to-city parameters $\alpha_{i \to j}$ on city-level covariates such as population, race, wealth, and age.

Dynamical systems

A latent variable time series model:

$$\boldsymbol{z}_{t,i} \sim f\left(\sum_{j \in \mathcal{I}} \alpha_{j \to i} \boldsymbol{z}_{t-1,j}\right) \tag{7}$$

$$\boldsymbol{y}_{t,i} \sim g(\boldsymbol{z}_{t,i}). \tag{8}$$

Lots of interesting special cases:

- **Kalman filter/smoother**: f and g are affine+Gaussian
- **Generalized Kalman**: f and g are affine+other[35]
- **Dynamic topic model**: g is a topic model, f models changes in topic *proportions*[36] or *content*[37]

---

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

[36] Nguyen et al. 2014.

[37] Gerrish and Blei 2010; Gerow et al. 2018.

Dynamical systems

A latent variable time series model:

$$\boldsymbol{z}_{t,i} \sim f\left(\sum_{j \in \mathcal{I}} \alpha_{j \to i} \boldsymbol{z}_{t-1,j}\right) \tag{7}$$

$$\boldsymbol{y}_{t,i} \sim g(\boldsymbol{z}_{t,i}). \tag{8}$$

Lots of interesting special cases:

- **Kalman filter/smoother**: f and g are affine+Gaussian
- **Generalized Kalman**: f and g are affine+other[35]
- **Dynamic topic model**: g is a topic model, f models changes in topic *proportions*[36] or *content*[37]

---

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

[36] Nguyen et al. 2014.

[37] Gerrish and Blei 2010; Gerow et al. 2018.

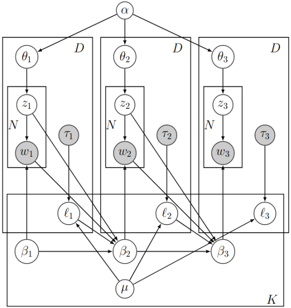
Influence on scientific topics[39]

Each document has:

- a bag-of-words generated from a latent topic model;
- an influence parameter, which is a function of its covariates.

Topics

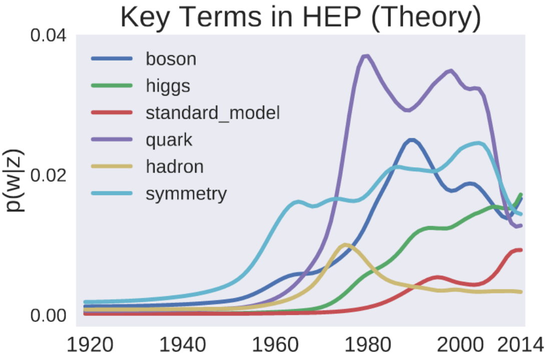
- are smoothly evolving vectors, $\boldsymbol{\beta}_{t,k} \in \mathbb{R}^V$, where
- each $\boldsymbol{\beta}_{t,k}$ depends on $\boldsymbol{\beta}_{t-1,k}$ and the words used in influential documents at $t-1$.



[39] Gerow et al. 2018.

Evolution of a scientific topic[40]

From a topic model of physics research papers:



[40] Gerow et al. 2018.

Attribution of leadership to authors[41]

From a topic model of the ACL anthology:

Parsing & Grammar	Information Retrieval
tree	document
grammar	term
node	query
structure	text
feature	topic
language	retrieval
figure	collection
form	result
constituent	information
set	sentence

Parsing & Grammar		Information Retrieval	
Parameter	**Value in $\hat{\mu}_k$**	**Parameter**	**Value in $\hat{\mu}_k$**
Gerald Gazdar	0.00013	Donna Harman	0.00017
Robert C. Berwick	0.00012	G. Vladutz	0.00014
Monique Rolbert	0.00012	Gerard Salton	0.00013
J. N. Verastegui-Carvajal	0.00011	Jade Goldstein	0.00012
C. Raymond Perrault	0.00001	Chris Buckley	0.00011
Jan Landsbergen	0.00009	David D. Lewis	0.00011
James Kilbury	0.00008	Tomek Strzalkowski	0.00010
Luis Damas	0.00008	K. L. Kwok	0.00009
R. C. Bainbridge	0.00008	John Broglio	0.00009
C. S. Mellish	0.00008	Andy Lauriston	0.00007

[41] Gerow et al. 2018.

Estimating influence from diachronic data

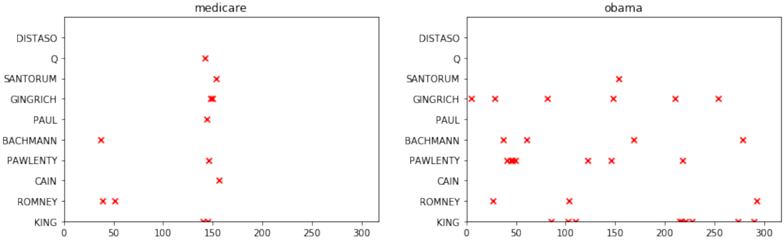
- Goal is to estimate parameters $\alpha_{i \to j}$ for all i, j in population $\mathcal{I}$.
- Given covariates $\{\boldsymbol{x}_i\}_{i \in \mathcal{I}}$, we may set $\alpha_{i \to j} = f(\boldsymbol{x}_i, \boldsymbol{x}_j)$.
- Observed data:
  - **Time series**: counts or proportions $\boldsymbol{y}_{t,i}$ at each $t = (1, 2, \ldots, T)$.
  - **Event cascades**: a set of events $\mathcal{E} = \{(s_n, t_n)\}$, where $s_n \in \mathcal{I}$ is the source and $t_n \in \mathbb{R}_+$ is the timestamp.
- Typically, we optimize $\{\alpha_{i \to j}\}$ to assign high likelihood to the observations in some probability model.

Estimating influence from diachronic data

- Goal is to estimate parameters $\alpha_{i \to j}$ for all i, j in population $\mathcal{I}$.
- Given covariates $\{\boldsymbol{x}_i\}_{i \in \mathcal{I}}$, we may set $\alpha_{i \to j} = f(\boldsymbol{x}_i, \boldsymbol{x}_j)$.
- Observed data:
  - **Time series**: counts or proportions $\boldsymbol{y}_{t,i}$ at each $t = (1, 2, \ldots, T)$.
  - **Event cascades**: a set of events $\mathcal{E} = \{(s_n, t_n)\}$, where $s_n \in \mathcal{I}$ is the source and $t_n \in \mathbb{R}_+$ is the timestamp.
- Typically, we optimize $\{\alpha_{i \to j}\}$ to assign high likelihood to the observations in some probability model.

Event cascade data

From a 2012 Republican Presidential Primary debate:



How can we discern leaders, followers, and resisters from this type of data?

The Poisson Process probability model

- Consider a set of event times, $\{t_n\}_{n=1}^{N}$.
- The number of events in any interval $(t_1, t_2]$ can be modeled as a draw from a Poisson distribution:

$$n(t_1, t_2) \sim \text{Poisson}(\mu_{t_1,t_2}) \tag{9}$$

$$\mu_{t_1,t_2} = \int_{t_1}^{t_2} \lambda(\tau) d\tau, \tag{10}$$

  where $\lambda(\tau)$ is an **intensity function**.
- This is a very general framework for modeling sequences of events. Let's explore a special case that is appropriate for our research questions.[42]

[42] For more on temporal point processes, see the 2018 ICML tutorial by Gomez-Rodriguez and Valera (2018).

The Hawkes Process

A Hawkes Process is a **self-exciting** Poisson process,[43] in which the intensity function depends on the previous events,

$$\lambda(\tau) = \lambda_0 + \alpha \sum_{t_n < \tau} \kappa(\tau - t_n), \tag{11}$$

where,

- the sum is over all events before time τ;
- $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ is a kernel decay function, typically monotonically decreasing in the temporal gap $\tau - t_n$, e.g. $\kappa(\delta t) = \beta \exp(-\delta t / \gamma)$.[44]
- $\alpha \in \mathbb{R}$ is a scalar excitement parameter.

---

[43] Hawkes 1971.

[44] This function can be much more expressive! (Du et al. 2015; Mei and Eisner 2017)

Multivariate Hawkes Processes

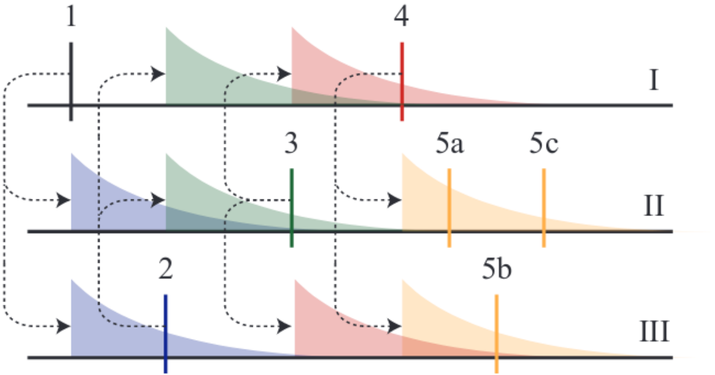
In a multivariate Hawkes Process, each event is marked with a **source** s_n, and each source exerts a different level of excitement on its neighbors:

$$\lambda_i(\tau) = \lambda_{i,0} + \sum_{t_n < \tau} \alpha_{s_n \rightarrow i} \kappa(\tau - t_n), \tag{12}$$

where

- $\lambda_i(\tau)$ is the excitement function for individual i;
- $\alpha_{s_n \rightarrow i}$ quantifies how much events from s_n tend to excite events in i.

A multivariate Hawkes Process[45]



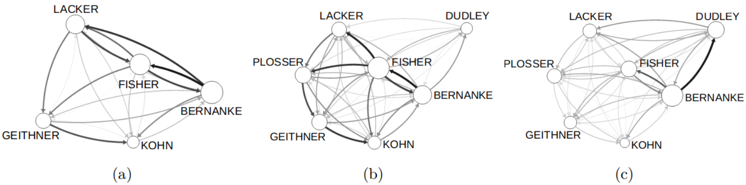
[45] Linderman and Adams 2014.

Notebook time!

Notebook: leaders and followers in the 2012 Republican Presidential Primary debates.

```
jupyter-notebook FollowTheLeader.ipynb
```

Point process models of language change[49]



- Accommodation and influence in multiparty dialogues[46]
- Spread of new words over the Twitter social network[47]
- Variable temporal kernels for different types of events in social media[48]

---

[46] Blundell, Beck, and Heller 2012; Guo et al. 2015.

[47] Goel et al. 2016.

[48] Du et al. 2015.

[49] Figure from Guo et al. (2015)

Outline

- Motivations and perspectives on language change
- Practical methods and case studies

1. Word frequency	Are you happier on the weekend?
2. Differences	Is polarization on the rise?
3. Word meaning	When did peers apply pressure?
4. **Leaders and followers**	Who is shaping the debate?
5. Cause and effect	Should we feed the trolls?

- Next steps

Outline

- Motivations and perspectives on language change
- Practical methods and case studies

1.	Word frequency	Are you happier on the weekend?
2.	Differences	Is polarization on the rise?
3.	Word meaning	When did peers apply pressure?
4.	Leaders and followers	Who is shaping the debate?
5.	**Cause and effect**	Should we feed the trolls?

- Next steps

Explanations and causes

- Did the restaurant close because of the bad reviews?
- Did the DARPA program cause people to change their research agendas?[50]
- Did the introduction of emojis cause people to use fewer emoticons?[51]
- Did the lobbying group's "model legislation" cause the bill to be written differently?[52]
- Did censorship cause people to change spelling?[53]

---

[50] Gerow et al. 2018.

[51] Pavalanathan and Eisenstein 2016.

[52] Burgess et al. 2016.

[53] Wang, Juffermans, and Du 2016; Stewart, Chancellor, et al. 2017.

The Randomized Controlled Trial (RCT)

The "gold standard" of causal inference.[54]

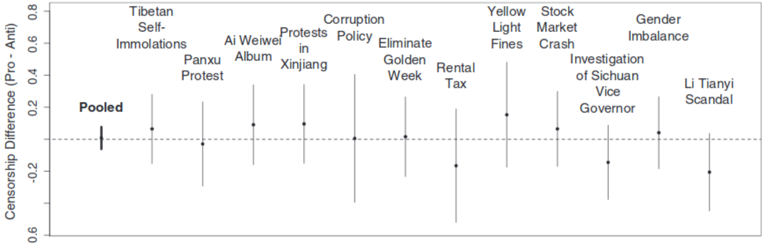
- Does the **treatment** affect the **outcome**?
  The challenge is that each unit (e.g., patient) is either treated or not; we do not observe the counterfactual.
- Solution: ensure treatment and **control** groups are similar on average, by assigning the treatment randomly.[55]
- **Average Treatment Effect**: difference in average outcome between the treatment and control groups.

[54] Rosenbaum 2017.

[55] Further assumptions: treatment of one unit does not affect outcome for another; all treatments are identical.

Example: Censorship in Chinese social media

- **Treatment**: whether the post is pro- or anti-government
- **Outcome**: whether the post is censored
- Many potential **confounds**: previous history of the author, their social network, the platform...
- Therefore King, Pan, and Roberts (2014) randomly assigned posts to (fake) user accounts.



Problems with RCTs

- **Impossible**: we cannot randomly assign legislation to be treated with lobbying.
- **Expensive**: King, Pan, and Roberts (2014) had to hire a team to write 1200 unique social media posts.
- **Unethical**:[56] it would be wrong to randomly assign bad reviews to restaurants.
- **Not representative**: RCTs typically rely on participants who are often drawn from pools that are notoriously different from the larger population.

[56] see also Kramer, Guillory, and Hancock 2014; Verma 2014.

Causation from observation

Impossible in general, but acceptable under some conditions.
A non-exhaustive list of approaches:

- **Matching**: try to enumerate all covariates, select treatment and control groups to be balanced.[57]
- **Natural experiments**: identify some exogeneous factor that affects probability of treatment but is assigned at random (e.g., the weather).[58]
- **Interrupted time series**: compare outcome before and after the treatment, correcting for temporal trends.[59]

---

[57] e.g., De Choudhury et al. 2016; Pavalanathan and Eisenstein 2016.

[58] e.g., Coviello et al. 2014.

[59] Bernal, Cummins, and Gasparrini 2017.

Causation from observation

Impossible in general, but acceptable under some conditions.
A non-exhaustive list of approaches:

- **Matching**: try to enumerate all covariates, select treatment and control groups to be balanced.[57]
- **Natural experiments**: identify some exogeneous factor that affects probability of treatment but is assigned at random (e.g., the weather).[58]
- **Interrupted time series**: compare outcome before and after the treatment, correcting for temporal trends.[59]

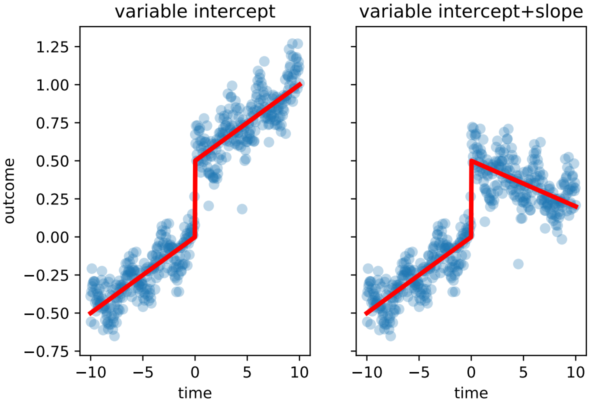
---

[57] e.g., De Choudhury et al. 2016; Pavalanathan and Eisenstein 2016.

[58] e.g., Coviello et al. 2014.

[59] Bernal, Cummins, and Gasparrini 2017.

Interrupted time series schematics



Variable intercept model as segmented regression

$$y_{t,i} \sim f(\boldsymbol{\beta} \cdot \boldsymbol{x}_{t,i} + \gamma\delta(t > \tau)), \tag{13}$$

where,

- f is a probability function;
- $y_{t,i}$ is the outcome for individual i at time t;
- $\boldsymbol{x}_{t,i}$ is a vector of covariates, including t;
- $\boldsymbol{\beta}$ is a vector of parameters;
- $\delta(t > \tau)$ indicates if t is in the post-treatment period;
- γ quantifies the causal impact of the treatment on the intercept.

Variable intercept+slope model

$$y_{t,i} \sim f(\boldsymbol{\beta} \cdot \boldsymbol{x}_{t,i} + \gamma\delta(t > \tau) + \omega\delta(t > \tau)t), \qquad (14)$$

where,

- ω quantifies the causal impact of the treatment on the slope;
- all other terms are the same as in the variable intercept model.

Assumptions

If $\gamma < 0$, can we conclude that the treatment causes the outcome to decrease? Some assumptions are needed:

- **Ignorability**: there is no latent covariate that is associated with both the treatment and the outcome.

  For example, suppose Reddit's decision to censor hate speech coincided with a widespread internet outage. This concern can be alleviated by adding a control group.

- **Correct specification**: the functional form of the regression model is correct.
- **No autocorrelation**: $y_{t,i}$ depends only on the covariates and treatment, and not on $y_{t'<t,i}$. Autocorrelation can usually be tested for and removed.

Replication time!

Replication: does banning hate speech forums reduce the overall amount of hate speech on Reddit?[60]

```
jupyter-notebook DontFeedTheTrolls.ipynb
```

[60] Chandrasekharan et al. 2018.

Granger causation

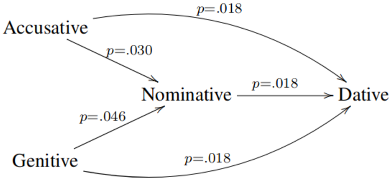
Granger causation is assessed by comparing two predictors of each element in a time series y_t:

1. the previous history $\boldsymbol{y}_{1:t-1}$;
2. the previous histories $\boldsymbol{x}_{1:t-1}, \boldsymbol{y}_{1:t-1}$.

- The time series $\boldsymbol{x}$ **Granger-causes** $\boldsymbol{y}$ if prediction 2 is significantly more accurate.
- Warning: it is possible to have Granger causation without "real" causation.

Granger causation of function shifts in case

Prototype embeddings of case groups in Icelandic predict changes in the embeddings of other case groups:[61]



This finding, which is based on an annotated corpus of nearly 1000 years of Icelandic, suggests that while the Icelandic case paradigms have not changed, their **functions** have undergone a gradual chain shift.

[61] Moscoso del Prado Martin and Brendel 2016.

Text as cause, text as consequence

Many possible roles for text in causal inference[62]:

- outcome (e.g., hate speech in Reddit);
- treatment (e.g., pro- or anti-government stance in Chinese social media);[63]
- deconfounding;[64]
- correcting for missing data and measurement error.[65]

Watch this space!

---

[62] Egami et al. 2018.

[63] see also Landeiro and Culotta 2018.

[64] Roberts, Stewart, and Nielsen 2018; Mozer et al. 2018; Veitch, Sridhar, and Blei 2019.

[65] Wood-Doughty, Shpitser, and Dredze 2018.

Outline

- Motivations and perspectives on language change
- Practical methods and case studies

1. Word frequency	Are you happier on the weekend?
2. Differences	Is polarization on the rise?
3. Word meaning	When did peers apply pressure?
4. Leaders and followers	Who is shaping the debate?
5. **Cause and effect**	Should we feed the trolls?

- Next steps

Outline

- Motivations and perspectives on language change
- Practical methods and case studies

1. Word frequency	Are you happier on the weekend?
2. Differences	Is polarization on the rise?
3. Word meaning	When did peers apply pressure?
4. Leaders and followers	Who is shaping the debate?
5. Cause and effect	Should we feed the trolls?

- **Next steps**

Broadening the scope

- Beyond English[66]
- Beyond the lexicon
- New collaborative models

[66] Hamilton, Leskovec, and Jurafsky 2016b; Moscoso del Prado Martin and Brendel 2016; Garley and Hockenmaier 2012.

Contrasting lexical and phonological change

Phonological change:

- measured by decades and centuries;
- (usually thought of as) acquired from peers and caretakers during early childhood;
- tightly interconnected system.

Lexical change:

- usually measured by months and years;
- acquired socially into adulthood;
- structural aspects are less clear.[67]

[67] but see Pierrehumbert 2010; Stewart and Eisenstein 2018.

Contrasting lexical and phonological change

Phonological change:

- measured by decades and centuries;
- (usually thought of as) acquired from peers and caretakers during early childhood;
- tightly interconnected system.

Lexical change:

- usually measured by months and years;
- acquired socially into adulthood;
- structural aspects are less clear.[67]

Is this one mechanism or two? What can students of these two phenomena teach each other?

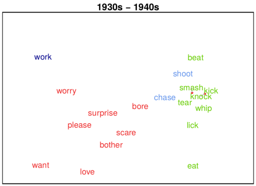
[67] but see Pierrehumbert 2010; Stewart and Eisenstein 2018.

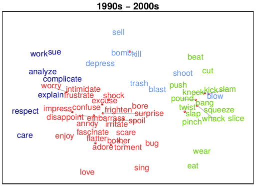
Challenges for future work

1. Elaborate the connection between lexical change and other forms of systemic language change.

Measuring syntactic change: specific constructions

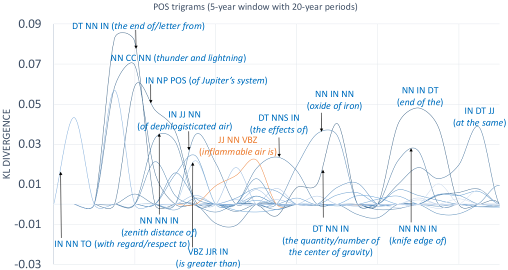
- Perek (2014):
  V the hell out of NP
  - snakes scare the hell out of me
  - you drove the hell out of it
- Increasing productivity throughout the 20th century, with new cases tending to occur in denser parts of embedding space.
- Identifying this construction requires part-of-speech tagging and lemmatization (Schmid 1995).





Measuring syntactic change: POS trigrams

Degaetano-Ortlieb and Teich (2018): KL-divergence of POS trigrams decreases over time in scientific writing.

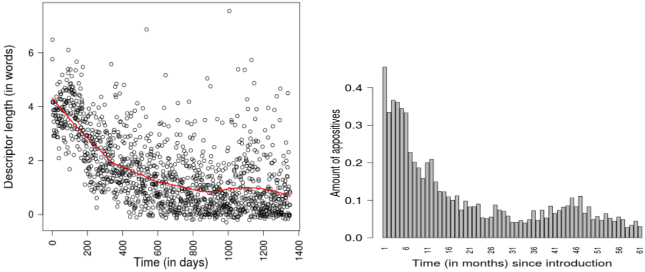


- "Grammatical consolidation" follows lexical expansion.
- Again, automated POS tagging required.[68]

[68] Schmid 1995; Baron and Rayson 2008.

Forms of referring expressions over entity lifespans

Staliūnaitė et al. (2018): referring expressions reflect information status of entities.



Uses CoreNLP for constituency parsing and coreference.[69]

[69] Manning et al. 2014.

How good is syntactic analysis on historical text?

We can assume some transfer loss when moving trained systems to historical text, but how much?

- There is not much evidence about this, as almost all labeled data is from the 20th century.
- **English**: Penn Parsed Corpus of Historical English: great resource, but idiosyncratic tagset and purchasing model.
- **Portuguese**: Tycho Brahe corpus uses a single annotation system over several centuries.[70]

Yang and Eisenstein (2015):
POS error rate doubles for older texts.

The concern: are syntactic effects real, or are they an artifact of changes in tagger performance?

---

[70] Galves and Faria 2010.

Challenges for future work

1. Elaborate the connection between lexical change and other forms of systemic language change.

Challenges for future work

1. Elaborate the connection between lexical change and other forms of systemic language change.
2. Make NLP systems more robust to language change, without labeling thousands of documents.

Construct validity challenges

Construct validity is not just a problem for syntax!

- We want to use diachronic corpora to make claims, e.g., about changing ideas, beliefs, agendas, etc.
- Some of these constructs will be visible to the tools we have (which are mostly lexical), some won't.
- In many historical corpora, digitization creates its own measurement artifacts.[71]
- Extant statistical methodology is designed to quantify uncertainty due to finite samples, not variable measurement error.

[71] e.g., Soni, Klein, and Eisenstein 2019, at this year's LaTeCH.

Challenges for future work

1. Elaborate the connection between lexical change and other forms of systemic language change.
2. Make NLP systems more robust to language change, without labeling thousands of documents.

Challenges for future work

1. Elaborate the connection between lexical change and other forms of systemic language change.
2. Make NLP systems more robust to language change, without labeling thousands of documents.
3. Better account for errors, especially when those errors are unevenly distributed over time.

Toward new collaborative models

- Many text mining tools are becoming "commoditized."
- There are good reasons for practitioners to prefer standard tools (like LDA and LIWC) over bespoke solutions.
- Yet there remains a need for computationally-augmented access to large-scale corpora **in human terms**:

lexicons, POS patterns, dependency subtrees, . . .

Challenges for future work

1. Elaborate the connection between lexical change and other forms of systemic language change.
2. Make NLP systems more robust to language change, without labeling thousands of documents.
3. Better account for errors, especially when those errors are unevenly distributed over time.

Challenges for future work

1. Elaborate the connection between lexical change and other forms of systemic language change.
2. Make NLP systems more robust to language change, without labeling thousands of documents.
3. Better account for errors, especially when those errors are unevenly distributed over time.
4. Allow non-specialists to tell NLP systems what they want without labeling thousands of examples.

Thank you!

Thank you!

- **Students** and **Collaborators**: **Sandeep Soni and Ian Stewart**, Eshwar Chandrasekharan, Munmun De Choudhury, Fernando Diaz, Lauren F. Klein, Eric Gilbert, Rahul Goel, Naman Goyal, Adam Glynn, Xiaochuang Han, Brendan O'Connor, John Paparrizos, Umashanthi Pavalanathan, Noah A. Smith, Hanna Wallach, Eric Xing, Yi Yang.
- **Sponsors**: National Endowment for the Humanities, National Science Foundation, National Institutes for Health, Air Force Office of Scientific Research, Google.

References I

Antoniak, Maria and David Mimno (2018). "Evaluating the stability of embedding-based word similarities". In: *Transactions of the Association of Computational Linguistics* 6, pp. 107–119.

Bamler, Robert and Stephan Mandt (2017). "Dynamic Word Embeddings". In: *Proceedings of the International Conference on Machine Learning (ICML)*.

Bantum, Erin O'Carroll and Jason E Owen (2009). "Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narratives.". In: *Psychological assessment* 21.1, p. 79.

Baron, Alistair and Paul Rayson (2008). "VARD2: A tool for dealing with spelling variation in historical corpora". In: *Postgraduate conference in corpus linguistics*.

Barron, Alexander TJ et al. (2018). "Individuals, institutions, and innovation in the debates of the French Revolution". In: *Proceedings of the National Academy of Sciences* 115.18, pp. 4607–4612.

Bernal, James Lopez, Steven Cummins, and Antonio Gasparrini (2017). "Interrupted time series regression for the evaluation of public health interventions: a tutorial". In: *International journal of epidemiology* 46.1, pp. 348–355.

References II

Blundell, Charles, Jeff Beck, and Katherine A Heller (2012). "Modelling Reciprocating Relationships with Hawkes Processes". In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira et al. Curran Associates, Inc., pp. 2600–2608.

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.

Brown, Peter F et al. (1992). "Class-based n-gram models of natural language". In: *Computational linguistics* 18.4, pp. 467–479.

Burgess, Matthew et al. (2016). "The Legislative Influence Detector: Finding Text Reuse in State Legislation". In: *Proceedings of Knowledge Discovery and Data Mining (KDD)*, pp. 57–66.

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)*.

Coviello, Lorenzo et al. (2014). "Detecting emotional contagion in massive social networks". In: *PloS one* 9.3, e90315.

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

References III

Danescu-Niculescu-Mizil, Cristian and Lillian Lee (2011). "Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs.". In: *Proceedings of the ACL Workshop on Cognitive Modeling and Computational Linguistics*.

Danescu-Niculescu-Mizil, Cristian, Robert West, et al. (2013). "No country for old members: User lifecycle and linguistic change in online communities". In: *Proceedings of the Conference on World-Wide Web (WWW)*, pp. 307–318.

Davies, Mark (2012). "Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English". In: *Corpora* 7.2, pp. 121–157.

Davies, Mark and Robert Fuchs (2015). "Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE)". In: *English World-Wide* 36.1, pp. 1–28.

De Choudhury, Munmun et al. (2016). "Discovering shifts to suicidal ideation from mental health content in social media". In: *Proceedings of Human Factors in Computing Systems (CHI)*, pp. 2098–2110.

Deerwester, Scott C. et al. (1990). "Indexing by latent semantic analysis". In: *Journal of the American society for information science* 41.6, pp. 391–407.

Degaetano-Ortlieb, Stefania and Elke Teich (2018). "Using relative entropy for detection and analysis of periods of diachronic linguistic change". In: *Proceedings of LaTeCH*.

References IV

- Del Tredici, Marco and Raquel Fernández (2018). "The Road to Success: Assessing the Fate of Linguistic Innovations in Online Communities". In: *Proceedings of the International Conference on Computational Linguistics (COLING)*, pp. 1591–1603.
- Demszky, Dorottya et al. (2019). "Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings". In: *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*.
- Doyle, Gabriel and Michael C Frank (2015). "Shared common ground influences information density in microblog texts". In: *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*.
- Du, Nan et al. (2015). "Dirichlet-hawkes processes with applications to clustering continuous-time document streams". In: *Proceedings of Knowledge Discovery and Data Mining (KDD)*, pp. 219–228.
- Dubossarsky, Haim, Daphna Weinshall, and Eitan Grossman (2017). "Outta Control: Laws of Semantic Change and Inherent Biases in Word Representation Models". In: *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, pp. 1136–1145.
- Egami, Naoki et al. (2018). "How to make causal inferences using texts". In: *arXiv preprint arXiv:1802.02163*.

References V

Eger, Steffen and Alexander Mehler (2016). "On the Linearity of Semantic Change: Investigating Meaning Variation via Dynamic Graph Models". In: *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 52–58.

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.

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.

Galves, Charlotte and Pablo Faria (2010). *Tycho Brahe Parsed Corpus of Historical Portuguese*. http://www.tycho.iel.unicamp.br/~tycho/corpus/en/index.html.

Garg, Nikhil et al. (2018). "Word embeddings quantify 100 years of gender and ethnic stereotypes". In: *Proceedings of the National Academy of Sciences* 115.16, E3635–E3644.

References VI

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.

Gentzkow, Matthew, Jesse Shapiro, and Matt Taddy (2016). *Measuring polarization in high-dimensional data: Method and application to congressional speech*. Tech. rep. 22423. NBER Working Papers.

Gerow, Aaron et al. (2018). "Measuring discursive influence across scholarship". In: *Proceedings of the National Academy of Sciences*, p. 201719792.

Gerrish, Sean and David M Blei (2010). "A Language-based Approach to Measuring Scholarly Impact". In: *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 375–382.

Goel, Rahul et al. (Nov. 2016). "The Social Dynamics of Language Change in Online Networks". In: *The International Conference on Social Informatics (SocInfo)*.

Golder, Scott A and Michael W Macy (2011). "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures". In: *Science* 333.6051, pp. 1878–1881.

Gomez-Rodriguez, Manuel and Isabel Valera (2018). ICML tutorial, `http://learning.mpi-sws.org/tpp-icml18/`.

References VII

Gulordava, Kristina and Marco Baroni (July 2011). "A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus.". In: *Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics*. Edinburgh, UK: Association for Computational Linguistics, pp. 67–71.

Guo, Fangjian et al. (2015). "The Bayesian Echo Chamber: Modeling social influence via linguistic accommodation". In: *Proceedings of Artificial Intelligence and Statistics (AISTATS)*.

Hall, David, Daniel Jurafsky, and Christopher D Manning (2008). "Studying the history of ideas using topic models". In: *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, pp. 363–371.

Hamilton, William L (2016).

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

References VIII

Hawkes, Alan G (1971). "Spectra of some self-exciting and mutually exciting point processes". In: *Biometrika* 58.1, pp. 83–90.

Hofman, Jake M, Amit Sharma, and Duncan J Watts (2017). "Prediction and explanation in social systems". In: *Science* 355.6324, pp. 486–488.

Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes (2012). "Affect, not ideologya social identity perspective on polarization". In: *Public opinion quarterly* 76.3, pp. 405–431.

Jurgens, David and Keith Stevens (Sept. 2009). "Event Detection in Blogs using Temporal Random Indexing". In: *Proceedings of the Workshop on Events in Emerging Text Types*. Borovets, Bulgaria: Association for Computational Linguistics, pp. 9–16.

King, Gary, Jennifer Pan, and Margaret E Roberts (2014). "Reverse-engineering censorship in China: Randomized experimentation and participant observation". In: *Science* 345.6199, p. 1251722.

Kooti, Farshad et al. (2012). "The Emergence of Conventions in Online Social Networks". In: *Proceedings of the International Conference on Web and Social Media (ICWSM)*, pp. 194–201.

Kramer, Adam DI, Jamie E Guillory, and Jeffrey T Hancock (2014). "Experimental evidence of massive-scale emotional contagion through social networks". In: *Proceedings of the National Academy of Sciences* 111.24, pp. 8788–8790.

References IX

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

Kutuzov, Andrey et al. (2018). "Diachronic word embeddings and semantic shifts: a survey". In: *Proceedings of the International Conference on Computational Linguistics (COLING)*, pp. 1384–1397.

Labov, William (2001). *Principles of Linguistic Change*. Vol. 2: Social Factors. Wiley-Blackwell.

Landeiro, Virgile and Aron Culotta (2018). "Robust Text Classification under Confounding Shift". In: *Journal of Artificial Intelligence Research* 63, pp. 391–419.

Lansdall-Welfare, Thomas et al. (2017). "Content analysis of 150 years of British periodicals". In: *Proceedings of the National Academy of Sciences* 114.4, E457–E465.

Leech, Geoffrey and Paul Rayson (2014). *Word frequencies in written and spoken English: Based on the British National Corpus*. Routledge.

Leskovec, Jure, Lars Backstrom, and Jon Kleinberg (2009). "Meme-tracking and the dynamics of the news cycle". In: *Proceedings of Knowledge Discovery and Data Mining (KDD)*, pp. 497–506.

Liberman, Mark (2010). *More on culturnomics*. accessed May 2019.

References X

Linderman, Scott and Ryan Adams (2014). "Discovering latent network structure in point process data". In: *International Conference on Machine Learning*, pp. 1413–1421.

Manning, Christopher D. et al. (2014). "The Stanford CoreNLP Natural Language Processing Toolkit". In: *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55–60.

Mei, Hongyuan and Jason M Eisner (2017). "The neural hawkes process: A neurally self-modulating multivariate point process". In: *Neural Information Processing Systems (NIPS)*, pp. 6754–6764.

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

Mihalcea, Rada and Vivi Nastase (2012). "Word epoch disambiguation: Finding how words change over time". In: *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 259–263.

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

Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig (2013). "Linguistic Regularities in Continuous Space Word Representations". In: *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pp. 746–751.

References XI

Milroy, Lesley (1991). *Language and Social Networks*. 2nd ed. Wiley-Blackwell.

Monroe, Burt L., Michael P. Colaresi, and Kevin M. Quinn (2008). "Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict". In: *Political Analysis* 16.4, pp. 372–403.

Moscoso del Prado Martin, Fermin and Christian Brendel (2016). "Case and Cause in Icelandic: Reconstructing Causal Networks of Cascaded Language Changes". In: *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 2421–2430.

Mozer, Reagan et al. (2018). "Matching with Text Data: An Experimental Evaluation of Methods for Matching Documents and of Measuring Match Quality". In: *arXiv preprint arXiv:1801.00644*.

Nguyen, Viet-An et al. (2014). "Modeling topic control to detect influence in conversations using nonparametric topic models". In: *Machine Learning* 95.3, pp. 381–421.

Pavalanathan, Umashanthi and Jacob Eisenstein (Nov. 2016). "More emojis, less :) The Competition for Paralinguistic Functions in Microblog Writing". In: *First Monday* 22.11.

Pechenick, Eitan Adam, Christopher M Danforth, and Peter Sheridan Dodds (2015). "Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution". In: *PloS one* 10.10, e0137041.

References XII

Perek, Florent (2014). "Vector spaces for historical linguistics: Using distributional semantics to study syntactic productivity in diachrony". In: *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 309–314.

Peterson, Andrew and Arthur Spirling (2018). "Classification accuracy as a substantive quantity of interest: Measuring polarization in Westminster systems". In: *Political Analysis* 26.1, pp. 120–128.

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.

Popescu, Octavian and Carlo Strapparava (2015). "Semeval 2015, task 7: Diachronic text evaluation". In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pp. 870–878.

Roberts, Margaret E, Brandon M Stewart, and Richard A Nielsen (2018). *Adjusting for confounding with text matching*.

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

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.

References XIII

Rudolph, Maja and David Blei (2018). “Dynamic Embeddings for Language Evolution”. In: *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, pp. 1003–1011.

Schmid, Helmut (1995). “Improvements in Part-of-Speech Tagging with an Application to German”. In: *Proceedings of SIGDAT*.

Shmueli, Galit (2010). “To explain or to predict?” In: *Statistical science* 25.3, pp. 289–310.

Soni, Sandeep, Lauren F. Klein, and Jacob Eisenstein (2019). “Correcting Whitespace Errors in Digitized Historical Texts”. In: *Proceedings of LaTECH*.

Staliūnaitė, Ieva et al. (2018). “Getting to “Hearer-old”: Charting Referring Expressions Across Time”. In: *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*.

Stewart, Ian, Stevie Chancellor, et al. (2017). “#anorexia, #anarexia, #anarexyia: Characterizing Online Community Practices with Orthographic Variation”. In: *Proceedings of IEEE Big Data: Workshop on Natural Language Processing for Social Media (SocialNLP)*.

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)*.

References XIV

Taddy, Matt (2013). "Multinomial inverse regression for text analysis". In: *Journal of the American Statistical Association* 108.503, pp. 755–770.

Tahmasebi, Nina, Lars Borin, and Adam Jatowt (2018). "Survey of Computational Approaches to Diachronic Conceptual Change". In: *arXiv preprint arXiv:1811.06278*.

Tang, Xuri (2018). "A state-of-the-art of semantic change computation". In: *Natural Language Engineering* 24.5, pp. 649–676.

Tausczik, Yla R and James W Pennebaker (2010). "The psychological meaning of words: LIWC and computerized text analysis methods". In: *Journal of Language and Social Psychology* 29.1, pp. 24–54.

Traugott, Elizabeth Closs and Richard B Dasher (2001). *Regularity in semantic change*. Cambridge University Press.

Tsur, Oren and Ari Rappoport (2015). "Don't Let Me Be# Misunderstood: Linguistically Motivated Algorithm for Predicting the Popularity of Textual Memes.". In: *ICWSM*, p. 426.

Underwood, Ted, David Bamman, and Sabrina Lee (2018). "The Transformation of Gender in English-Language Fiction". In: *Cultural Analytics*.

Veitch, Victor, Dhanya Sridhar, and David M. Blei (2019). *Using Text Embeddings for Causal Inference*. https://arxiv.org/abs/1905.12741.

References XV

Verma, Inder M. (2014). "Editorial Expression of Concern: Experimental evidence of massivescale emotional contagion through social networks". In: *Proceedings of the National Academy of Sciences* 111.29, pp. 10779–10779. eprint: https://www.pnas.org/content/111/29/10779.1.full.pdf.

Wang, Xuan, Kasper Juffermans, and Caixia Du (2016). "Harmony as language policy in China: An Internet perspective". In: *Language Policy* 15.3, pp. 299–321.

Weinreich, Uriel, William Labov, and Marvin Herzog (1968). "Empirical foundations for a theory of language change". In: *Directions for historical linguistics*. Ed. by W. P. Lehmann and Y. Malkiel. University of Texas Press, pp. 97–188.

Wendlandt, Laura, Jonathan K. Kummerfeld, and Rada Mihalcea (2018). "Factors Influencing the Surprising Instability of Word Embeddings". In: *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pp. 2092–2102.

Wijaya, Derry Tanti and Reyyan Yeniterzi (2011). "Understanding semantic change of words over centuries". In: *Proceedings of the 2011 international workshop on DETecting and Exploiting Cultural diversiTy on the social web*. ACM, pp. 35–40.

Wood-Doughty, Zach, Ilya Shpitser, and Mark Dredze (2018). "Challenges of Using Text Classifiers for Causal Inference". In: *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, pp. 4586–4598.

References XVI

Yang, Yi and Jacob Eisenstein (2015). "Unsupervised Multi-Domain Adaptation with Feature Embeddings". In: *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL).*