Modeling the Social Dynamics Using Language Change

Sandeep Soni

School of Interactive Computing

Thesis Proposal
May 06, 2020

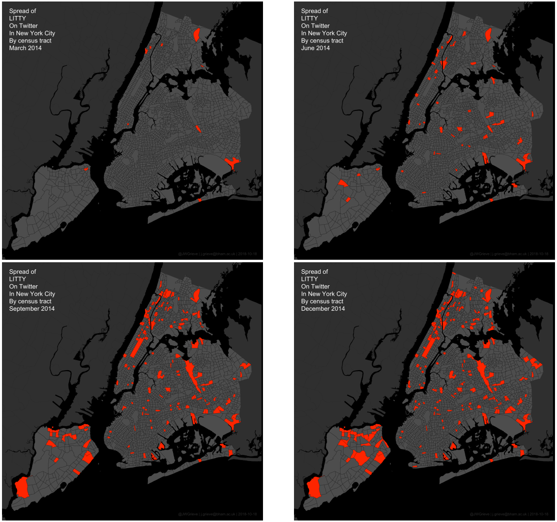
Language *changes* over time

15c English	Our Lord gouerneth me, and nothyng shal defailen to me
19c English	The Lord is my shepherd, I lack nothing.

Language *changes* over time

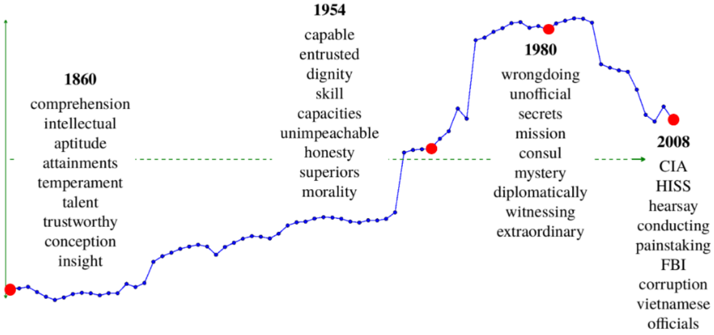
15c English	Our Lord gouerneth me, and nothyng shal defailen to me
19c English	The Lord is my shepherd, I lack nothing.
Today ?	Gawd, he's ma dude, I dun need nuffin.

Lexical change



Diffusion of litty in NY; figure from Grieve (2018)

Semantic change



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

Figure from Rudolph and Blei (2018)

Language change is social

- Language change is intimately linked with social factors such as influence and identity.
- This offers a chance to study how language evolves but also to infer influence relationships or unpack the social meaning of attributes such as race, gender, age, *etc.*

Diachronic text data

1881 On Matter as a form of Energy
1892 Non-Euclidean Geometry
1900 On Kathode Rays and Some Related Phenomena
1917 ``Keep Your Eye on the Ball''
1920 The Arrangement of Atoms in Some Common Metals
1933 Studies in Nuclear Physics
1943 Aristotle, Newton, Einstein. II
1950 Instrumentation for Radioactivity
1965 Lasers
1975 Particle Physics: Evidence for Magnetic Monopole Obtained
1985 Fermilab Tests its Antiproton Factory
1999 Quantum Computing with Electrons Floating on Liquid Helium

- Diachronic text data can be used to track what is changing in a language.[1]
- Computational methods can consume large amounts of diachronic text and additional metadata to reconstruct the latent social structure.

[1] Many works have used diachronic text to gain insights, e.g., Hall et al. 2008; Danescu-Niculescu-Mizil et al. 2013; Eisenstein et al. 2014; Garg et al. 2018; Gerow et al. 2018.

Thesis statement

The computational modeling of diachronic text can reveal social dynamical patterns of influence and identity.

Studies

The thesis comprises five studies:[2]

Peer influence	
1. word adoption	Which network ties matter in the spread of words?
2. influence detection	When can social influence be detected in event cascades?
Semantic lead	
3. semantic progressiveness	Which documents are semantically progressive?
4. semantic leadership	Which newspapers lead or lag a specific semantic change?
Identity signaling	
5. word abandonment	Why are words abandoned?

[2] blue denotes completed work and red denotes proposed work

Thesis contributions

- **Models**

  scalable point process models, discriminative model to detect social influence in event cascades, extensions to diachronic word embedding models

- **Metrics**

  semantic progressiveness of documents, semantic leadership between a pair of sources

- **Linguistic insights**

  language change is a social contagion, linguistic leaders get more citations, linguistic abandonment because of mainstream adoption

Studies

The thesis comprises five studies:[2]

Peer influence

1. **word adoption** — Which network ties matter in the spread of words?
2. influence detection — When can social influence be detected in event cascades?

Semantic lead

3. **semantic progressiveness** — Which documents are semantically progressive?
4. semantic leadership — Which newspapers lead or lag a specific semantic change?

Identity signaling

5. word abandonment — Why are words abandoned?

[2] blue denotes completed work and red denotes proposed work

The social dynamics of language change in online networks[3]

[3]Rahul Goel, Sandeep Soni, Naman Goyal, John Paparrizos, Hanna Wallach, Fernando Diaz, and Jacob Eisenstein (2016). "The social dynamics of language change in online networks". In: *International Conference on Social Informatics*. Springer, pp. 41–57

Hypotheses

This work uses large scale Twitter data to test the following hypotheses:

- H1: Language change spreads through online social network connections.
- **H2:** Densely embedded ties are more linguistically influential.
- **H3:** Geographically local ties are more linguistically influential.

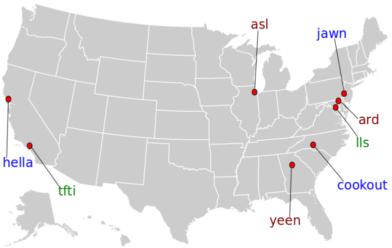
Data

Key characteristics of the dataset

- **Period**: All public messages from US between June 2013 and June 2014.
- **Scale**: A total of 4.35 million twitter users with geolocation and social network metadata.
- **Linguistic variables**: 16 non-standard words divided into 3 linguistic categories.

Data: linguistic variables and geography

- These words are known to be strongly associated with the selected areas.



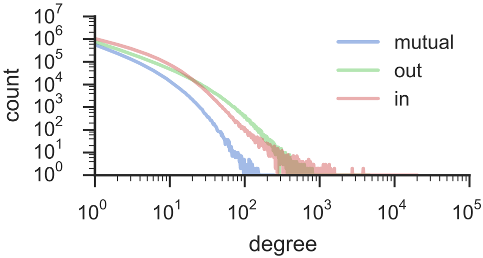
For example:

- ard : phonetic spelling for all right.
- hella: lexical word typically means extremely.
- lls: abbreviation for laughing like shit.

Data: social network

A social network connection exists between two users if they have mentioned each other.

- mary: @mark going to the game?
- mark: @mary yea



The mentions network is more easily available and is socially meaningful than follower-followee network.[4]

[4]Huberman et al. 2008.

Dataset summary

Social network

Mark	Mary
Mark	Mickey
Mary	Todd
Todd	Barney
...	...

Locations

Mark	Los Angeles
Mickey	Los Angeles
Mary	Atlanta
Todd	Chicago
...	...

Language

Mark	jawn	Jun 1, 2013, 13:45
Mickey	jawn	Jun 1, 2013, 13:50
Todd	hella	Jun 1, 2013, 18:15
Mark	lls	Jun 2, 2013, 07:30
Mickey	lls	Jun 2, 2013, 07:40
...	...	...

Multivariate Hawkes process[5]

- For every word, a cascade of events about when the word was used and who used it is $\{t_n, m_n\}_{n \in 1 \ldots N}$.
- Every author has an intensity function to define how likely it is to use a word at any time.

$$\lambda_m(t) = \mu_m + \sum_{t \leq t_n} \alpha_{m_n \to m} \kappa(t - t_n), \qquad (1)$$

  - μ_m is the base rate of user m ($\mu_m > 0$);
  - $\alpha_{m_n \to m}$ is excitation from events by user m_n on m ($\alpha_{m_n \to m} > 0$);
  - $\kappa(\delta t) = \mathrm{e}^{-\gamma \delta t}$ and γ is the time scale.

---

[5] Hawkes 1971.

Parametric Hawkes process

- **But!** number of parameters is quadratic in the number of users.
- Make α a function of shared features between each pair of individuals.[6]

$$\alpha_{m_1 \to m_2} = \theta^T \mathbf{f}(m_1 \to m_2) \tag{2}$$

- Now need to estimate only $\#|\theta|$ parameters instead of M^2

[6] Li and Zha 2014.

Features and model comparison

Model	Feature name	$f(m_1, m_2) = 1$ if
F1	**Self-excitation**	$m_1 = m_2$
F2	**Mutual friend**	$m_1 \leftrightarrow m_2$
F3	**Tie strength**	$m_1 \leftrightarrow m_2$ **and** have a *strong* tie.[7]
F4	**Locality**	$m_1 \leftrightarrow m_2$ **and** in same metro area.

[7] measured by the Adamic-Adar index (Adamic and Adar 2003)

Features and model comparison

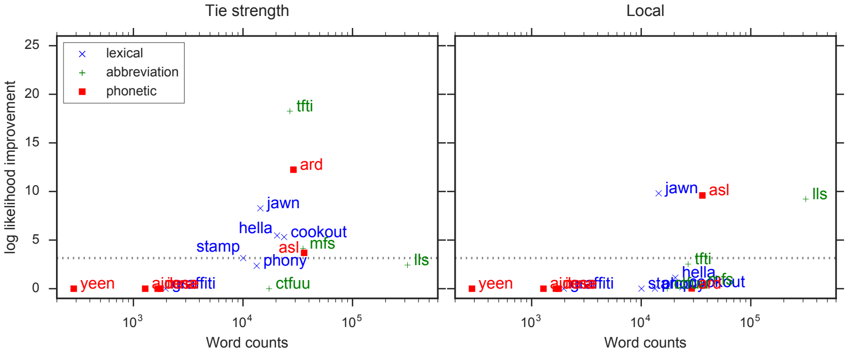
Model	Feature name	$f(m_1, m_2) = 1$ if
F1	**Self-excitation**	$m_1 = m_2$
F2	**Mutual friend**	$m_1 \leftrightarrow m_2$
F3	**Tie strength**	$m_1 \leftrightarrow m_2$ **and** have a *strong* tie.[7]
F4	**Locality**	$m_1 \leftrightarrow m_2$ **and** in same metro area.

Test hypotheses using nested model comparison:

- Null model: F1 + F2
- Null + F3 : are densely embedded ties more linguistically influential?
- Null + F4 : are geographically local ties more linguistically influential?

[7] measured by the Adamic-Adar index (Adamic and Adar 2003)

Results



- Linguistic influence exerted across densely embedded ties greater than linguistic influence exerted across other ties.
- But little evidence to suggest that linguistic influence is exerted across geographically local ties.

Summary of this part

- **Social network matters** : Authors are likely to use a new word if their friends have used it before.
- **Strong ties matter more** : Close friends have more linguistic influence.
- **Shared geography matters less** : Weak evidence that Twitter users pay special attention to local ties.

Follow the Leader: Documents on the Leading Edge of Semantic Change Get More Citations[8]

[8] Sandeep Soni et al. (2019). "Follow the Leader: Documents on the Leading Edge of Semantic Change Get More Citations". In: *in submission JASIST*.

Semantic change

- Languages change as new meanings are attached to known signs.[9]

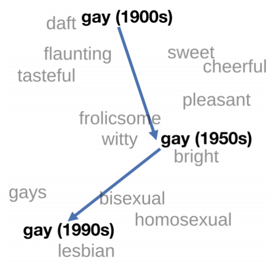


Figure: Hamilton et al. (2016)

- Temporal word representations can help track these changes.

[9] Traugott and Dasher 2001.

Detecting semantic changes

A general recipe is,

1. **Model:** Learn diachronic word embeddings.[10]

   by estimating word embeddings on timesliced corpus and then aligning the embeddings

2. **Measure:** Combine the temporal embeddings to score each word for its shift.[11]

   by quantifying the lack of overlap in the distributional neighborhood across timeslices

3. **Evaluate:** Verify the correctness of detected changes.[12]

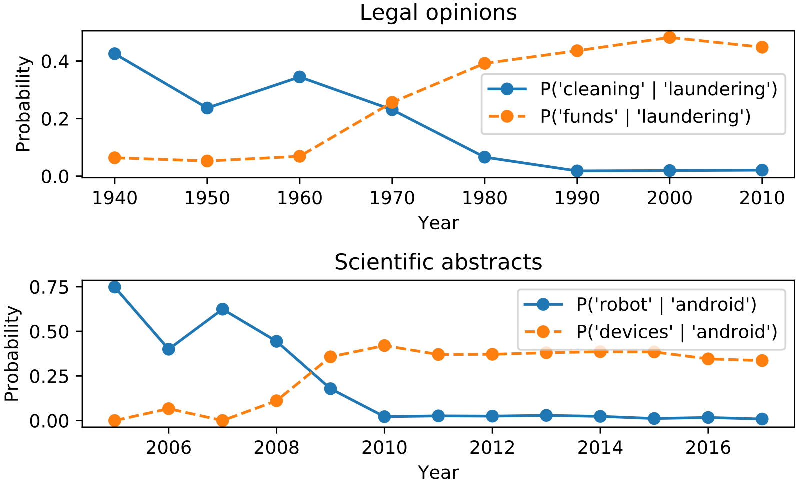
---

[10] e.g., Kulkarni et al. 2015; Hamilton et al. 2016b.

[11] Hamilton et al. 2016a.

[12] for other evaluation schemes, see Mihalcea and Nastase 2012; Shoemark et al. 2019.

Examples


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

What's still missing?

- How to identify **documents** that are leading a change?
- Are leading documents particularly influential?

Just following the general recipe is not enough to answer these questions

Identifying progressive usages

- **Core idea:** For every instance of a word, assess if it is used with a "new" or "old" meaning.
- In the skipgram model, the probability of context word around a target word is:

$$P(w_{i+k} \mid w_i) \propto \exp\left(\mathbf{v}_{w_{i+k}} \cdot \mathbf{u}_{w_i}\right), \tag{3}$$

- Progressiveness of w_i is defined as the log-odds ratio

$$r_{w_i} \triangleq \sum_k \log \frac{P^{(\text{new})}(w_{i+k} \mid w_i)}{P^{(\text{old})}(w_{i+k} \mid w_i)}. \tag{4}$$

- Per instance progressiveness can be aggregated to obtain a progressiveness score for a document.

Examples

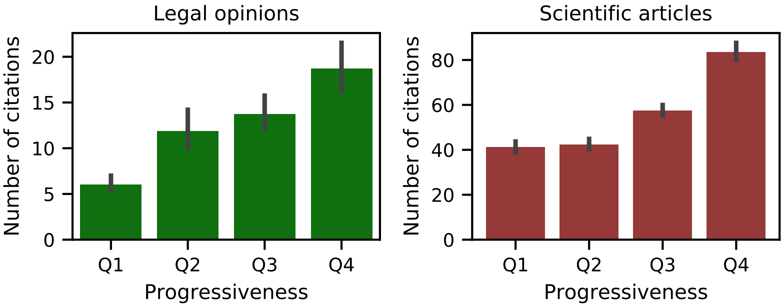
Doc. type	Innovations	Example document with new usage
Legal	laundering	United States v. Talmadge G. Rauhoff, (7th Cir. 1975)
	asylum	Bertrand v. Sava, (S.D.N.Y. 1982)
	fertilization	Planned Parenthood vs Casey (505 U.S. 833)
Science	ux	Hassenzahl and Tractinsky (2006)
	surf	Bay et al. (2008)
	android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

- . . . $15,000 as part of the **'laundering'** process . . .

.

- . . . first step in the successful **laundering** of the funds. . .

Are leading documents more influential?



For both document types, the citations increase for more progressive documents. This is true even in a multivariate regression when controlling for document age, length, # authors, document content, *etc.*

Summary of this part

- Identifying progressive semantic usages helps us zoom into parts of a document and find progressive documents.
- Semantic progressiveness of a document is strongly correlated to the number of citations it gets.
- **Caution!** Progressive language use does not necessarily *cause* the documents to get more citations.

Studies

The thesis comprises five studies:[2]

<u>Peer influence</u>

1. **word adoption** — Which network ties matter in the spread of words?
2. influence detection — When can social influence be detected in event cascades?

<u>Semantic lead</u>

3. **semantic progressiveness** — Which documents are semantically progressive?
4. semantic leadership — Which newspapers lead or lag a specific semantic change?

<u>Identity signaling</u>

5. word abandonment — Why are words abandoned?

[2] blue denotes completed work and red denotes proposed work

Studies

The thesis comprises five studies:[2]

Peer influence

1. word adoption — Which network ties matter in the spread of words?
2. influence detection — When can social influence be detected in event cascades?

Semantic lead

3. semantic progressiveness — Which documents are semantically progressive?
4. **semantic leadership** — Which newspapers lead or lag a specific semantic change?

Identity signaling

5. **word abandonment** — Why are words abandoned?

[2] blue denotes completed work and red denotes proposed work

Proposed work

Discovering semantic leadership in abolitionist newspapers[13]

[13] Work in progress with Lauren Klein and Jacob Eisenstein

Beyond document semantic leaders

- Many timestamped collections have additional metadata.

  court opinions have jurisdictions, science articles are from journals, tweets have locations, *etc.*

- *Who* is a semantic leader with respect to a change can also be asked at the source level.
- But diachronic word embeddings are not yet fully equipped to answer this question.

Abolitionist newspapers

- Dynamical patterns of text borrowing and reuse is seen in nineteenth century newspapers.[14]
- This opens up research questions such as:
  - In a pair of newspapers, which one was a leader for a given change?
  - Which newspaper was a semantic leader overall or for a category of semantic changes?
  - Were semantic leaders different from lexical leaders?
- Answers to these questions can help quantitatively evaluate the influence and role these newspapers played.

[14] Cordell 2015.

Data

- **Period**: 1827-1865
- **Scale**: 200K newspaper articles, $\approx$ 130 million tokens
- **Newspapers**[15]: 14

Douglass Monthly, Frank Leslie's Weekly, The Frederick Douglass Paper, The Freedoms Journal, Godey's Lady's Book, The National Anti-Slavery Standard, The Provincial Freeman, The Christian Recorder, The Colored American, The Liberator, The Lily, The National Era, The North Star, The Weekly Advocate

[15] Susan Maret (2016). "Accessible archives". In: *The Charleston Advisor* 18.2, pp. 17–20.

Technical plan

The steps are to:

1. Estimate source-conditional temporal embeddings,
2. Identify leaders for specific changes,
3. Aggregate into global influence network,
4. Compare with network of lexical innovations.

Technical plan

The steps are to:

1. Estimate source-conditional temporal embeddings,
2. Identify leaders for specific changes,
3. Aggregate into global influence network,
4. Compare with network of lexical innovations.

Source-conditional diachronic embeddings

- Decompose temporal embedding into a base embedding and a temporal deviation.[16]

$$\mathbf{w}_t = \mathbf{w}_b + \mathbf{r}_t, \tag{5}$$

- Extend temporal embedding by adding a source-specific deviation

$$\mathbf{w}_{s,t} = \mathbf{w}_b + \mathbf{r}_t + \mathbf{v}_{s,t}, \tag{6}$$

- Embeddings and deviations can be learned by optimizing the *word2vec* objective.[17]
- Learn the temporal-only model separately to identify semantic changes and then the source-conditional model.

---

[16] adapted from Bamman et al. 2014; Gillani and Levy 2019.

[17] Mikolov et al. 2013.

Technical plan

The steps are to:

1. Estimate source-conditional temporal embeddings,
2. Identify leaders for specific changes,
3. Aggregate into global influence network,
4. Compare with network of lexical innovations.

Technical plan

The steps are to:

1. Estimate source-conditional temporal embeddings,
2. Identify leaders for specific changes,
3. Aggregate into global influence network,
4. Compare with network of lexical innovations.

Leadership scoring scheme

Assume w changed between t_1 and t_2 ($t_1 < t_2$).

Leadership scoring scheme

Assume w changed between t_1 and t_2 $(t_1 < t_2)$.

- For a pair of newspapers s_1 and s_2, the lead of s_1 over s_2 can be defined as:

$$\text{lead}(s_1 \rightarrow s_2) \triangleq \frac{\mathbf{w}_{s_1,t_1} \cdot \mathbf{w}_{s_2,t_2}}{\mathbf{w}_{s_2,t_1} \cdot \mathbf{w}_{s_2,t_2}} \tag{7}$$

Leadership scoring scheme

Assume w changed between t_1 and t_2 ($t_1 < t_2$).

- For a pair of newspapers s_1 and s_2, the lead of s_1 over s_2 can be defined as:

$$\text{lead}(s_1 \rightarrow s_2) \triangleq \frac{\mathbf{w}_{s_1,t_1} \cdot \mathbf{w}_{s_2,t_2}}{\mathbf{w}_{s_2,t_1} \cdot \mathbf{w}_{s_2,t_2}} \tag{7}$$

- Main intuition: If s_1 leads then it must have used w with a new meaning before s_2 caught up (high numerator, low denominator).

Technical plan

The steps are to:

1. Estimate source-conditional temporal embeddings,
2. Identify leaders for specific changes,
3. Aggregate into global influence network,
4. Compare with network of lexical innovations.

Technical plan

The steps are to:

1. Estimate source-conditional temporal embeddings,
2. Identify leaders for specific changes,
3. Aggregate into global influence network,
4. Compare with network of lexical innovations.

Network induction for semantic changes

- Induce a directed, weighted network between the newspapers by aggregating over all semantic changes.[18]
- Network view enables finding overall leaders and core-periphery structures between the leaders.
- This can help in discovering influential newspapers and gauge their central role in shaping opinions at that time.

[18] Few works have attempted to induce linguistic influence networks(e.g. Smith et al. 2013; Guo et al. 2015; He et al. 2015) but on lexical or topical changes

Technical plan

The steps are to:

1. Estimate source-conditional temporal embeddings,
2. Identify leaders for specific changes,
3. Aggregate into global influence network,
4. Compare with network of lexical innovations.

Technical plan

The steps are to:

1. Estimate source-conditional temporal embeddings,
2. Identify leaders for specific changes,
3. Aggregate into global influence network,
4. Compare with network of lexical innovations.

Network induction for lexical changes

- Repeat same network induction steps for lexical changes.
- Quantify node and structural differences between the two networks.
- Differences could be key in distinguishing between informational and opinion leaders among newspapers.

Evaluation

Evaluation is challenging in the absence of ground truth.

- One approach is to group semantic changes thematically and establish leaders for each group.
- The discovered relationships can then be validated against expert judgments.

Recap of this proposed work

- Source-conditional diachronic embeddings to find leaders with respect to semantic changes,
- Network induction and analysis of 19c antislavery newspapers for both semantic and lexical changes,
- Evaluation to verify the role of these newspapers in shaping the discourse.

Proposed work

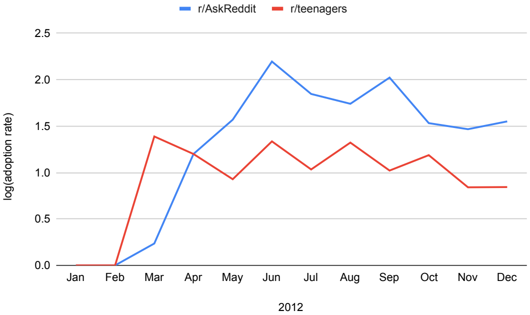
Linguistic abandonment as a form of identity signaling on Reddit[19]

[19] in collaboration with Jonah Berger and Jacob Eisenstein

The case of yolo

Image credit: FiveThirtyEight

The case of yolo



yolo originated among young (and AA) redditors but they gradually abandoned the word as it was picked by mainstream subreddits.

Linguistic abandonment

- Why language changes at a specific time is a fundamental problem in sociolinguistics.[20]
- Language change as a social contagion suggests stable adoption across all population groups.
- The strategic decision-making in the abandonment of linguistic innovations is an underexplored research topic.

[20]Weinreich et al. 1968.

Suggested reasons for abandonment

- Language choices are a means to socially differentiate.[21]
- Abandonment is a form of identity signaling by the adopters after evaluating innovation use in mainstream.[22]

  Linguistic and non-experimental evidence is not available.

[21] Bourdieu 1984.

[22] Berger 2008.

Hypothesis

Identity relevant linguistic innovations are more likely to be abandoned in subreddits in which they were salient as a result of their adoption in mainstream subreddits.

Technical plan

The steps are:

1. Identify identity relevant and mainstream subreddits,

   find clique-like or modular components in a cross-posting subreddit network[23]

2. Identify relevant linguistic innovations,

   find terms that are distinctive to a community[24]

3. Model the temporal dynamics of innovations.

---

[23] Newman 2006.

[24] Eisenstein et al. 2014.

Technical plan

The steps are:

1. Identify identity relevant and mainstream subreddits,

   find clique-like or modular components in a cross-posting subreddit network[23]

2. Identify relevant linguistic innovations,

   find terms that are distinctive to a community[24]

3. Model the temporal dynamics of innovations.

[23] Newman 2006.

[24] Eisenstein et al. 2014.

Modeling the temporal dynamics

Model a linguistic innovation as an event cascade over the subreddit network using regressive point process models.[25]

- Recall in multivariate HP, $\mu > 0$ and $\alpha > 0$

$$\lambda_m(t) = \mu_m + \sum_{t \leq t_n} \alpha_{m_n \to m} \kappa(t - t_n), \tag{8}$$

- If innovation use in one subreddit inhibits the use in another, allow μ and α to be negative.

$$\lambda^*_m(t) = f(\lambda_m(t)), \tag{9}$$

where $\lambda^*(t)$ is the new HP intensity function and $f(\cdot)$:$\mathbb{R} \to \mathbb{R}^+$ ensures it is positive at any time.

- Hypothesis testing: $\alpha_{\mathcal{M} \to \mathcal{S}} < 0$

[25] see variations in Mei and Eisner 2017; Apostolopoulou et al. 2019.

Summary of this proposed work

- The proposed work tests whether linguistic abandonment in a community is modulated by mainstream adoption.
- This is an entry point in studying more granular linguistic hypotheses related to cultural appropriation and semantic bleaching.

Thesis summary

- Language change can inform us about the latent social structures

  who talks to whom? who leads? who follows? who diverges?

Thesis summary

- Language change can inform us about the latent social structures

  who talks to whom? who leads? who follows? who diverges?

- Computational modeling of language change using diachronic text can help uncover these latent patterns.

Thesis summary

- Language change can inform us about the latent social structures

  who talks to whom? who leads? who follows? who diverges?

- Computational modeling of language change using diachronic text can help uncover these latent patterns.
- This thesis offers a few concrete ways to model timestamped text and demonstrates the complex interplay between language change, influence, and identity.

Timeline

May 2020	[Semantic leadership] Analysis
May 20, 2020	[Semantic progressiveness] Resubmission to JASIST
Jun 1, 2020	[Semantic leadership] Submit to EMNLP
Jun - Sep, 2020	[Word abandonment] Model and analysis
Oct 15, 2020	[Word abandonment] Submit to WebConf
Oct - Nov, 2020	Thesis writing
Dec 15, 2020	Thesis defense

Thank you!

Thank you!

- Computational Linguistics Lab members

- **Collaborators**: Fernando Diaz, Lauren F. Klein, Rahul Goel, Naman Goyal, John Paparrizos, Hanna Wallach, Kristina Lerman, Jonah Berger, Shawn Ling Ramirez.

References I

Adamic, Lada A and Eytan Adar (2003). "Friends and neighbors on the web". In: *Social networks* 25.3, pp. 211–230.

Apostolopoulou, Ifigeneia, Scott Linderman, Kyle Miller, and Artur Dubrawski (2019). "Mutually Regressive Point Processes". In: *Advances in Neural Information Processing Systems*, pp. 5116–5127.

Bamman, David, Chris Dyer, and Noah A Smith (2014). "Distributed representations of geographically situated language". In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 828–834.

References II

- Bay, Herbert, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool (2008). "Speeded-up robust features (SURF)". In: *Computer vision and image understanding* 110.3, pp. 346–359.
- Berger, Jonah (2008). "Identity signaling, social influence, and social contagion". In: *Understanding peer influence in children and adolescents*, pp. 181–199.
- Bourdieu, Pierre (1984). "Distinction: A Social Critique of the Judgement of Taste". In:
- Cordell, Ryan (2015). "Reprinting, circulation, and the network author in antebellum newspapers". In: *American Literary History* 27.3, pp. 417–445.

References III

- Danescu-Niculescu-Mizil, Cristian, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts (2013). "No country for old members: User lifecycle and linguistic change in online communities". In: *Proceedings of the 22nd international conference on World Wide Web*, pp. 307–318.
- Eisenstein, Jacob, Brendan O'Connor, Noah A Smith, and Eric P Xing (2014). "Diffusion of lexical change in social media". In: *PloS one* 9.11, e113114.
- Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou (2018). "Word embeddings quantify 100 years of gender and ethnic stereotypes". In: *Proceedings of the National Academy of Sciences* 115.16, E3635–E3644.

References IV

Gerow, Aaron, Yuening Hu, Jordan Boyd-Graber, David M Blei, and James A Evans (2018). "Measuring discursive influence across scholarship". In: *Proceedings of the National Academy of Sciences* 115.13, pp. 3308–3313.

Gillani, Nabeel and Roger Levy (2019). "Simple dynamic word embeddings for mapping perceptions in the public sphere". In: *Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science*, pp. 94–99.

Goel, Rahul, Sandeep Soni, Naman Goyal, John Paparrizos, Hanna Wallach, Fernando Diaz, and Jacob Eisenstein (2016). "The social dynamics of language change in online networks". In: *International Conference on Social Informatics*. Springer, pp. 41–57.

Grieve, Jack (2018). "Mapping emerging words in New York city". In: *New ways of analyzing variation*.

References V

Guo, Fangjian, Charles Blundell, Hanna Wallach, and Katherine Heller (2015). "The bayesian echo chamber: Modeling social influence via linguistic accommodation". In: *Artificial Intelligence and Statistics*, pp. 315–323.

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

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*. Vol. 2016, p. 2116.

References VI

- Hamilton, William L, Jure Leskovec, and Dan Jurafsky (2016b). "Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vol. 1, pp. 1489–1501.
- Hassenzahl, Marc and Noam Tractinsky (2006). "User experience-a research agenda". In: *Behaviour & information technology* 25.2, pp. 91–97.
- Hawkes, Alan G (1971). "Spectra of some self-exciting and mutually exciting point processes". In: *Biometrika* 58.1, pp. 83–90.

References VII

- He, Xinran, Theodoros Rekatsinas, James Foulds, Lise Getoor, and Yan Liu (2015). "Hawkestopic: A joint model for network inference and topic modeling from text-based cascades". In: *International conference on machine learning*, pp. 871–880.
- Huberman, Bernardo A, Daniel M Romero, and Fang Wu (2008). "Social networks that matter: Twitter under the microscope". In: *arXiv preprint arXiv:0812.1045*.
- Kulkarni, Vivek, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena (2015). "Statistically significant detection of linguistic change". In: *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, pp. 625–635.

References VIII

- Li, Liangda and Hongyuan Zha (2014). "Learning parametric models for social infectivity in multi-dimensional hawkes processes". In: *Twenty-Eighth AAAI Conference on Artificial Intelligence*.
- Maret, Susan (2016). "Accessible archives". In: *The Charleston Advisor* 18.2, pp. 17–20.
- Mei, Hongyuan and Jason M Eisner (2017). "The neural hawkes process: A neurally self-modulating multivariate point process". In: *Advances in Neural Information Processing Systems*, pp. 6754–6764.
- Mihalcea, Rada and Vivi Nastase (2012). "Word epoch disambiguation: Finding how words change over time". In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 259–263.

References IX

- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean (2013). "Distributed representations of words and phrases and their compositionality". In: *Advances in neural information processing systems*, pp. 3111–3119.
- Newman, Mark EJ (2006). "Modularity and community structure in networks". In: *Proceedings of the national academy of sciences* 103.23, pp. 8577–8582.
- Rudolph, Maja and David Blei (2018). "Dynamic embeddings for language evolution". In: *Proceedings of the 2018 World Wide Web Conference*, pp. 1003–1011.
- Shabtai, Asaf, Yuval Fledel, Uri Kanonov, Yuval Elovici, Shlomi Dolev, and Chanan Glezer (2010). "Google android: A comprehensive security assessment". In: *IEEE Security & Privacy*, pp. 35–44.

References X

Shoemark, Philippa, Farhana Ferdousi Liza, Dong Nguyen, Scott Hale, and Barbara McGillivray (2019). "Room to Glo: A Systematic Comparison of Semantic Change Detection Approaches with Word Embeddings". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 66–76.

Smith, David A, Ryan Cordell, and Elizabeth Maddock Dillon (2013). "Infectious texts: Modeling text reuse in nineteenth-century newspapers". In: *2013 IEEE International Conference on Big Data*. IEEE, pp. 86–94.

Soni, Sandeep, Kristina Lerman, and Jacob Eisenstein (2019). "Follow the Leader: Documents on the Leading Edge of Semantic Change Get More Citations". In: *in submission JASIST*.

References XI

Traugott, Elizabeth Closs and Richard B Dasher (2001). *Regularity in semantic change*. Vol. 97. Cambridge University Press.

Weinreich, Uriel, William Labov, and Marvin Herzog (1968). "Empirical foundations for a theory of language change". In: *Directions for historical linguistics*, pp. 97–188.

The role of social networks

Close-knit networks have the capacity to maintain and even enforce local conventions and norms including linguistic norms

- J.Milroy and L.Milroy

The leaders of linguistic change are people at the center of their social networks

- William Labov

Corpora

Statistic	Legal opinions	Scientific abstracts
Number of documents	3,854,738	2,408,010
Years	1754–2018	1949–2018
Average number of citations (in-degree)	7.84	39.19
Average number of references (out-degree)	7.80	9.49
Length (number of unique word types per document)	632.22	93.10

Table: Descriptive summary of the two datasets. Legal opinions are from US Courts. Scientific abstracts are from DBLP.

Multivariate regression: scientific abstracts

Predictors	M1	M2	M3	M4
Constant	1.983	1.943	2.032	1.770
Outdegree	0.009	0.009	0.009	0.009
# Authors	0.055	0.054	0.054	0.054
Age	0.079	0.079	0.078	0.073
Length	0.002	0.002	0.002	0.002
BoWs	0.000	0.000	0.000	0.000
# Innovs		0.028	-0.010	-0.034
Prog.			0.137	
Prog. Q2				0.179
Prog. Q3				0.431
Prog. Q4				0.698
Log Lik.	-13.07	-13.06	-12.93	-**12.82**

Table: Poisson regression analysis of citations to scientific abstracts. Log likelihood is in millions of nats.

Multivariate regression: legal opinions

Predictors	M1	M2	M3	M4
Constant	1.614	1.421	1.476	1.168
Outdegree	0.022	0.020	0.021	0.020
Age	0.009	0.011	0.010	0.010
Length	0.000	-0.000	-0.000	-0.000
BoWs	-0.000	-0.000	-0.000	-0.000
# Innovs		0.054	0.045	0.042
Prog.			0.094	
Prog. Q2				0.384
Prog. Q3				0.382
Prog. Q4				0.470
Log Lik.	-415195	-410601	**-406843**	-408031

Table: Poisson regression analysis of citations to legal opinions