Applications of semantic change research in other domains

Haim Dubossarsky, h.dubossarsky@qmul.ac.uk

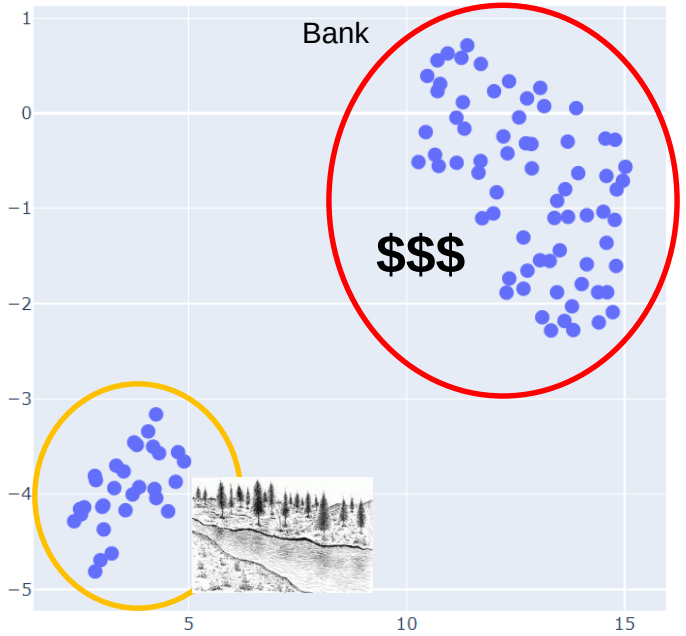
Previously in the course....

Synchronic models of meaning

PPMI

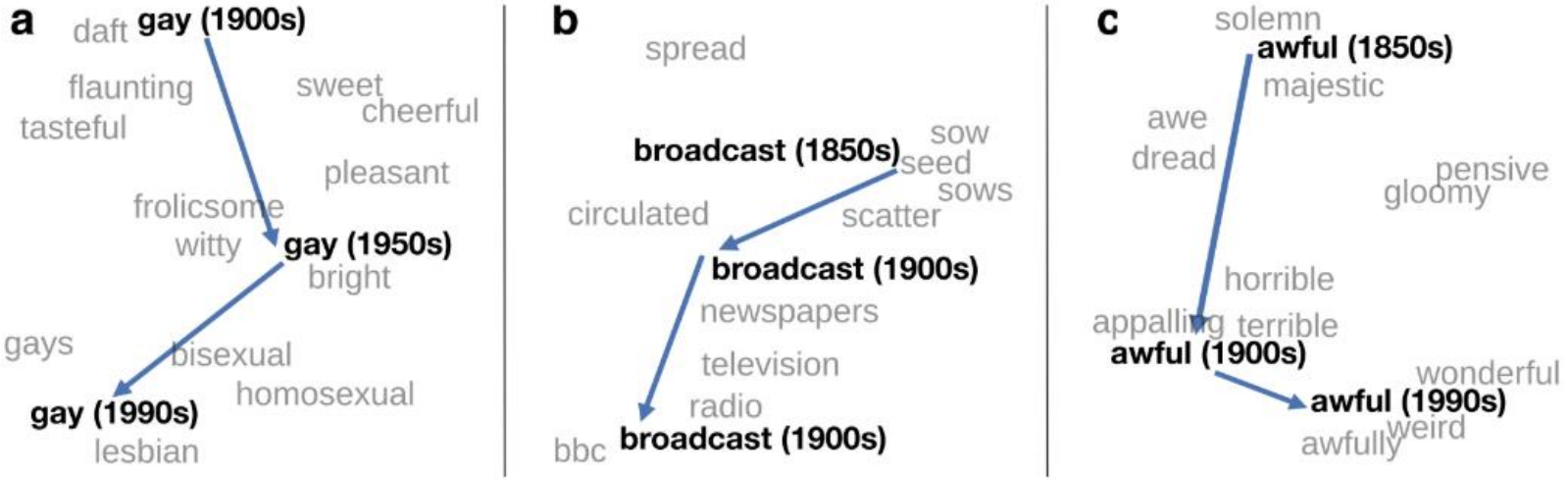
You shall know a word by the company it keeps (Firth)

Contextualized models



Words in similar contexts tend to have similar meanings (Harris)

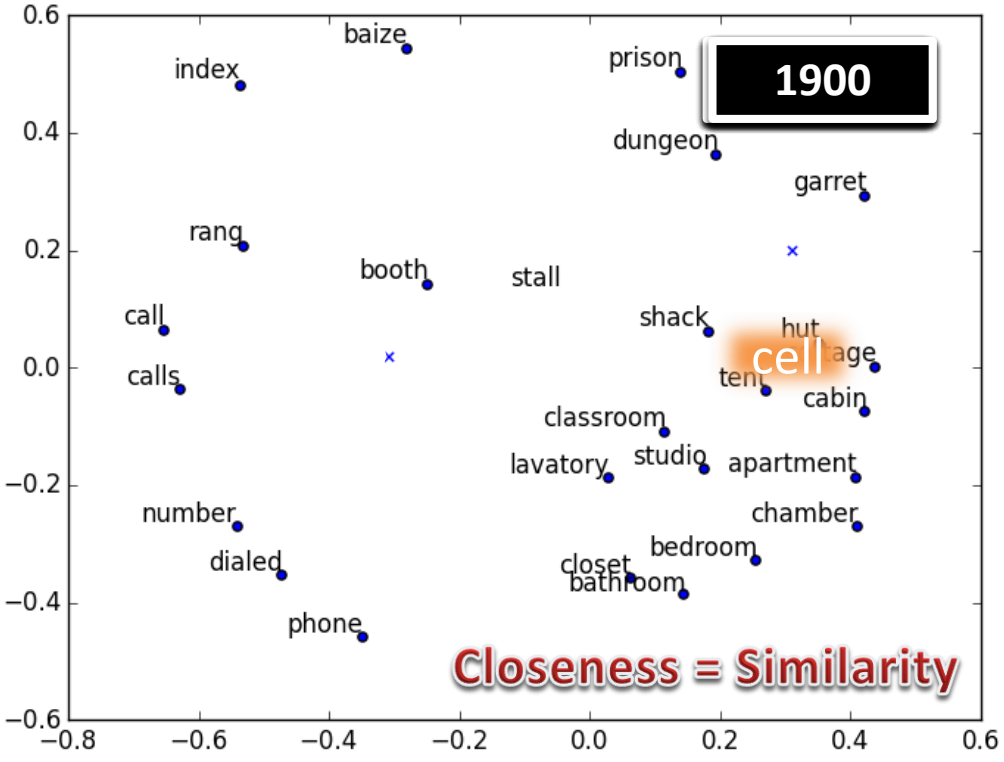
How does meaning change through time



From Hamilton et al. 2016

How does meaning change through time

Meaning change of *cell*	
Time	**Change score**
1900-10	
1910-20	
1920-30	
1930-40	
1940-50	
1950-60	
1960-70	
1970-80	
1980-90	
1990-2000	

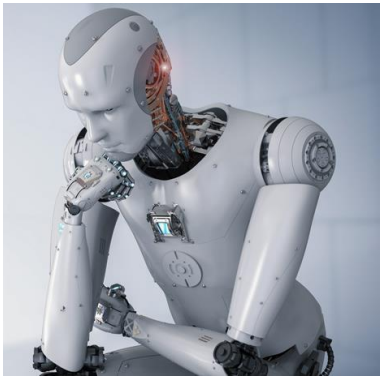


What can we do with it?

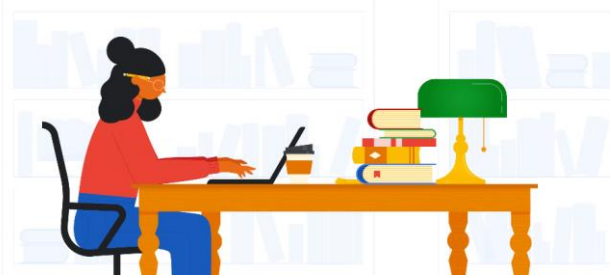
Why change (and variation) matters

If you are a linguist

If you work on AI



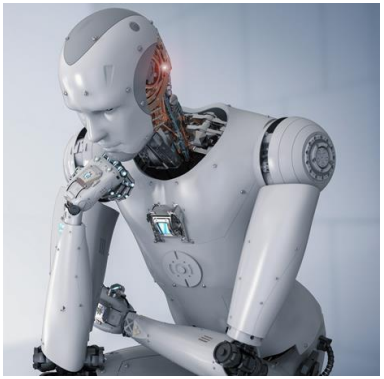
If you are a historian, sociologist or interested in societal changes



Why change (and variation) matters

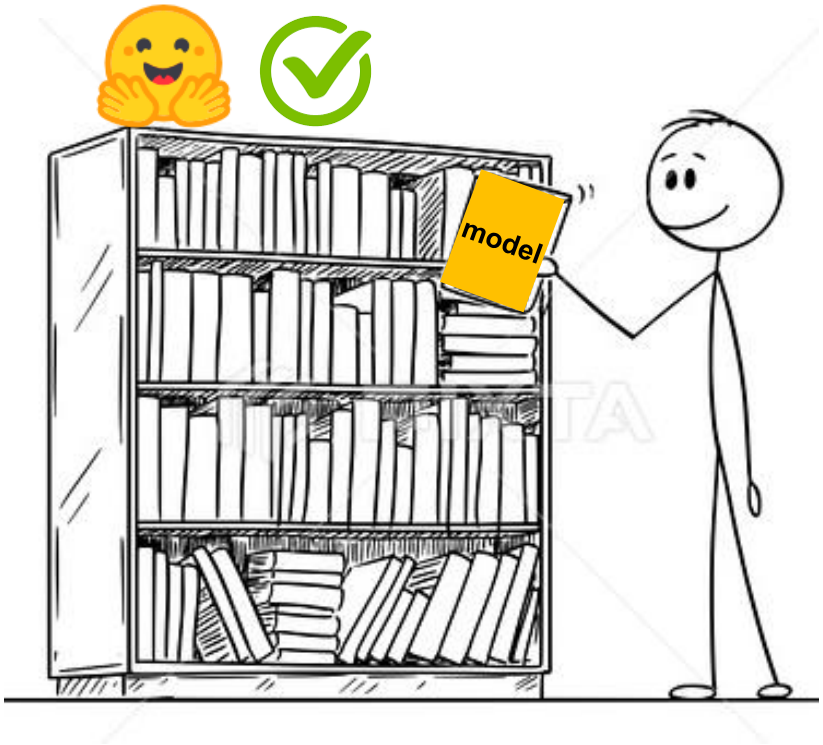
If you are a linguist

If you work on AI



If you are a historian, sociologist or interested in societal changes

Current modus operandi



Works well iff:

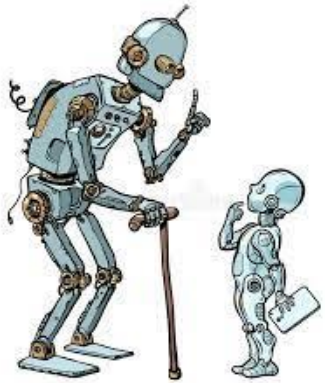
- Domain is very similar to the training dataset
- Finetuning on suitable dataset is possible

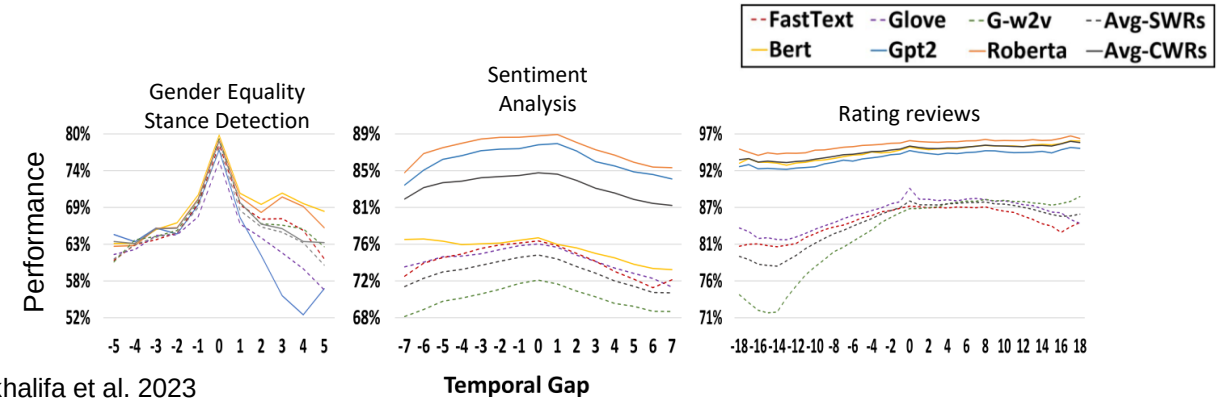
Problematic in cases of:

- **Language change**
- **Language variation**
- **Any domain change**

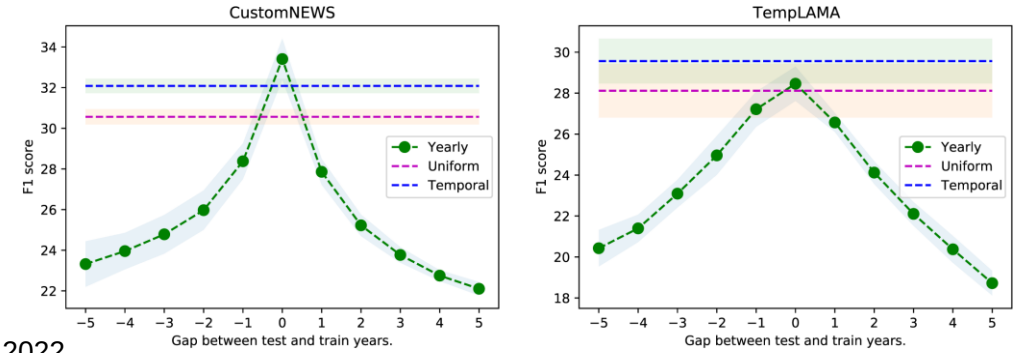
Semantic change causes bias

Performance drop as a function of language change and variation





From Alkhalifa et al. 2023



Dhingra et al. 2022

Semantic change causes bias

The case of reclaimed language

Standard hate speech problem

Hateful	Non hateful
I just love it when all of the **brown people** leave my office so I don't have to look at them	the **bl**

Reclamation takes deroga
group, and consciously use
thus turning a hurtful term into

True for **any** minority or marginalized community: women, immigrants, religious groups, elderly, etc.

Elicit negative attitudes, dehumanization, and discrimination, arginalized communities groups.

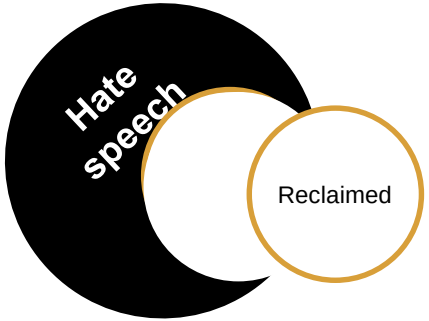
Among the LGBTQ+ community

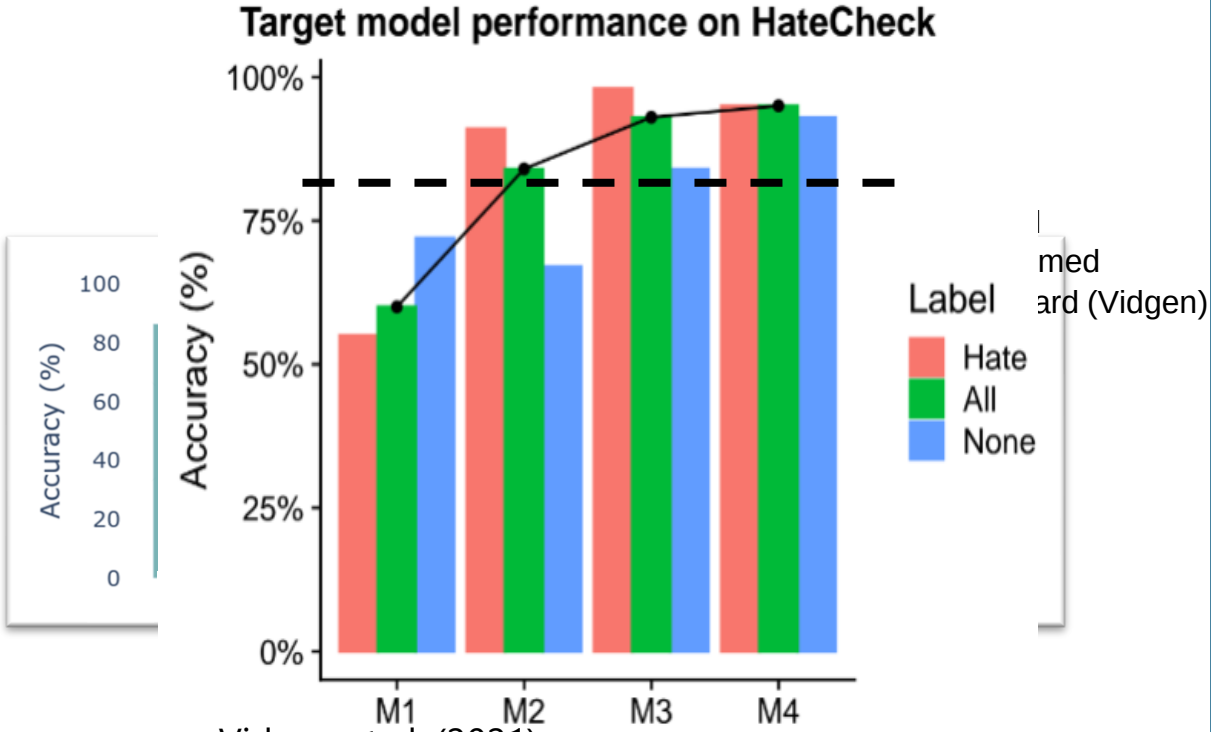
Hateful	Reclaimed
Look, that boy's crying! What a sissy!	Sissy That Walk, queens!

Marginalize members can change meaning and deconstruct *power*-relationships

Semantic change causes bias

The case of reclaimed language

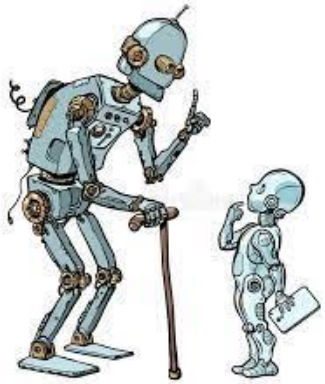


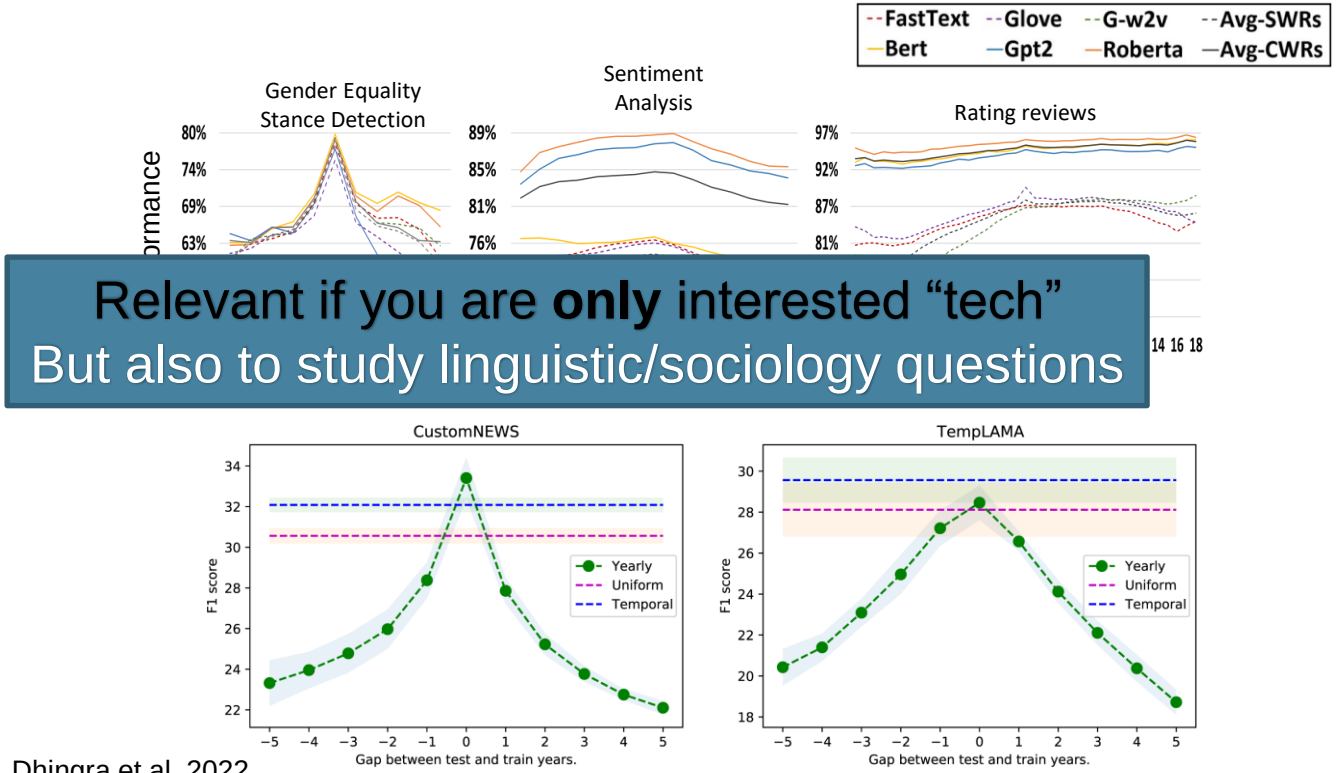


Vidgen et al. (2021)

Semantic change causes bias

Performance drop as a function of language change and variation





Relevant if you are **only** interested “tech”
But also to study linguistic/sociology questions

Dhingra et al. 2022

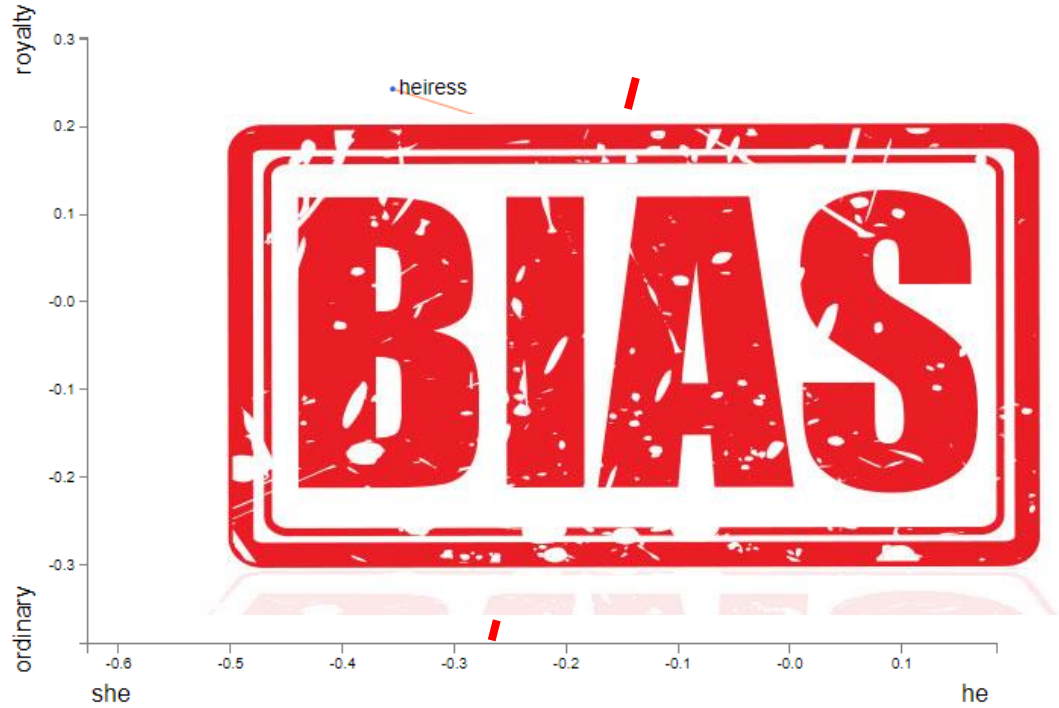
Why change (and variation) matters

If you are a linguist

If you work on AI

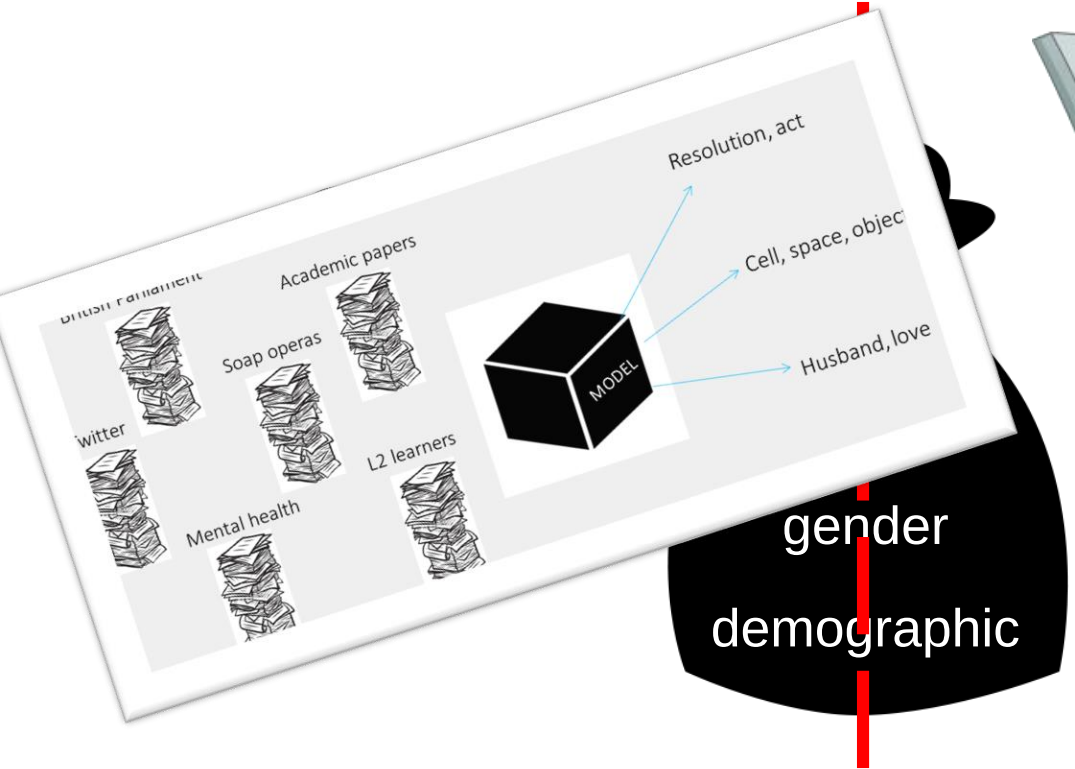
If you are a historian, sociologist or interested in societal changes

Models are biased

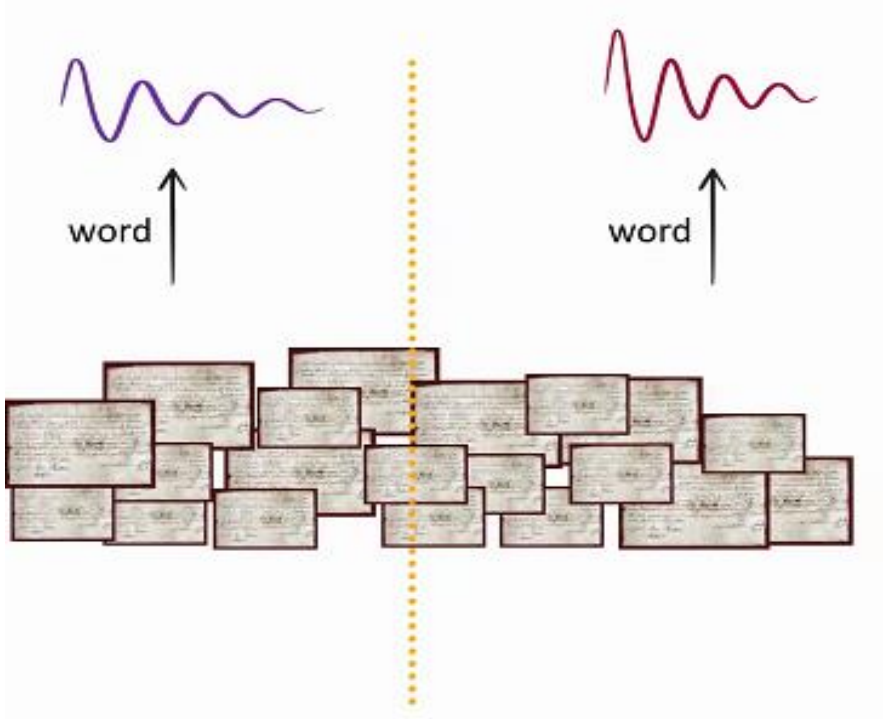


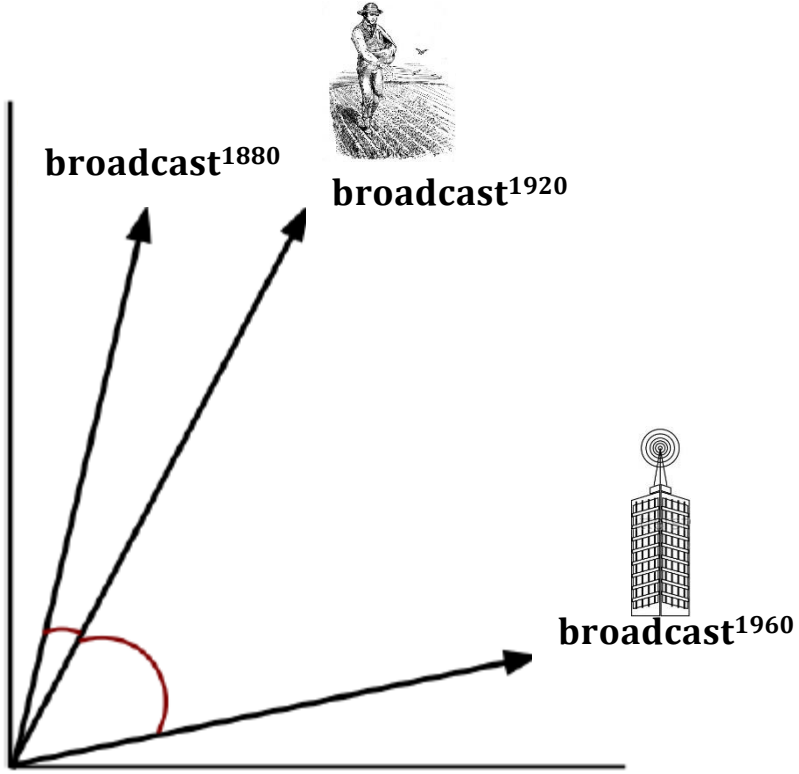
https://lamyiowce.github.io/word2viz/

Bias is neither good or bad, but a mirror to society

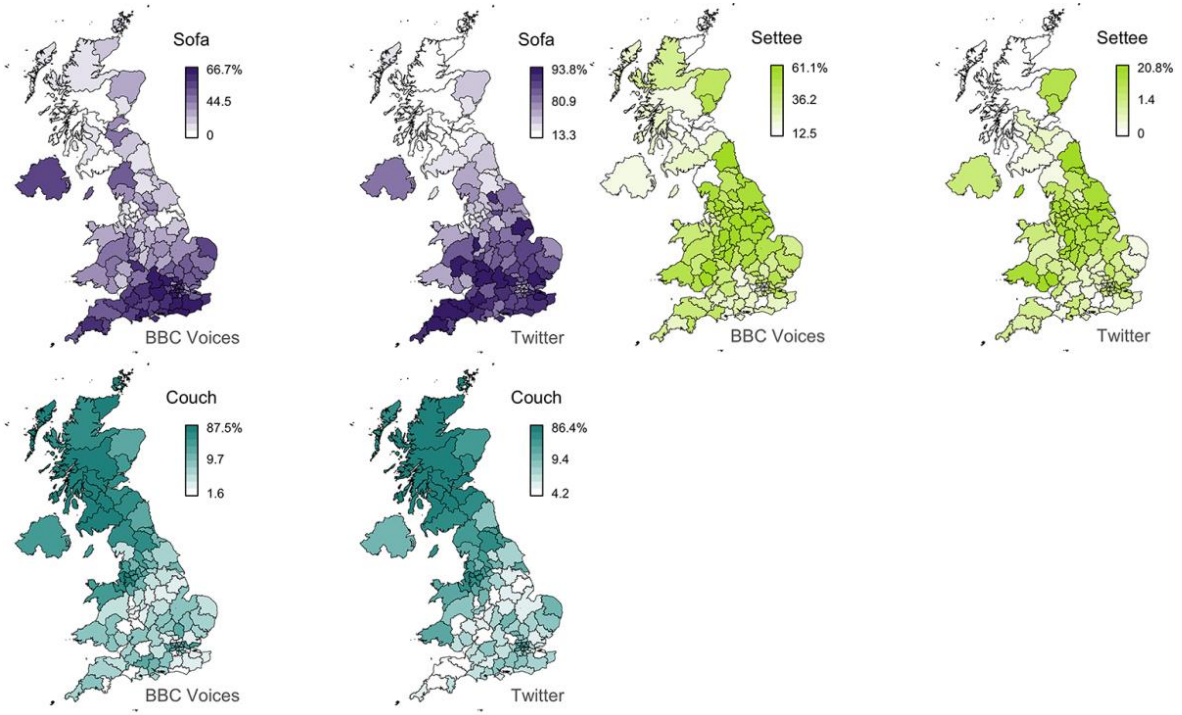


Study how bias changes over time & across domains



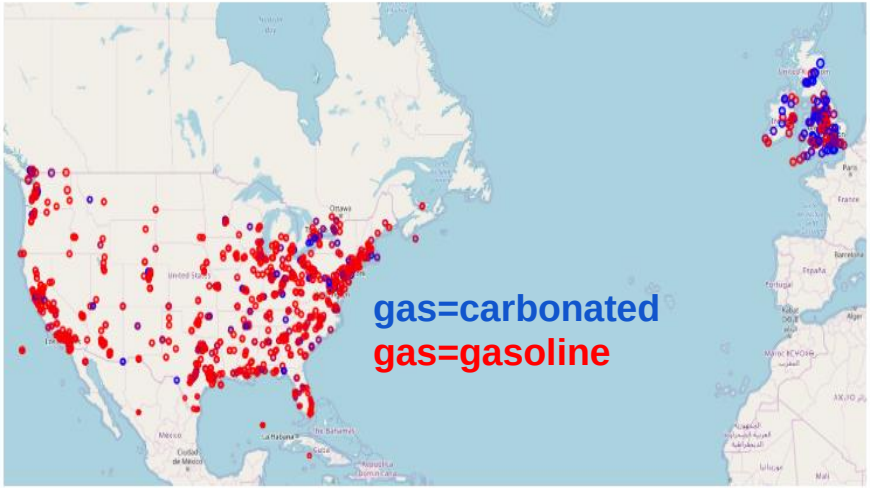


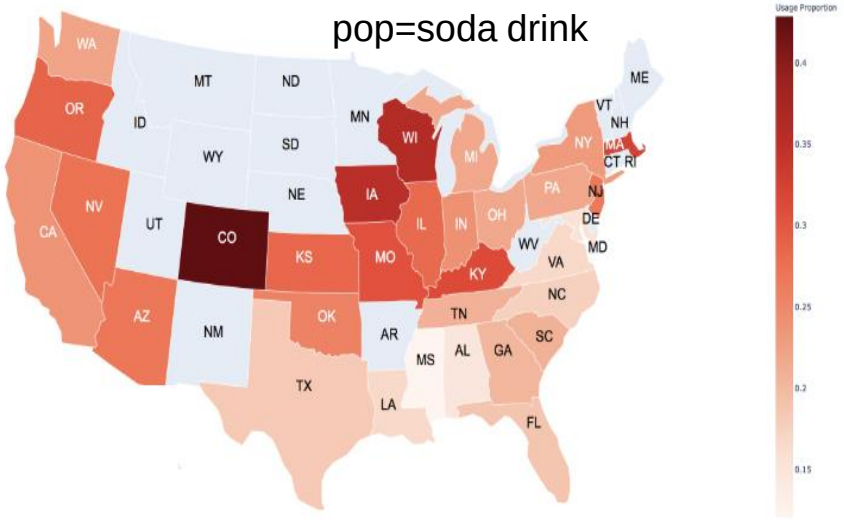
Language variation and dialectology as semantic change



Grive et. al 2019

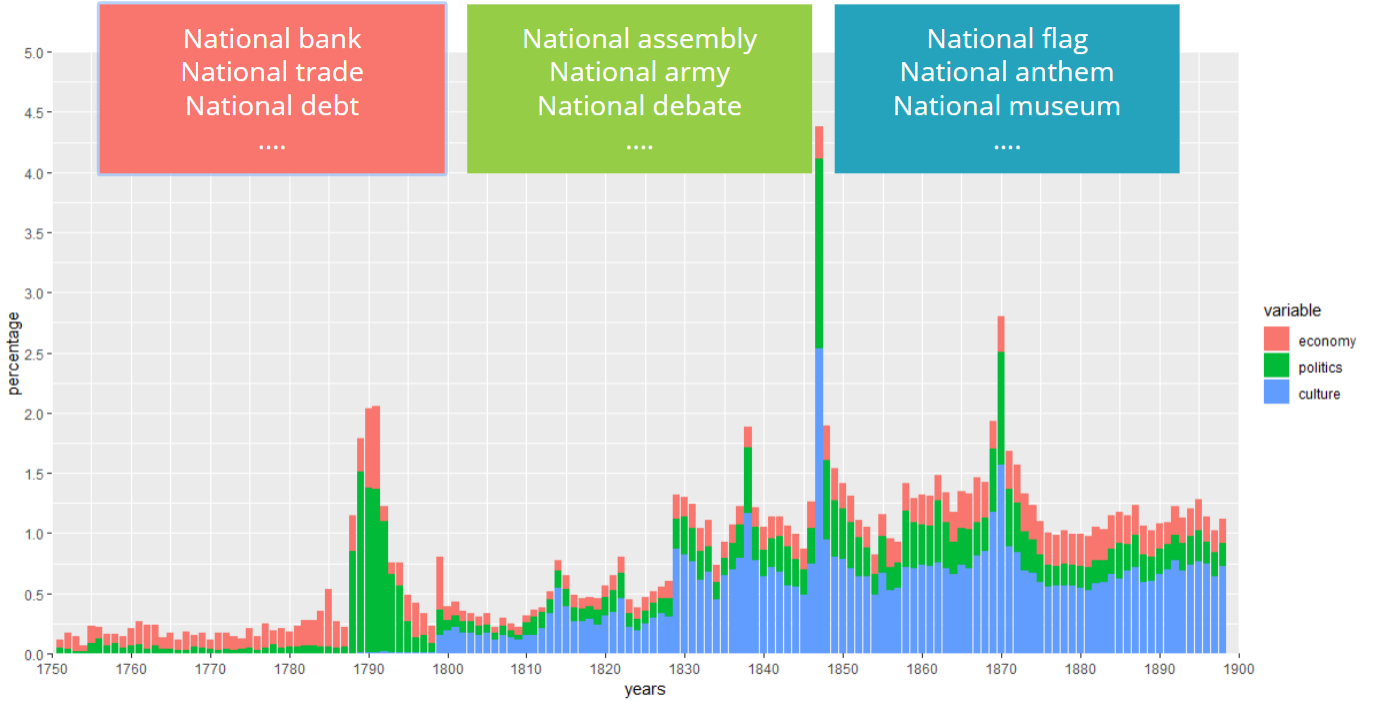
Language variation and dialectology as semantic change





Jiang et al. 2020

Semantic change in politics and sociology



From Henchen et al. 2019

Semantic change in history and technology

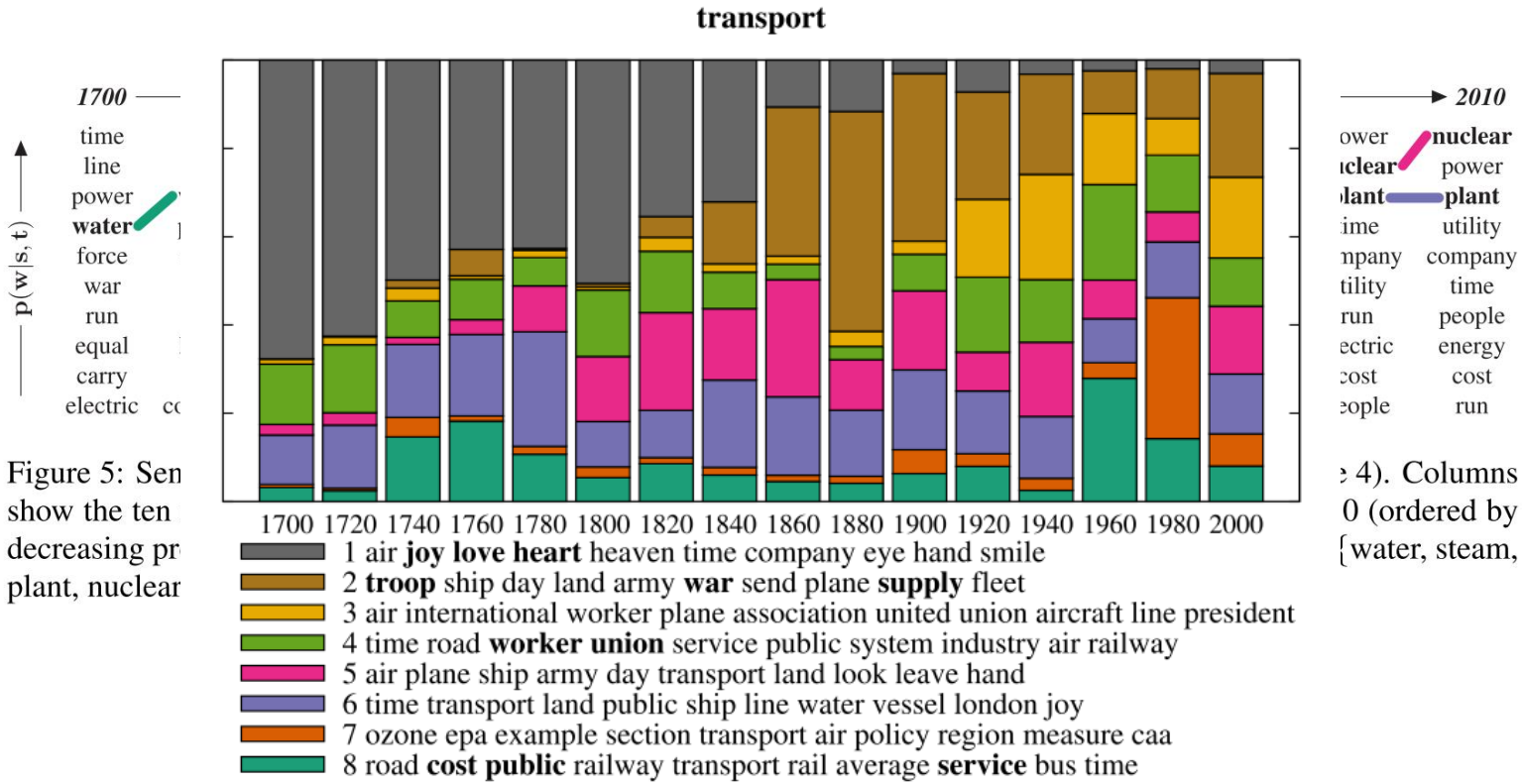


Figure 5: Sen … 4). Columns show the ten … 0 (ordered by decreasing pr … {water, steam, plant, nuclear

Frermann & Lapata, 2016

Semantic change in the finance domain

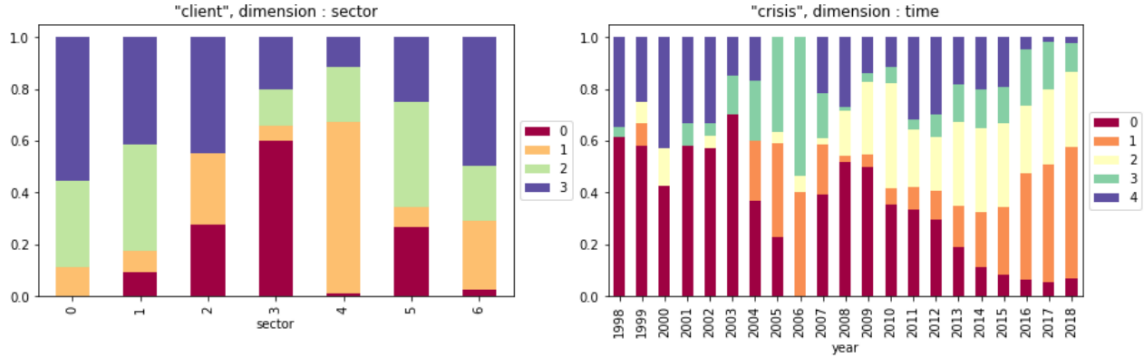


Figure 1: Distribution of clusters per Office for the word *client* (left) and per year for the word *crisis* (right) in the SEC-Edgar corpus. The Offices are described in Table 4

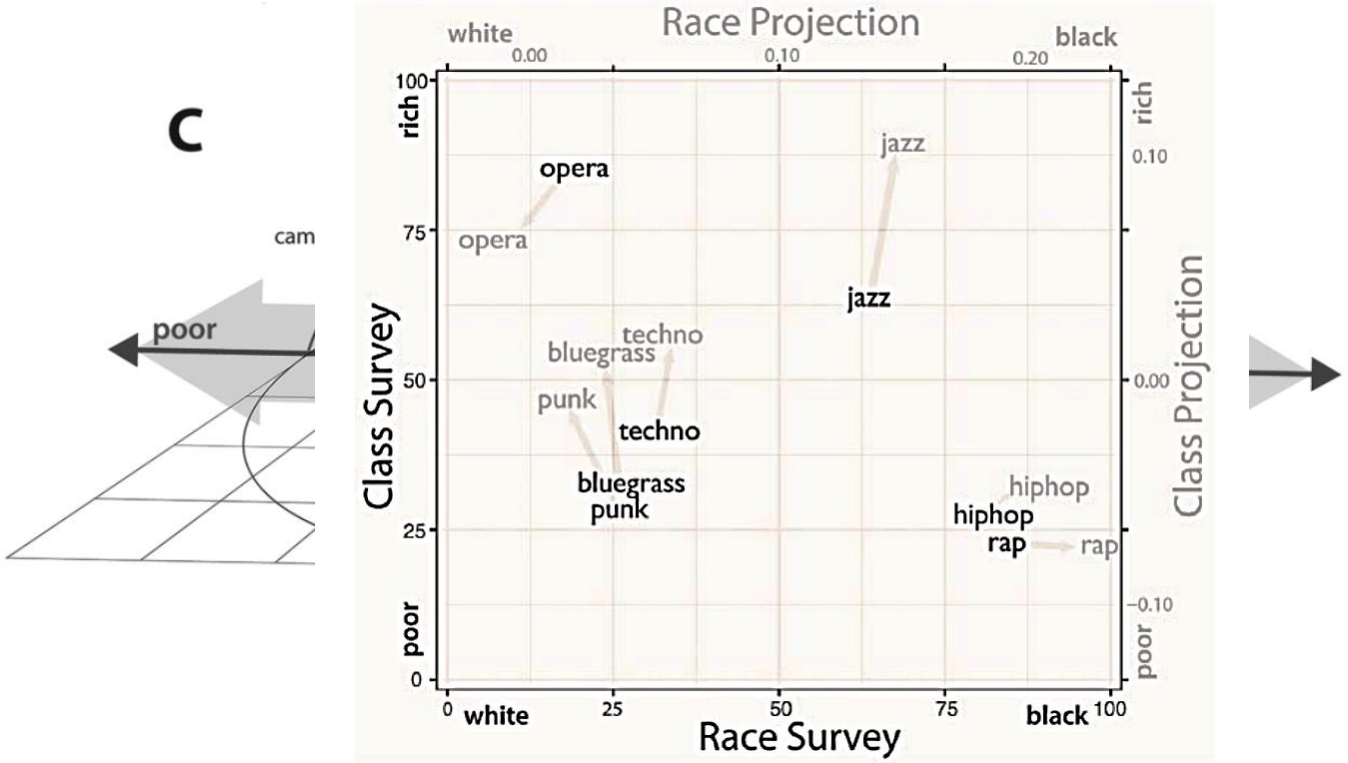
N^o	Keyword examples - Word = *client*
0	server, products, data, applications, services, systems
1	revenue, contract, risk, costs, loss, business, fees
2	assets, funds, cash, interest, balances, investment
3	services, business, revenue, growth, management, products

N^o	Keyword examples - Word = *crisis*
0	liquidity, funding, contingency, cash, collateral, outflows
1	marketing, business, management, design, advertising, media
2	european, debt, credit, sovereign, countries, eurozone, banks
3	financial, accident, capital, regulatory, loss, liquidity, funding
4	credit, financial, global, markets, debt, european, recession

Table 5: List of clusters and keyword examples for the words *client* (left) and *crisis* (right) in the SEC-Edgar Corpus

From Montariol et al., 2020

Semantic change in the cultural domain



From Kozlowski et al., 2019

Why change (and variation) matters

If you are a linguist

If you work on AI

If you are a historian, sociologist or interested in societal changes

Exploratory research

$\mathcal{H}$: Do certain word classes change more than others?

utability effect (Gentner 1981)

Cognitive Science, 2022

VERB METAPHORIC EXTENSION UNDER SEMANTIC STRAIN

Daniel King & Dedre

Northwestern Univ

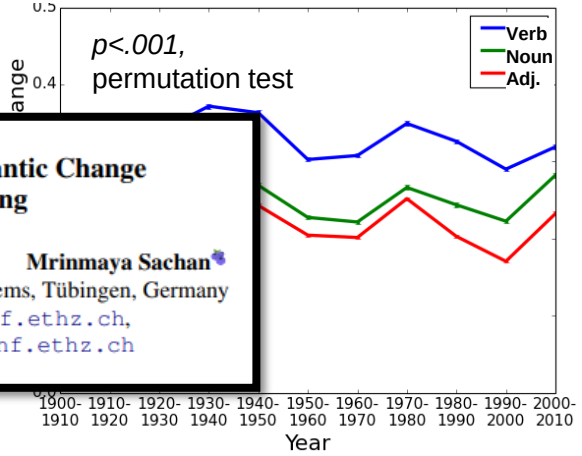
NOTE: This is an unpublished preprint c

tion between object-reference concepts and relati

Slangvolution: A Causal Analysis of Semantic Change and Frequency Dynamics in Slang

Daphna Keidar* Andreas Opedal* Zhijing Jin Mrinmaya Sachan
ETH Zürich, Max Planck Institute for Intelligent Systems, Tübingen, Germany
dkeidar@ethz.ch, andreas.opedal@inf.ethz.ch,
zjin@tue.mpg.de, mrinmaya.sachan@inf.ethz.ch

Word-class effect (Dubossarsky et al. 2016)

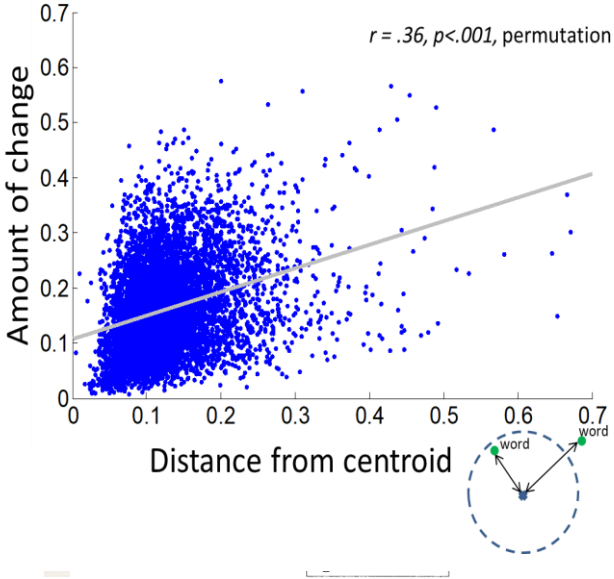


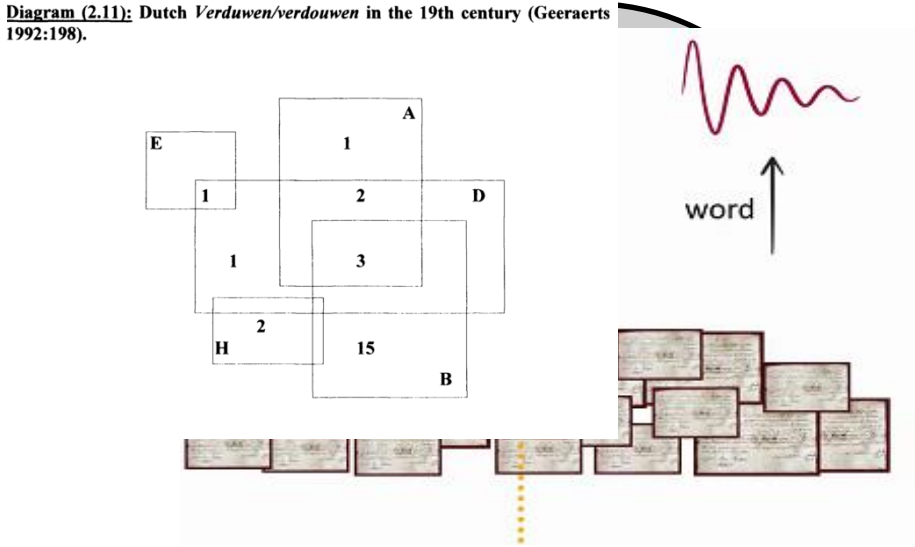
Dubossarsky, Weinshall, & Grossman, *Lingue e linguaggio*, 2016

Testing linguistic theories

$\mathcal{H}$: can category membership explain semantic change?

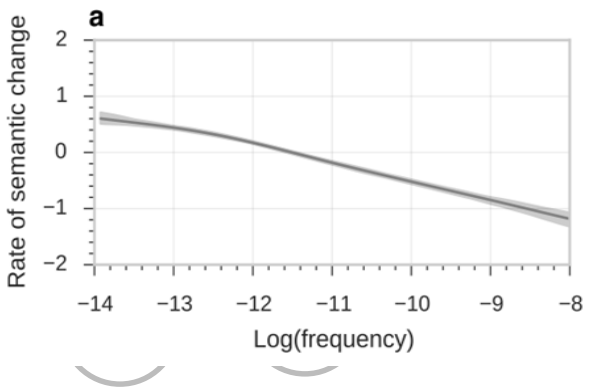
Law of prototypicality

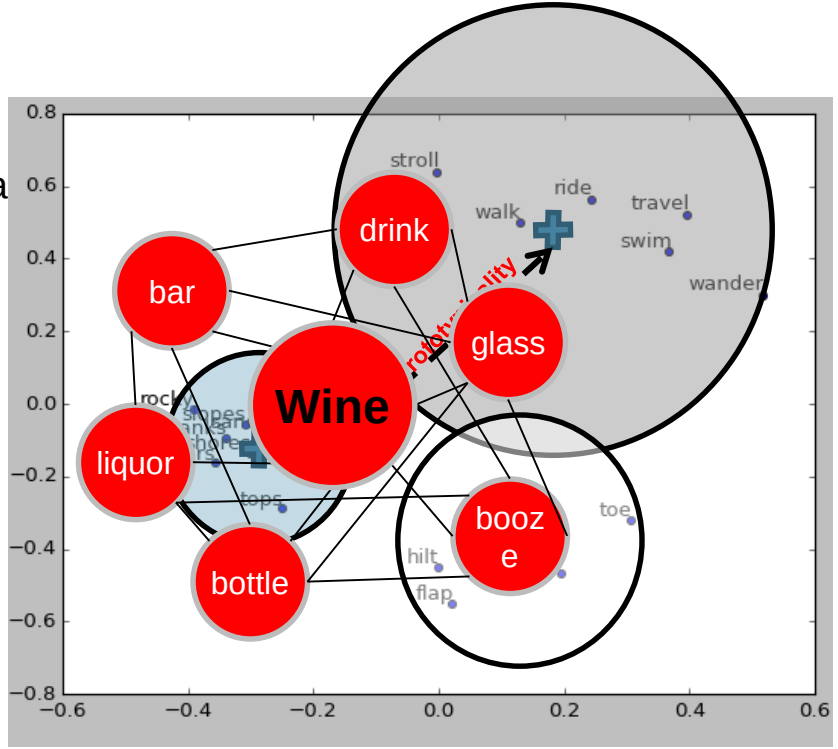




Critical analysis: laws of semantic change

- Law of Prototypicality (Dubossarsky et. al.
- Law of Innovation (Polysemy, Hamilton et. a
- Law of Confo (Frequency, Hamilton et.





Critical analysis: laws of semantic change

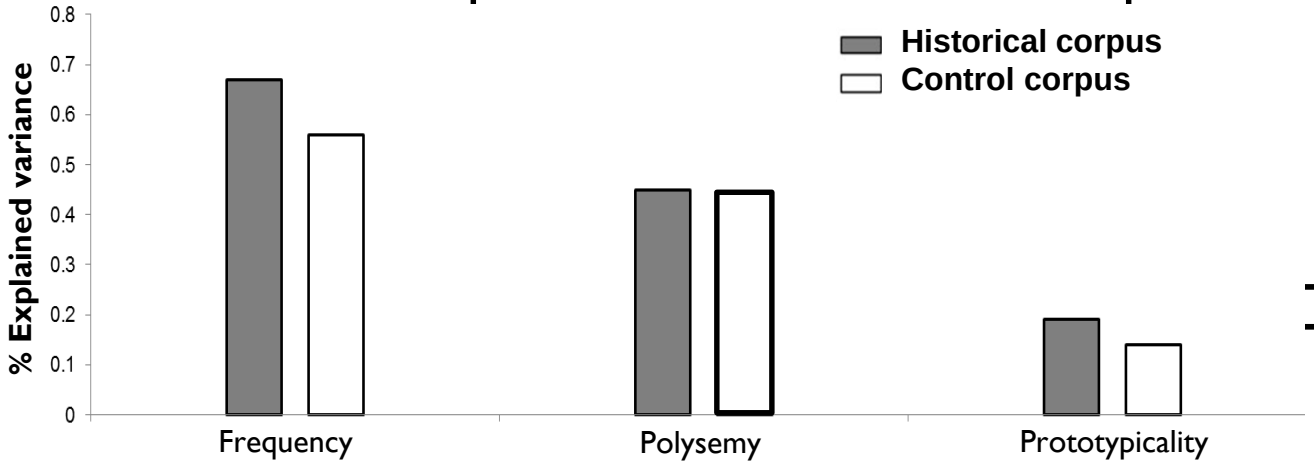
- Law of Prototypic
- Law of Innovatio
- Law of Conformit

Theoretically

In a control condition:

$$\mathbb{E}[x_{t^1}] = \mathbb{E}[x_{t^2}] = \mathbb{E}[x]$$

So, semantic change of a word is,



The search for laws of semantic change continues

Revisiting Statistical Laws of Semantic Shift in Romance Cognates

The return of the “law of conformity”
Good research is one that steers a discussion

	Coe			
Intercept	0.00	0.03	0.00	1.00
FREQ$_{lat}$	-0.08	0.04	-1.82	0.07
POLY$_{lat}$	**0.10**	**0.04**	**2.28**	**0.02**
LEN$_{lat}$	**-0.21**	**0.03**	**-6.29**	**0.00**
FREQ$_{rom}$	**-0.54**	**0.03**	**-18.40**	**0.00**
NORM$_{rom}$	–	–	–	–
EDIT	**0.13**	**0.03**	**4.07**	**0.00**

Table 3: Results of regression analysis on distance scores of French–Spanish cognate pairs ($N = 794$, $Adj.R^2 = 0.35$). NORM$_{rom}$ was kept out by model selection methods.

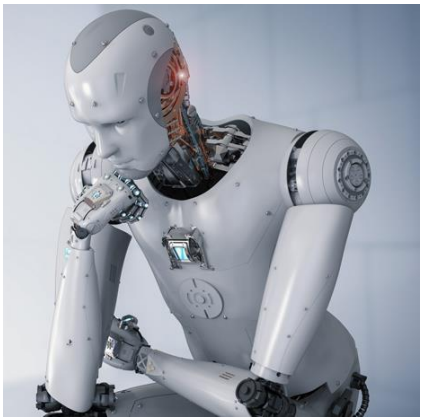
	pair	$Adj.R^2$	N
French	French–Italian	0.29	812
	French–Spanish	0.35	794
	Italian–Spanish	0.35	842
Italian	French–Italian	0.29	812
	French–Spanish	0.33	794
	Italian–Spanish	0.38	842
Spanish	French–Italian	0.27	812
	French–Spanish	0.35	794
	Italian–Spanish	0.39	842

Table 4: Adjusted R-squared for respective language pairs in different embedding spaces.

Summary

If you are a linguist

If you are a historian, sociologist or interested in societal changes

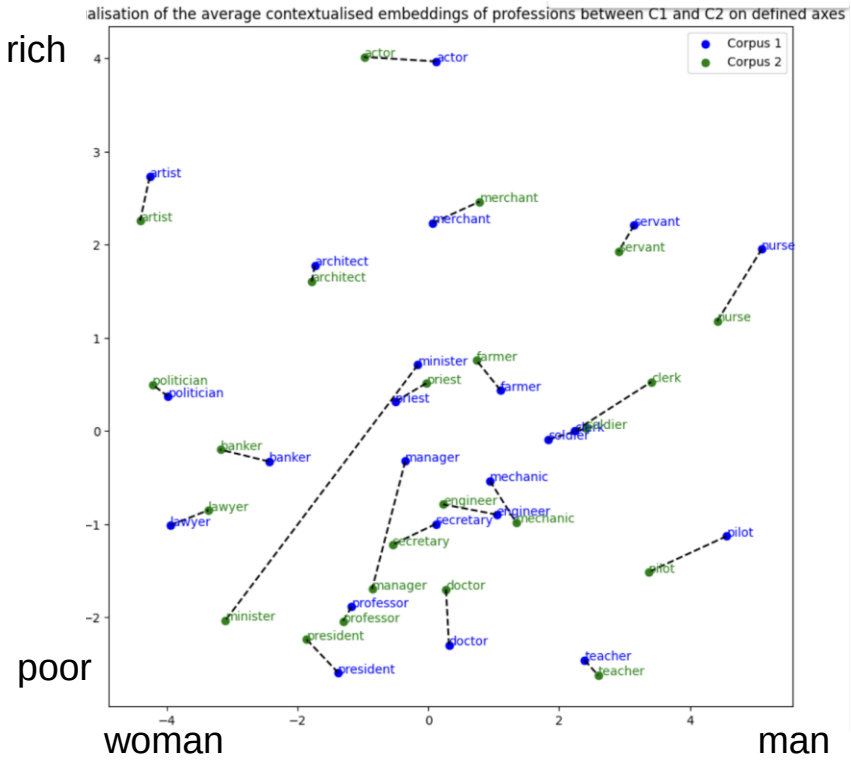


If you work on AI

Hands on

https://colab.research.google.com/drive/1NSCdYExjLjSV15wvQ-msCyoShZKpkAeu?usp=sharing

Hands on


alisation of the average contextualised embeddings of professions between C1 and C2 on defined axes
Corpus 1
Corpus 2
rich
poor
woman
man
actor
actor
artist
artist
merchant
merchant
servant
servant
nurse
nurse
architect
architect
minister
minister
farmer
farmer
priest
priest
politician
politician
clerk
soldier
soldier
banker
banker
manager
mechanic
engineer
lawyer
lawyer
secretary
secretary
pilot
pilot
manager
doctor
professor
professor
president
president
doctor
teacher
teacher