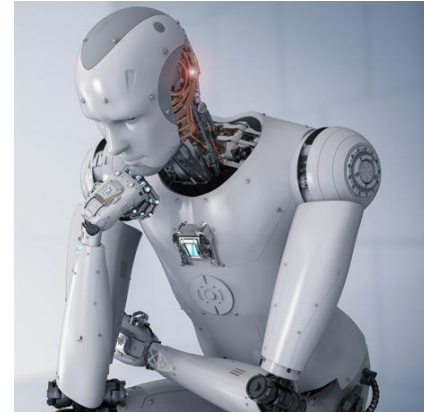# Insights & Applications

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

Why change matters?

If you are a linguist

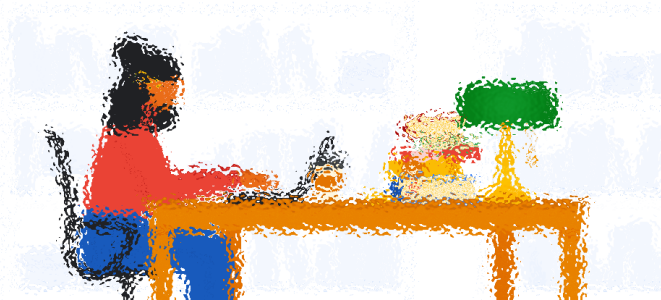If you are a historian, sociologist or interested in societal changes

If you work on AI

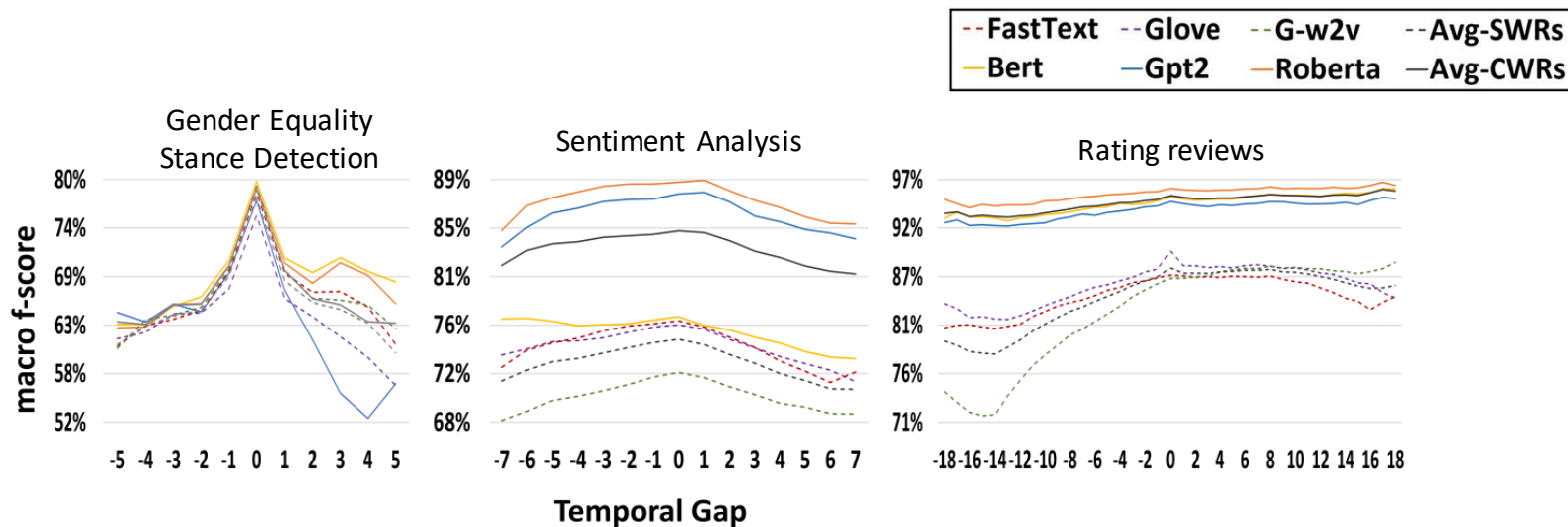2

Why change matters?

If you are a linguist

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

If you work on AI

Models age: Performance drop with temporal gap

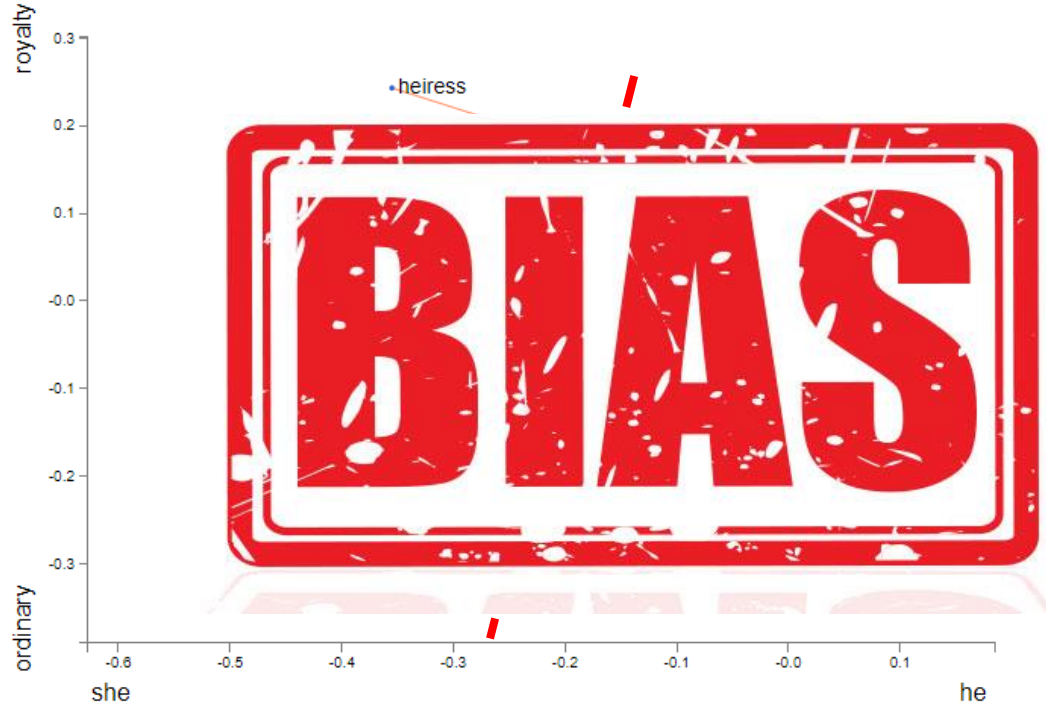From Alkhalifa et al. 2023

Why change matters?

If you are a linguist

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

If you work on AI

Models are biased

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

Democratizing AI

Poor support
Good Support
N/A

linguistic

gender

demographic

Academic papers

Soap operas

Twitter

Mental health

L2 learners

MODEL

Resolution, act

Cell, space, objec

Husband, love
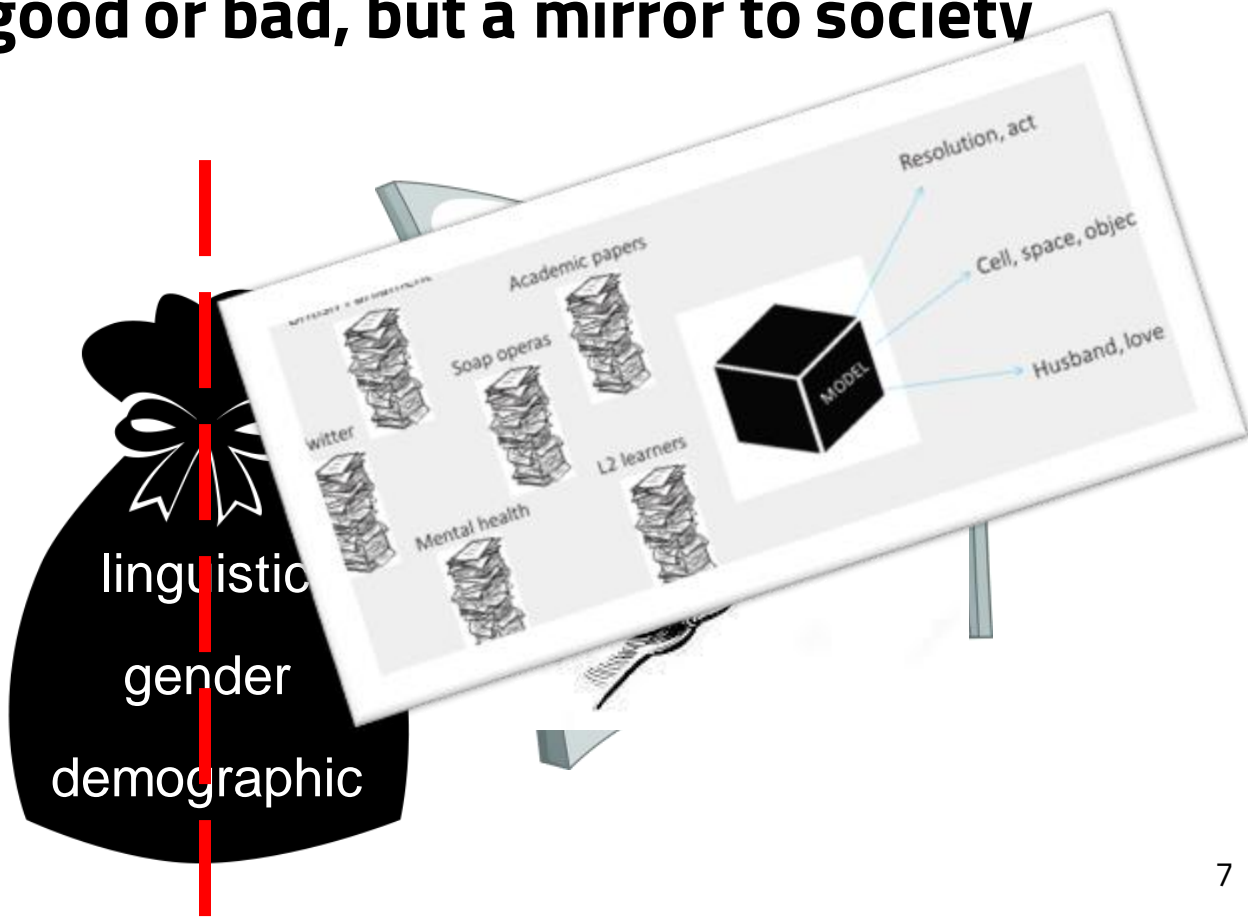
Study how bias changes over time & across domains

word

word

broadcast[1880]

broadcast[1920]

broadcast[1960]

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°	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°	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

9

In the cultural domain

Why change matters?

If you are a linguist

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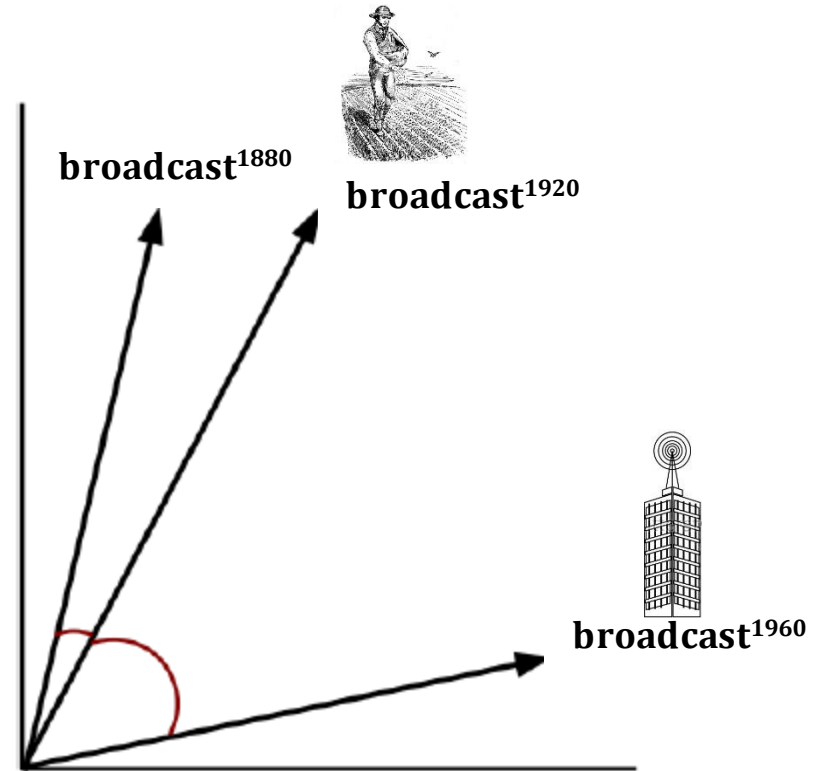

If you work on AI

Exploratory research

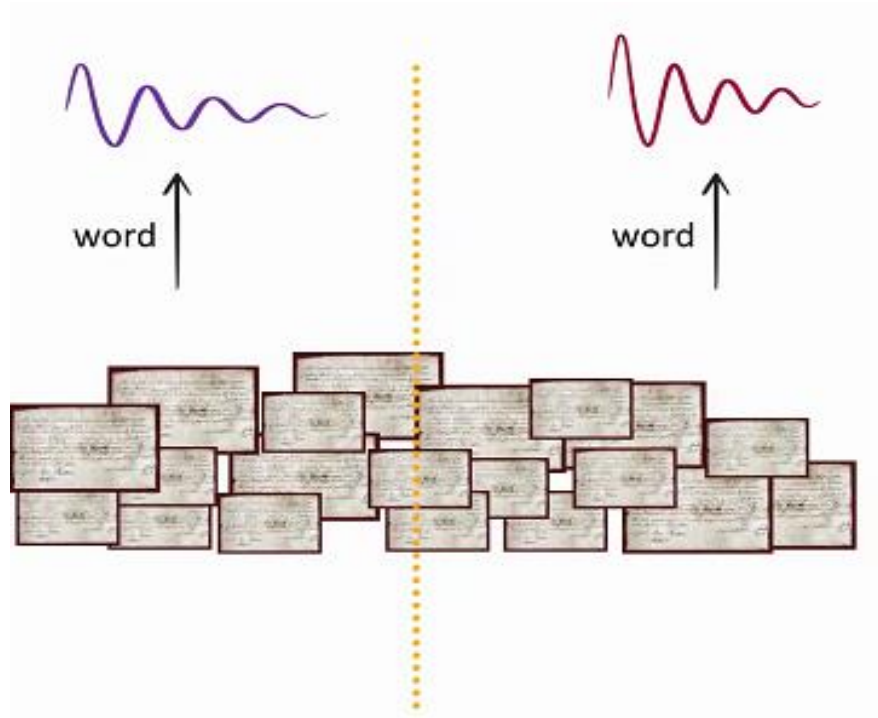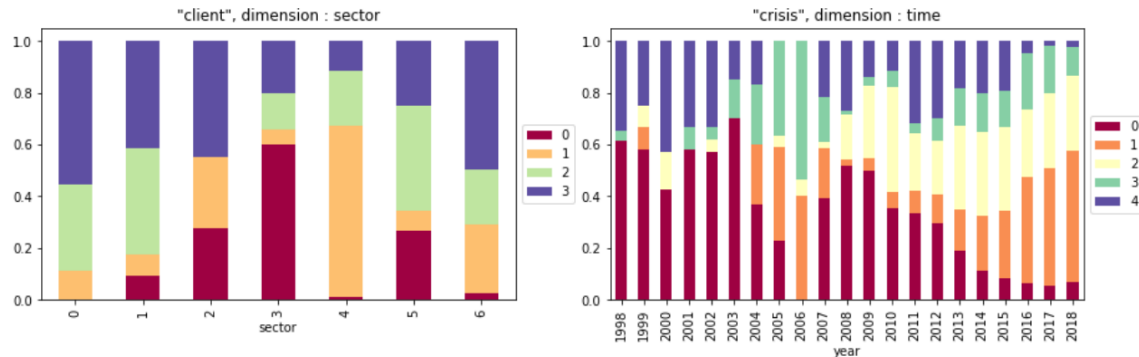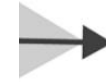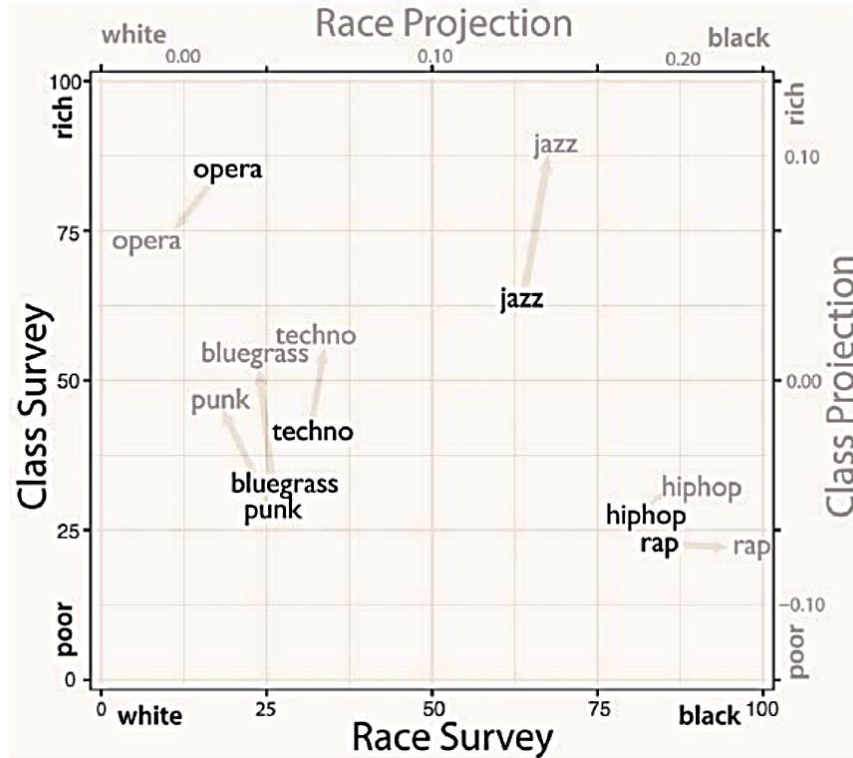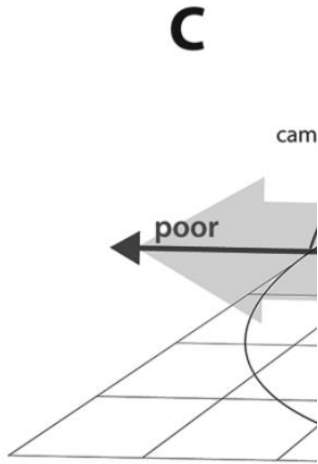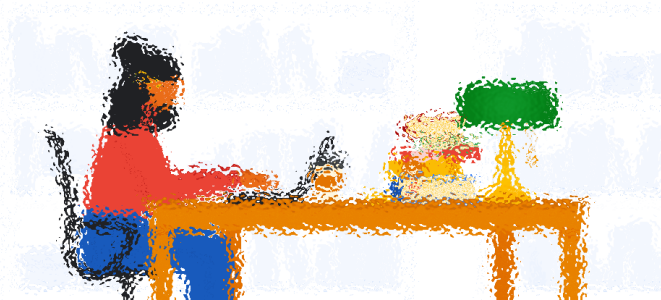$\mathcal{H}$: Do certain word classes change more than others?

mutability effect (Gentner 1981)

Cognitive Science, 2022

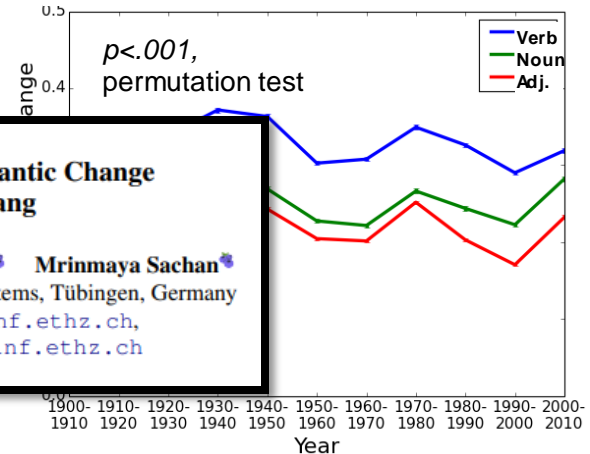VERB METAPHORIC EXTENSION UNDER SEMANTIC STRAIN

Daniel King & Dedre

Northwestern Univ

NOTE: This is an unpublished preprint

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

tion between object-reference concepts and relational concepts. Object-reference

Word-class effect (Dubossarsky et al. 2016)

p<.001,
permutation test

Verb
Noun
Adj.

1900- 1910- 1920- 1930- 1940- 1950- 1960- 1970- 1980- 1990- 2000-
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010
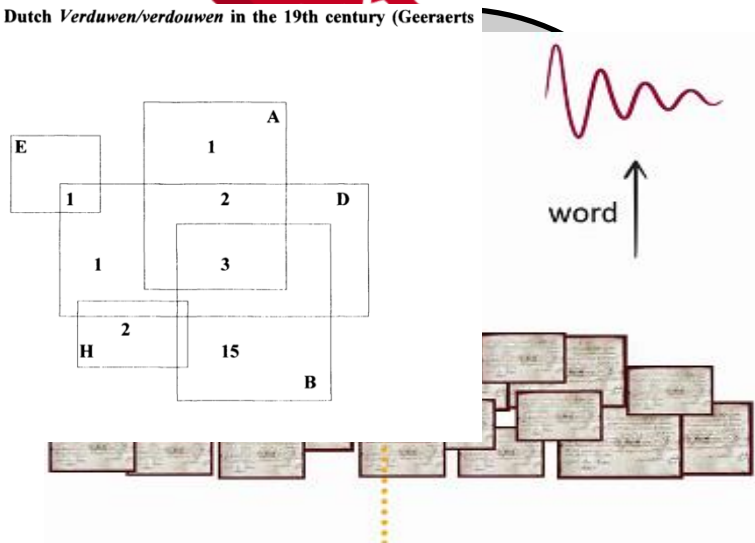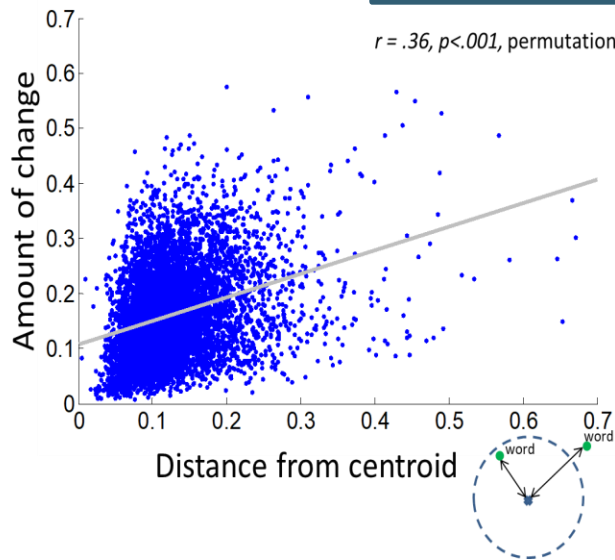Year

Testing linguistic theories

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

Law of prototypicality

r = .36, p<.001, permutation test

Amount of change

Distance from centroid

Diagram (2.11): Dutch *Verduwen/verdouwen* in the 19th century (Geeraerts 1992:198).

word

Dubossarsky, Tsvetkov, Dyer, & Grossman, *Word Structure and Word Usage*, 2015

Critical analysis of our models: laws of semantic change

- Law of Prototypicality (Dubossarsky et. al.
- Law of Innovation (Polysemy, Hamilton et. a
- Law of Confo river (Frequency, Hamilton et.
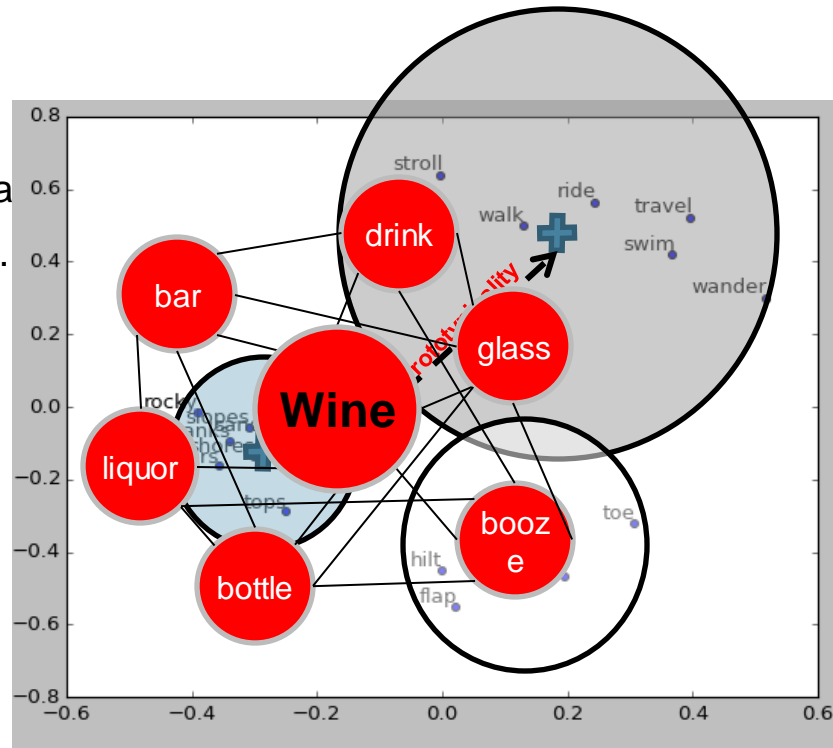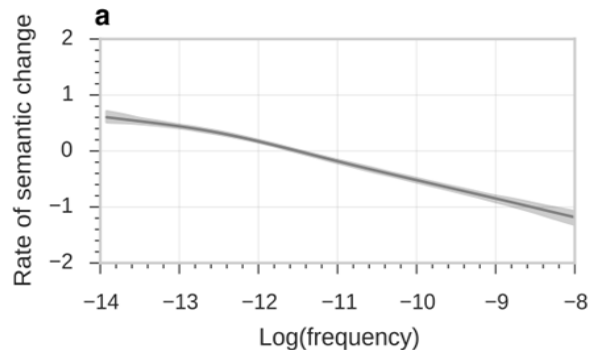
Dubossarsky, Grossman, & Weinshall, *EMNLP*, 2017

Critical analysis of our models: laws of semantic change

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

Theoretically
In a control condition:

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

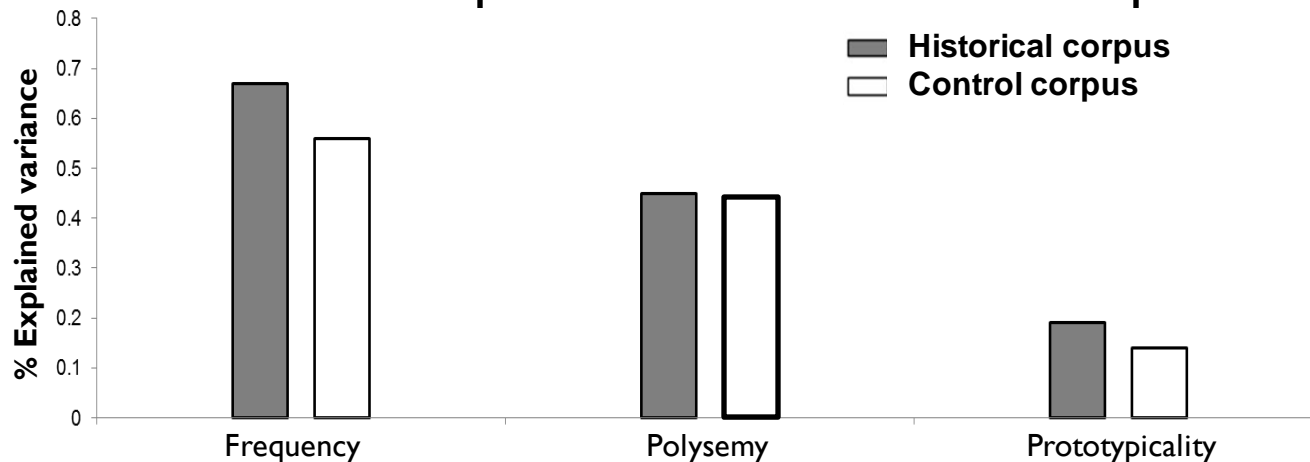So, semantic change of a word is,

Historical corpus
Control corpus

% Explained variance

Frequency   Polysemy   Prototypicality

Dubossarsky, Grossman, & Weinshall, *EMNLP*, 2017

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.

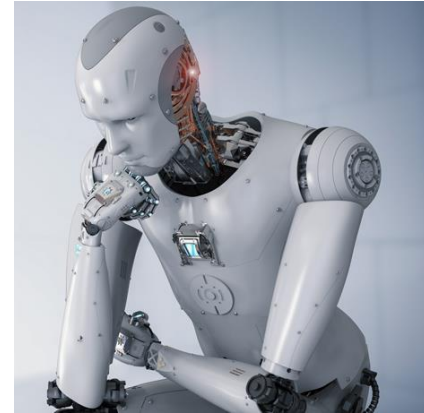|	...g space	...g...g...pair	$Adj.R^2$	N
French		French–Italian	0.29	812
		French–Spanish	0.35	794
		Italian–Spanish	0.35	842
Italian		French–Italian	0.29	812
		French–Spanish	0.33	794
		Italian–Spanish	0.38	842
Spanish		French–Italian	0.27	812
		French–Spanish	0.35	794
		Italian–Spanish	0.39	842

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

16

Summary

If you are a linguist

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

If you work on AI

17