Introduction to Semantic Change Research & its foundations in Distributional Semantics

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

Hello!

Haim Dubossarsky
[1]Queen Mary University of London (QMUL)
[2]Language Technology Lab, University of Cambridge
[3]The Alan Turing Institute
h.dubossarsky@qmul.ac.uk

Researcher @ Change is Key!

Hello!

Pierluigi Cassotti

Researcher @ Change is Key!
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg (Sweden)
pierluigi.cassotti@gu.se

Hello!

Roksana Goworek

2nd year PhD student
Queen Mary University of London (QMUL)
r.goworek@qmul.ac.uk

Schedule & expectations



- Introduction & foundations
- Evaluation
- Models of change
- Research applications
- LSC in your research

Some familiarity with NLP

Worked with text before

Interest in semantic change

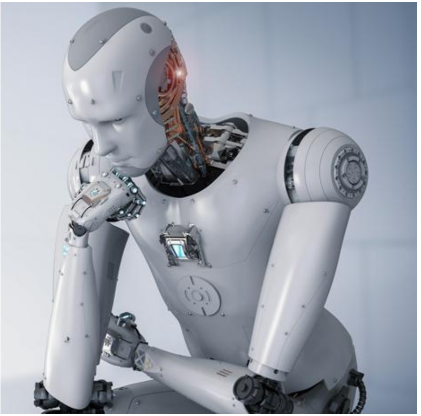
Have your ideas for research applications

We are here for **YOU**: approachable anytime in ESSLLI

Motivation: Why change (and variation) matters

If you are a linguist

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



If you work on AI

Semantic change is fundamental to human language

He was an **awesome** leader!

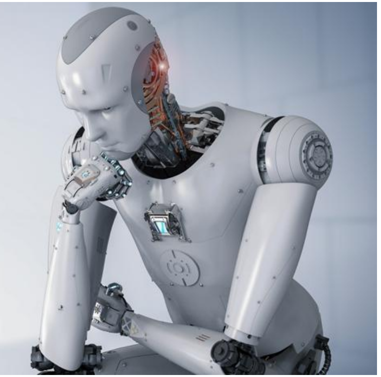
He was an **awesome** leader!

time

Motivation: Why change (and variation) matters

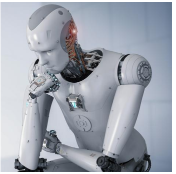
If you are a linguist

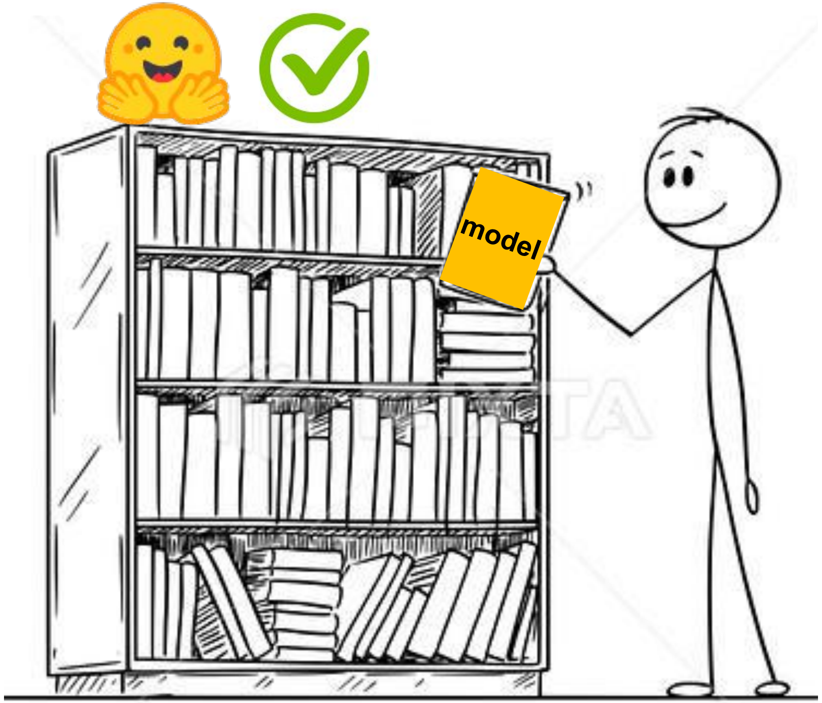
If you work on AI



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

Current modus operandi





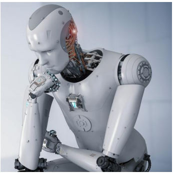
Works well iff:

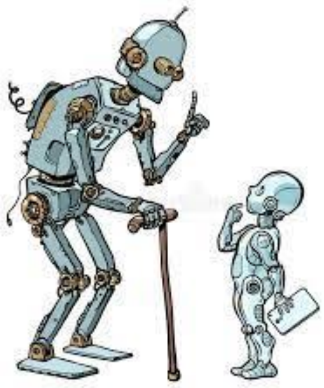
- Domain is very similar to the training dataset
- Finetuning on suitable dataset is possible

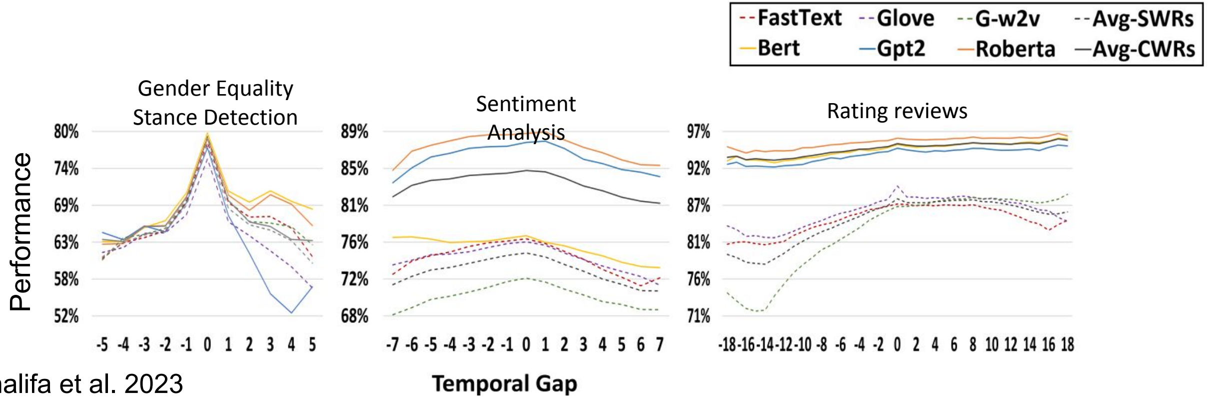
Problematic in cases of:

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

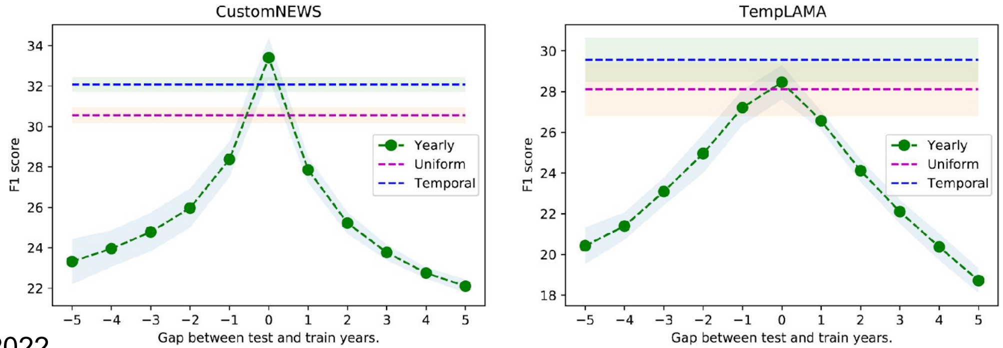
Models do not age well







From Alkhalifa et al. 2023



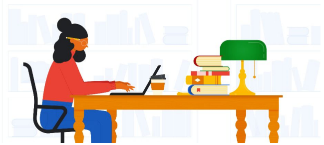
Dhingra et al. 2022

Motivation: Why change (and variation) matters

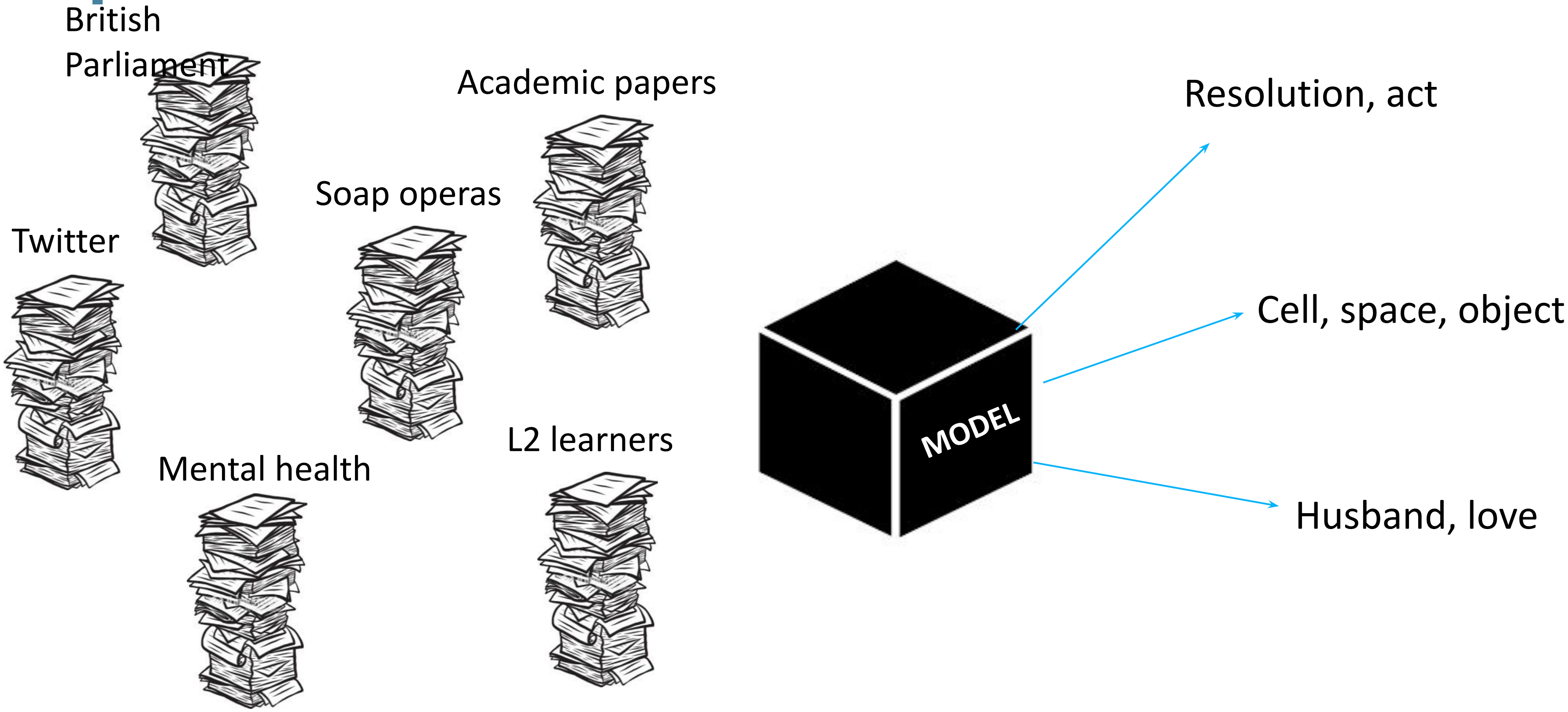
If you are a linguist

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

If you work on AI



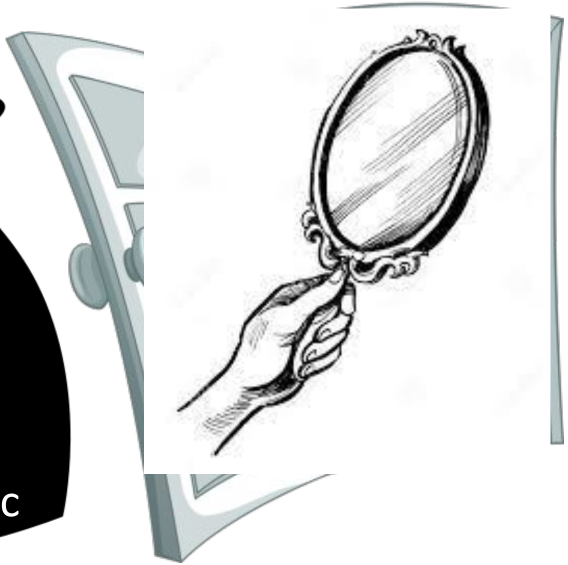
Text is a litmus paper of our societies and culture



Biases in Language Models

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

Bias is neither good or bad



- Introduction
- Evaluation
- Models of change
- **Research applications**
- LSC in your research

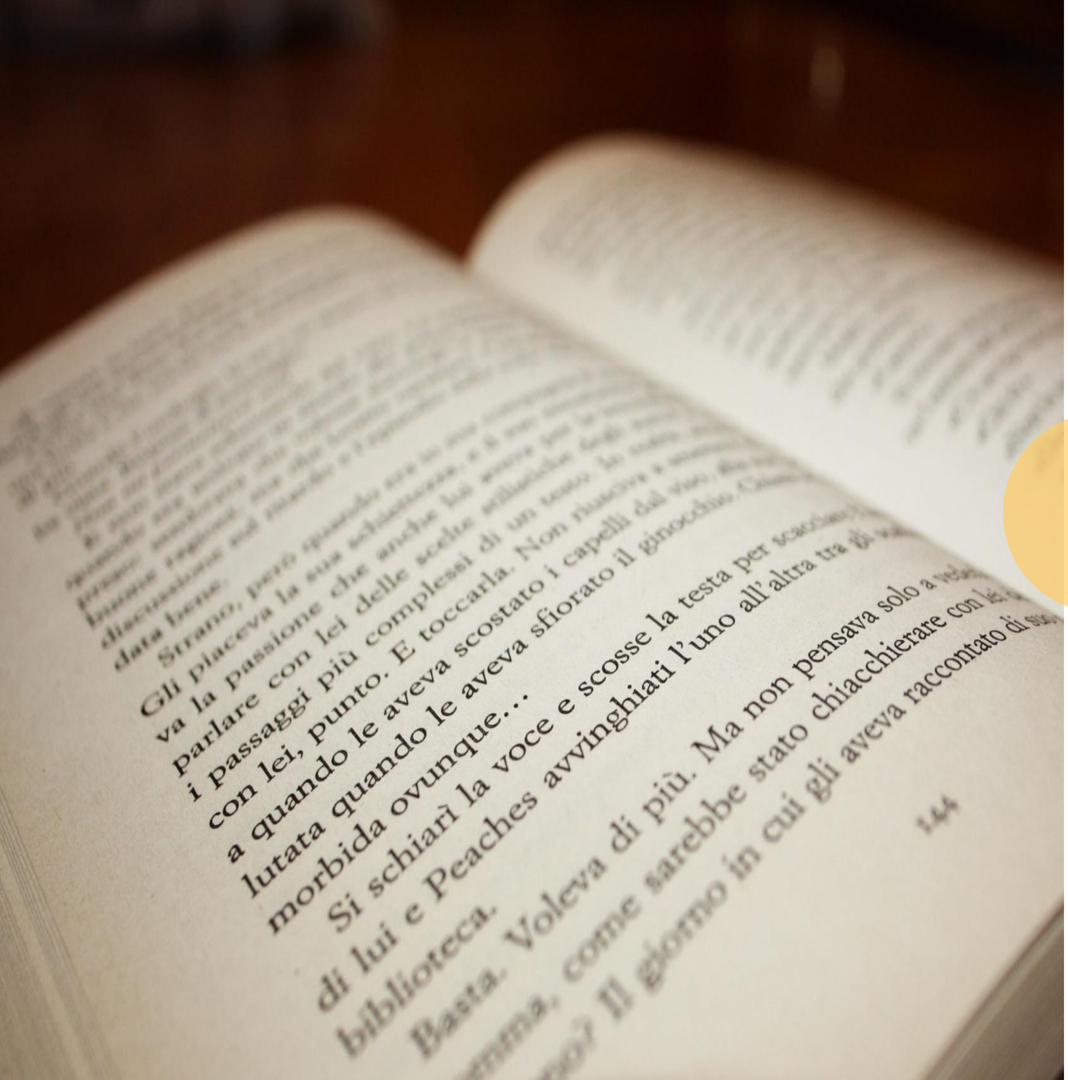
Why change (and also variation) matters?

If you are a linguist

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

If you work on AI

How did we do it before?



By close reading!

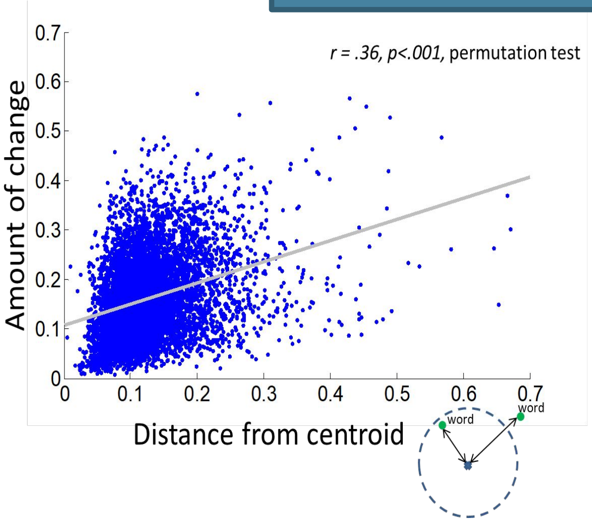
How do we do it now?

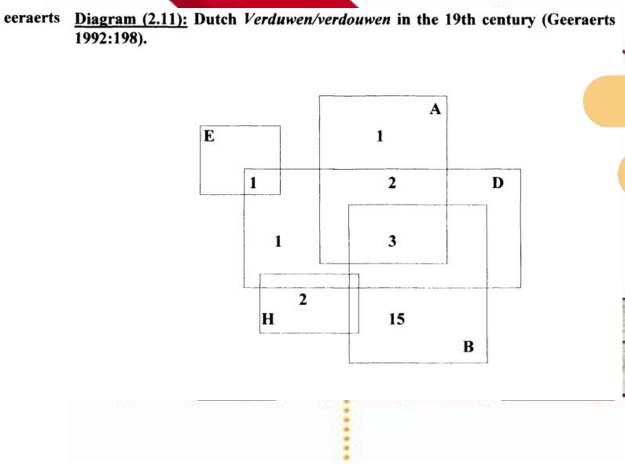
Advance linguistic research

𝒳: can category membership explain semantic change?

Dirk Geeraerts (1987, 1992)

Law of prototypicality





Manners of motion

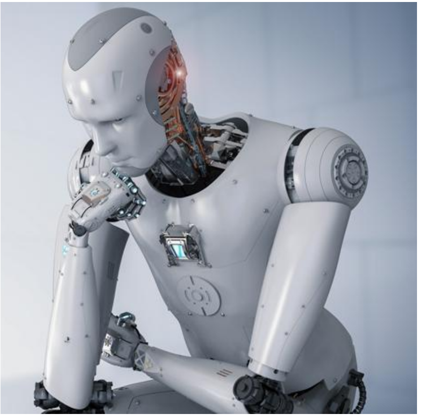
- Introduction
- Evaluation
- Models of change
- **Research applications**
- LSC in your research

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

Questions on the motivation?

If you are a linguist

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



If you work on AI

Change is Key!

The study of contemporary and historical societies

https://www.changeiskey.org/

Change is Key!
Facts

6	years
6	partner universities
4	Members from 4 countries
6	Countries, with advisors
17	People including a SE & PM
33.5	MSek from RJ + 10.5MSek from the University, Faculty, Dept

Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung

Queen Mary
University of London

KU LEUVEN

LINKÖPING UNIVERSITY

LUNDS UNIVERSITET

GÖTEBORGS UNIVERSITET

Activities in the community

EACL Tutorial
Malta, Mar 21, 2024

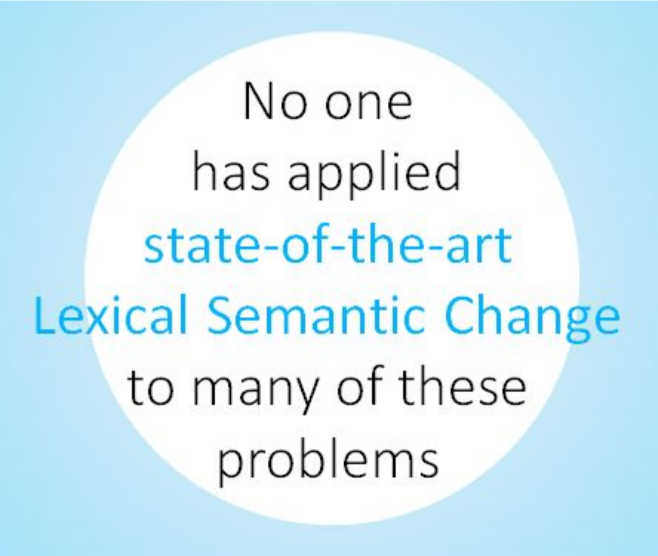
Evolang workshop
Madison US, May 18, 2024

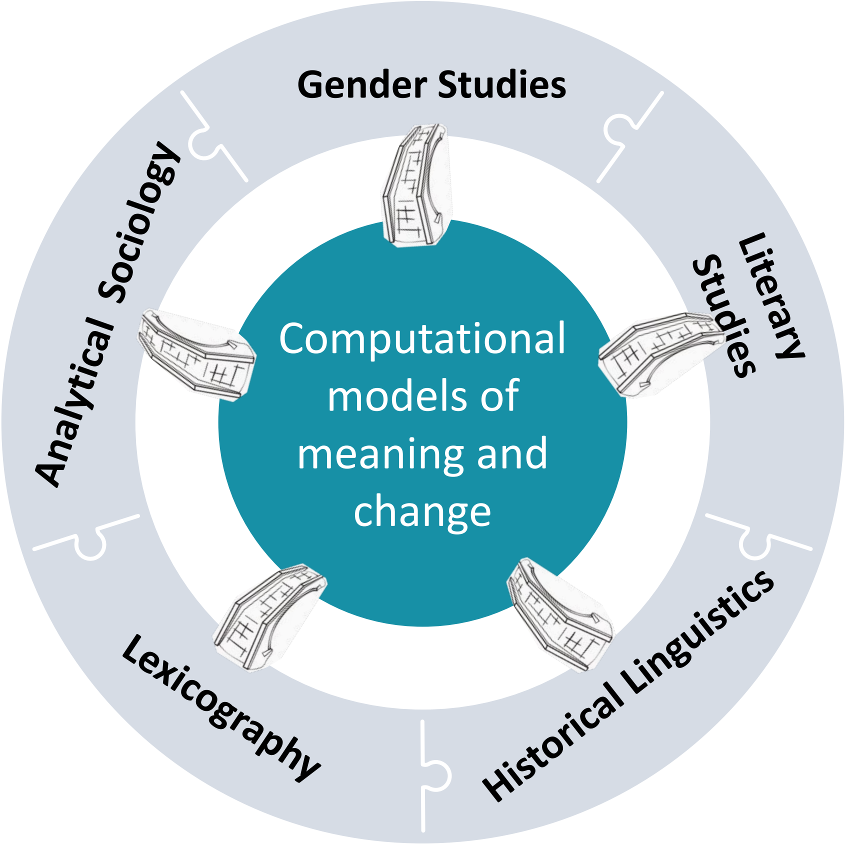
LChange'24
Bangkok, Aug 15, 2024

Change is Key!
Goals

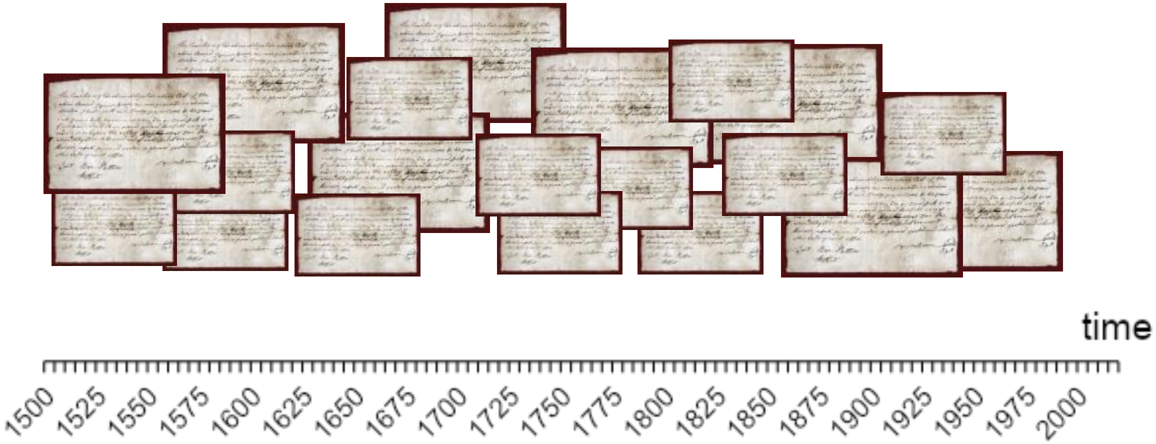
Language level change

Societal level change





How does semantic change research work?



Day 1: Comparing Representations

https://colab.research.google.com/drive/1JBW5pQ3-HxilyiJuZMREG8uBzzkQwuPS?usp=sharing

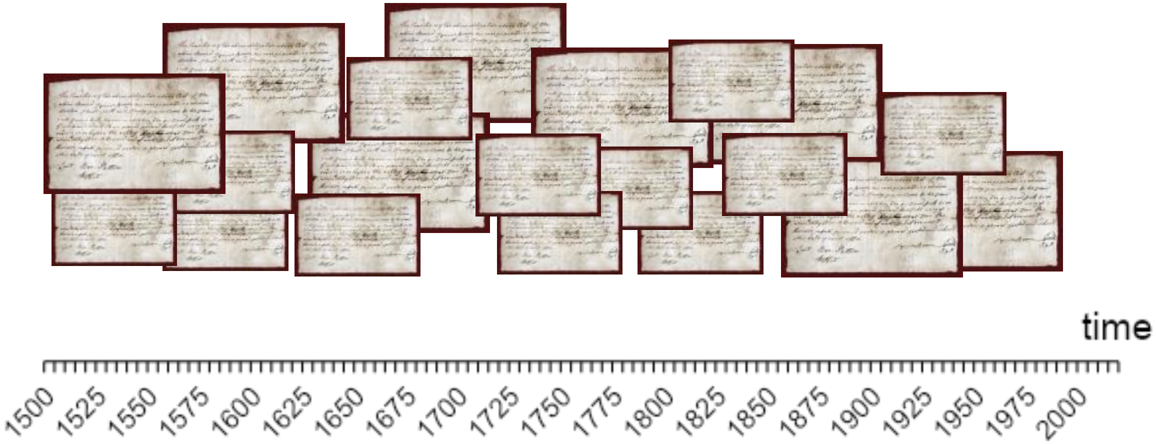
We need a model for ....

it's everywhere,
its effects can be felt,
but you cannot see or touch it.

Meaning change

-> meaning is the dark matter of language

How does semantic change research work?



Today | **Wednesday (by Pierluigi)**

Models of meaning + Models of change

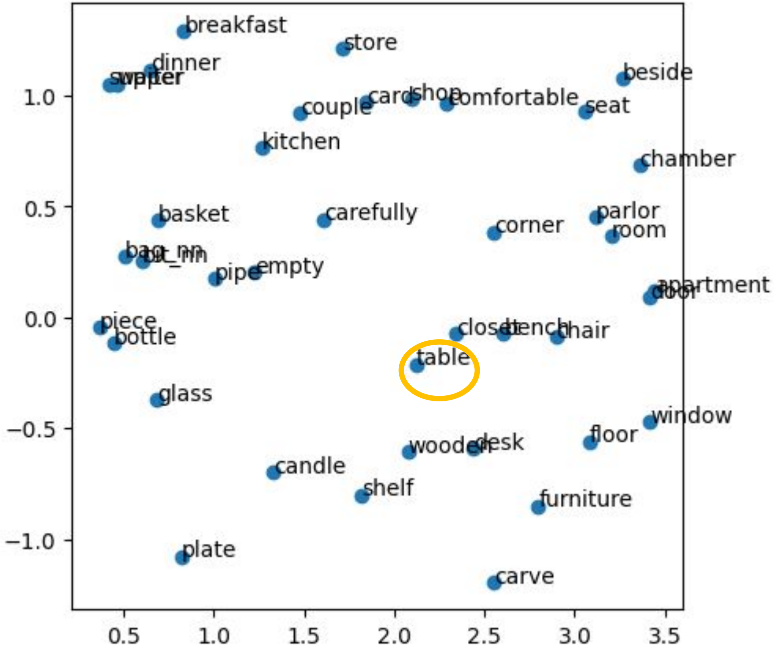
The distributional hypothesis

Words occurring in similar contexts tend to have similar meanings (Zellig Harris, 1954)

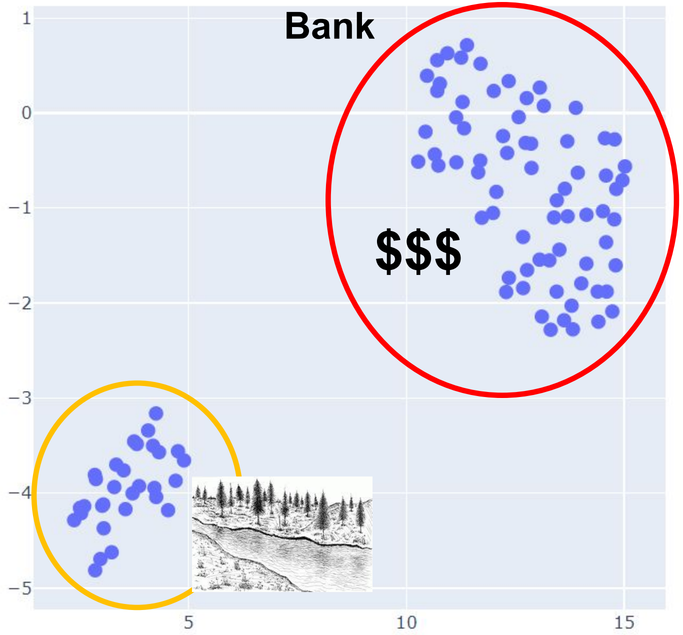
You shall know a word by the company it keeps (Firth, J. R. 1957:11)

Meaning "perceived" by computational models

Two operationalizations of the distributional hypothesis



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



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

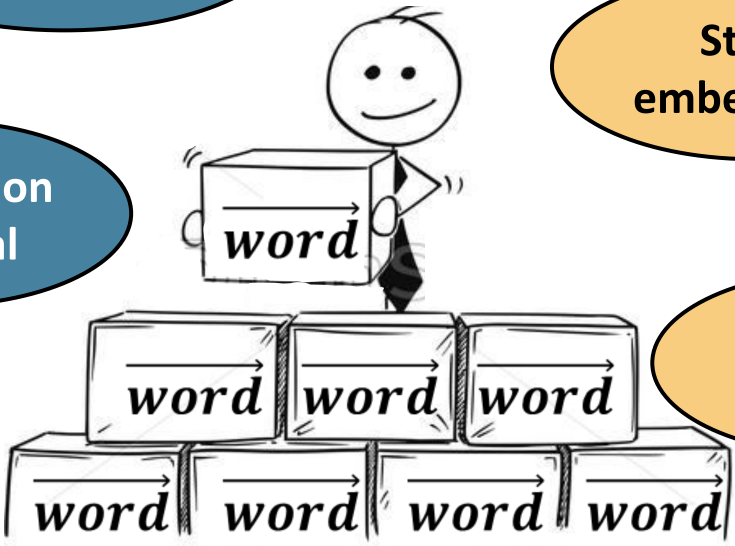
DH is the root of all computational models of meaning

Machine translation

Sentiment analysis

Information retrieval

Named Entity Recognition



Explicit models

Static embeddings

Contextualized models

Topic models

Count-based models

Simple co-occurrence models within a context window
Very sparse

Years

Context words

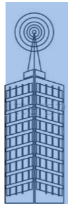
CONTEXT	ALL	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s
COMPUTER	171						1	75	67	28
BIG	170	2	26	8	2	7	31	43	29	22
PIE	148	4	3	22	18	16	29	31	17	8
TREE	94	6	18	16	17	15	9	9	1	3
ADAM	59	8	12	9	7	9	8	3	3	
JOBS	57						1	14	23	19
TREES	51	4	10	18	5	6	3	4	1	
JUICE	45	1	2		3	5	9	7	11	7
MACINTOSH	43							24	13	6
IBM	41							24	17	

Taken from https://corpus.byu.edu/

Count-based models

Years

Context words

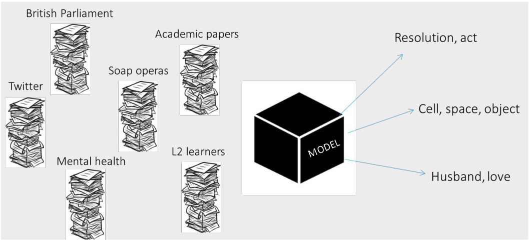


CONTEXT	ALL	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
RADIO	454									1			15	42	98	69	52	50	54	38	35
TELEVISION	161												4	3	6	28	15	24	33	29	19
NEWS	161					1	[illegible]	[illegible]	1			3	3	7	27	[illegible]	[illegible]	30	19	18	20
STATION	115						[illegible]	[illegible]					4	11	35	[illegible]	[illegible]	10	9	9	5
STATIONS	111						[illegible]	[illegible]					5	13	23	[illegible]	[illegible]	7	12	10	12
TV	79						[illegible]	[illegible]								[illegible]	[illegible]	7	11	28	19
LIVE	77						[illegible]	[illegible]								[illegible]	[illegible]	15	18	16	18
NETWORK	61						[illegible]	[illegible]						5	14	[illegible]	[illegible]	6	5	10	14
SPEECH	61						[illegible]	[illegible]					2	18	16	[illegible]	[illegible]	4	4	2	2
PROGRAM	60						[illegible]	[illegible]					1	15	15	[illegible]	[illegible]	7	3	3	1
SCATTERED	44				3	5	[illegible]	[illegible]	7	4	7	5	2	1	1	[illegible]	[illegible]	1			
TONIGHT	43						[illegible]	[illegible]					3	5	15	[illegible]	[illegible]	1	1	3	1
ADDRESS	41						[illegible]	[illegible]					4	11	15	[illegible]	[illegible]	1	4		1
THROUGHOUT	37				1	2	[illegible]	[illegible]			2		7	6	1	[illegible]	[illegible]	3	2	2	3
LOCAL	37						[illegible]	[illegible]			1		4			[illegible]	[illegible]	5	3	8	5
NBC	36						[illegible]	[illegible]					1	8	5	[illegible]	[illegible]	5	4	8	4
MEDIA	36						[illegible]	[illegible]								[illegible]	[illegible]	4	6	10	11
SUNDAY	36						[illegible]	[illegible]						5	14	[illegible]	[illegible]	1	3	1	1
WEEKLY	34						[illegible]	[illegible]					1	4	14	[illegible]	[illegible]	1	2	3	1
SPREAD	34				2		[illegible]	[illegible]	3	2	7	12	4			[illegible]	[illegible]	1			1
MOSCOW	31												1	3	10	9	4	2	1		1
MATERIAL	31												1	1	20	1	1	4		3	
CABLE	30												1						6	12	11
MESSAGE	30													4	7	5	3	6	3		2
SOWN	29			1	2	2	7	1	2	7	2	1	2		1	1					

Count-based models

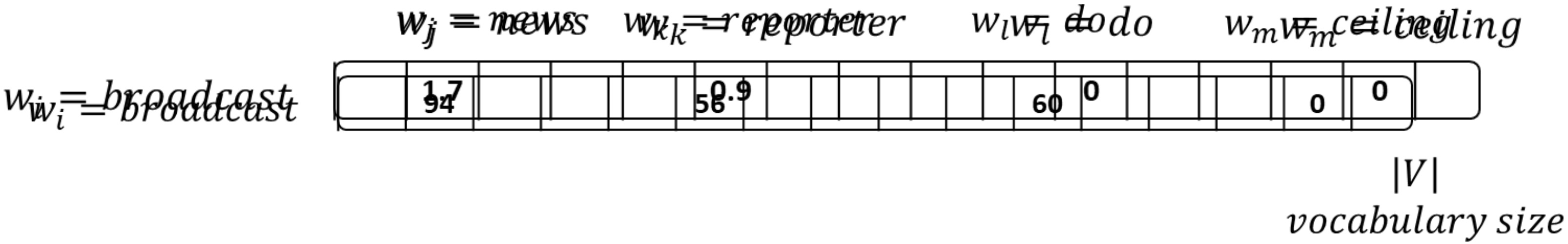
- Can be computed on a very small "local" corpus
  -> Highly dependent and reflects the meaning of the corpus/domain
  -> More apparent here as modern LM require massive corpus for pre-training, therefore more "global" in nature.

- **Problem**: Highly skewed for frequent collocates
  - Prepositions, function words (stopwords)
  - Solution: PPMI



Positive Pointwise Mutual Information (PPMI)

- Co-occurrence models **with a twist**
  - **Twist**: Mutual information measures the strength of association between the target word and its co-occurring words



- **Learn associativity by informativity**

Positive Pointwise Mutual Information (PPMI)

- Only "strong" co-occurring words are retained, hence "positive" PMI

$$\text{PPMI}(w,c) = \max(\log_2 \frac{P(w,c)}{P(w)P(c)}, 0)$$

Count(w,context)

	computer	data	pinch	result	sugar
apricot					
apple					
digital					
informatio					

p(w,context)

	computer	data	pinch	result	sugar	p(w)
apricot						
apple						
digital						
informatio						
p(context)						

PPMI(w,context)

	computer	data	pinch	result	sugar
appricot	-	-	2.25	-	2.25
apple	-	-	2.25	-	2.25
digital	1.66	0.00	-	0.00	-
information	0.00	0.57	-	0.47	-

Advantages of explicit models (count-based & PPMI)

$w_j = news$ $w_k = reporter$ $w_l = do$ $w_m = ceiling$

$w_i = broadcast$

	1.7				0.9					0				0	

- Enables a finer analysis of change (association level)
  - Long history in CL research: Stefanowitsch & Gries Collostructions Analysis (2003)

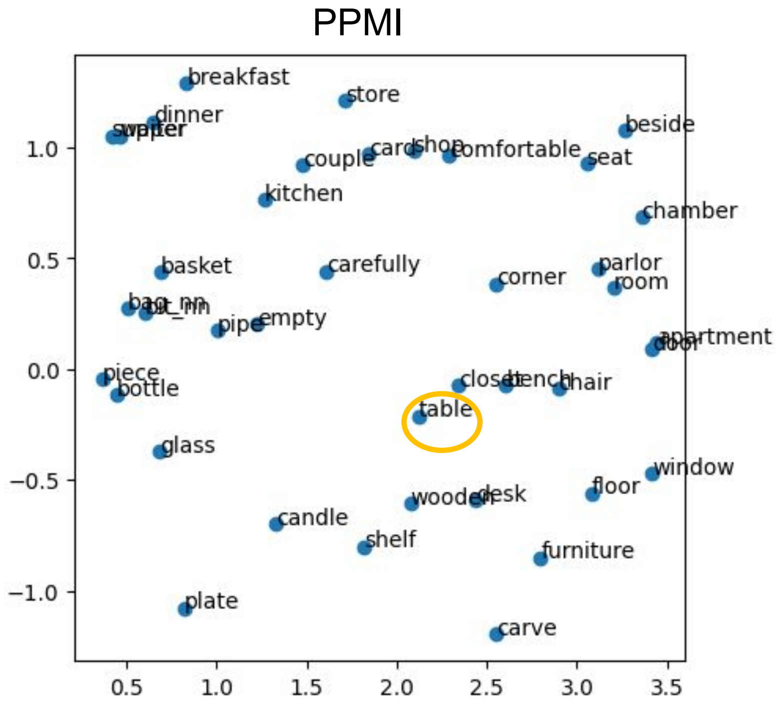
PPMI for *prime*[1990]
ministe
suspe
cut
numbe

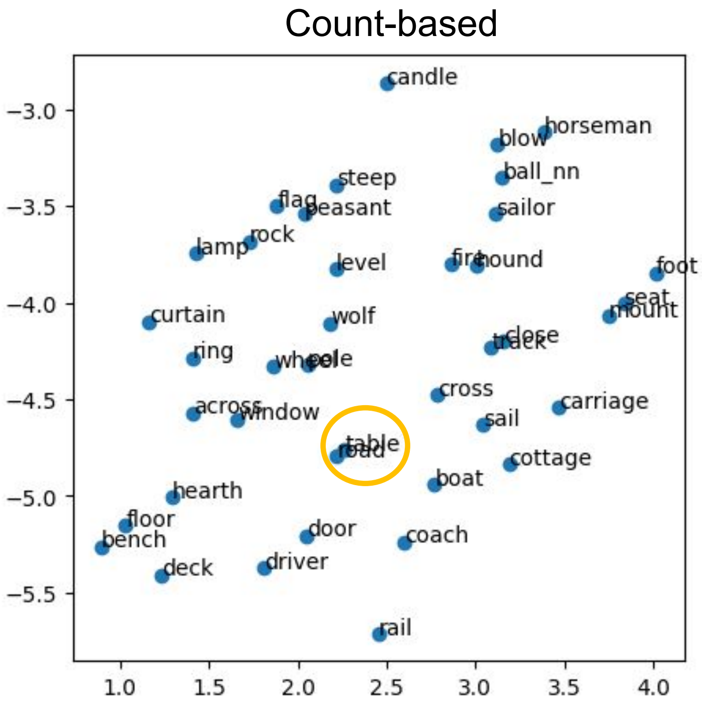
PPMI for *prime*[2010]	
minister	11.26
numbers	9.51
cut	10.1

PPMI for *heart*[medical]
attack
chest
pacemaker

PPMI for *heart*[standard]	
attack	13.4
emotion	**4.9**
central	**4.5**
warmth	**3.2**

PPMI vs. simple count-based



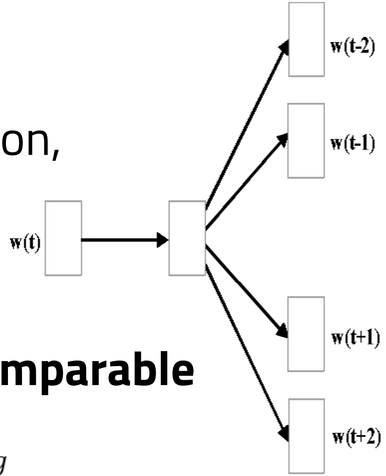


Implicit (predictive) models
Language Models (LM)

Predictive models (word2vec)

- Word2vec (Mikolov et al. 2013) is a Neural Network model
  - Other highly similar models exist (FastText, Glove)
  - Shallow network: 1 layer
  - Uses known NN machinery: MLM, objective function,

$$J_\theta = \frac{1}{T}\sum_{t=1}^{T} \sum_{-n\leq j\leq n, \neq 0} \log p\left(w_{t+j} \mid w_t\right)$$



- Vectors are opaque “implicit” & **vector spaces are incomparable**

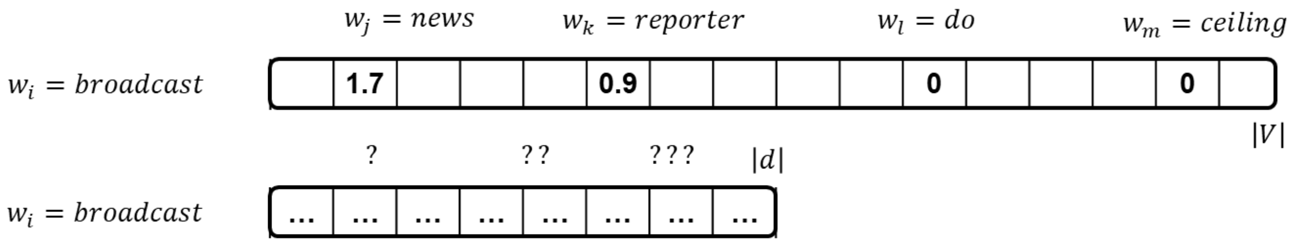
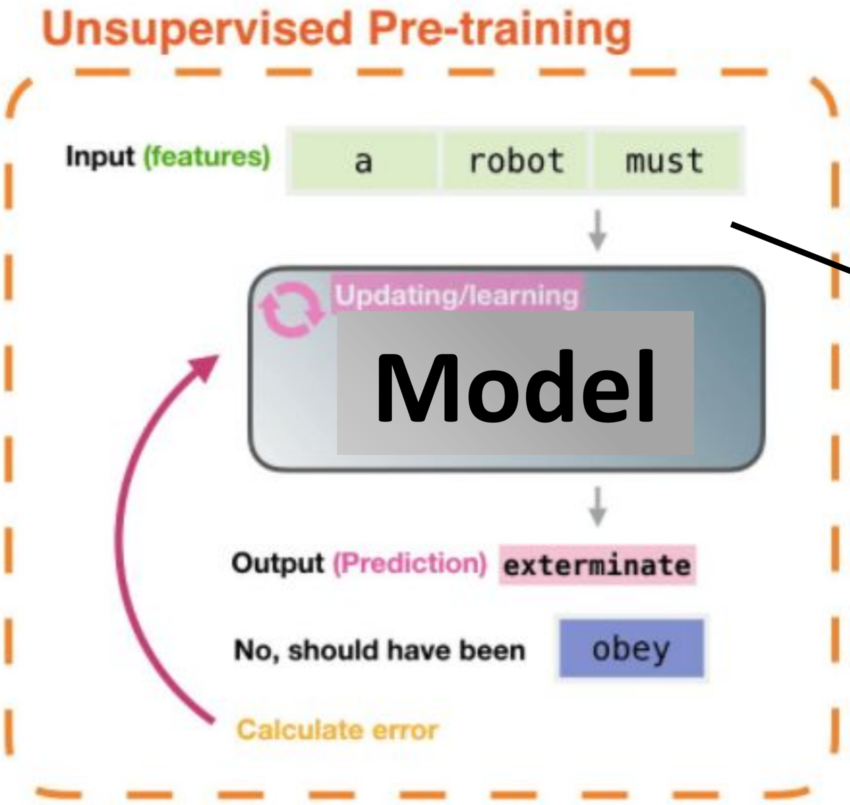


Image from Jay Alammar's blog

Implicit (predictive) models
Language Models (LM)

Training a Language Model



Correct output (label):

obey [redacted]

context [redacted]

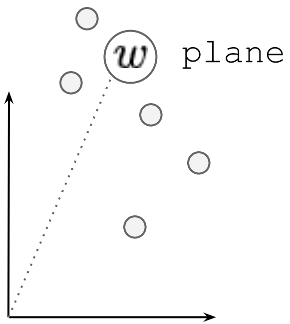
What is the closest vector?

exterminate [redacted]

Courtesy of Francesco Periti

Implicit (predictive) models
Language Models (LM)

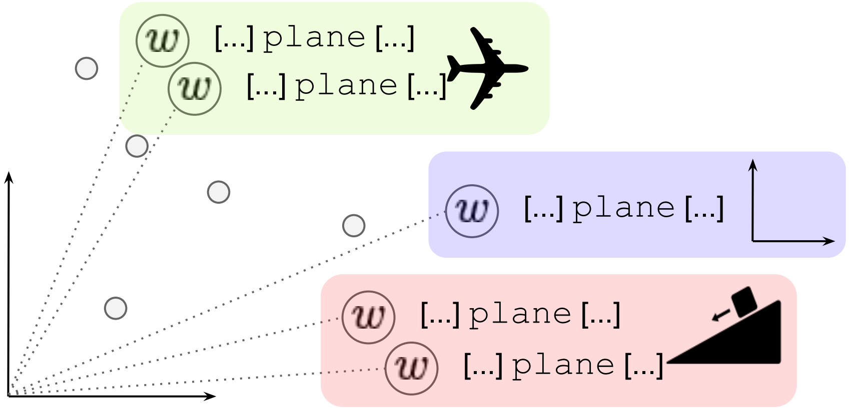
Static+count vs. contextualized models



trained / pre-trained

Word2Vec
FastText
GloVe

word-level approaches



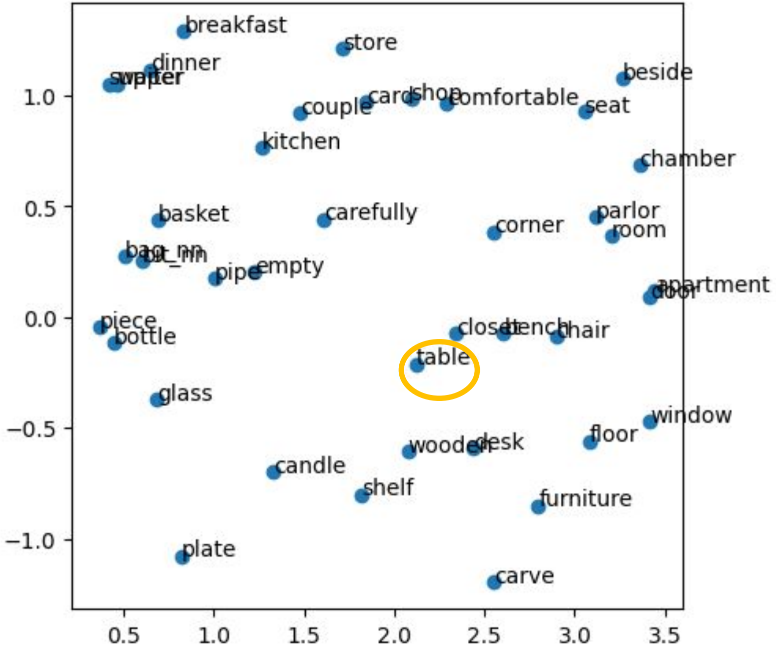
pre-trained / fine-tuned

BERT - mBERT
RoBERTa - XLM-R
XL-LEXEME

word usage-level approaches

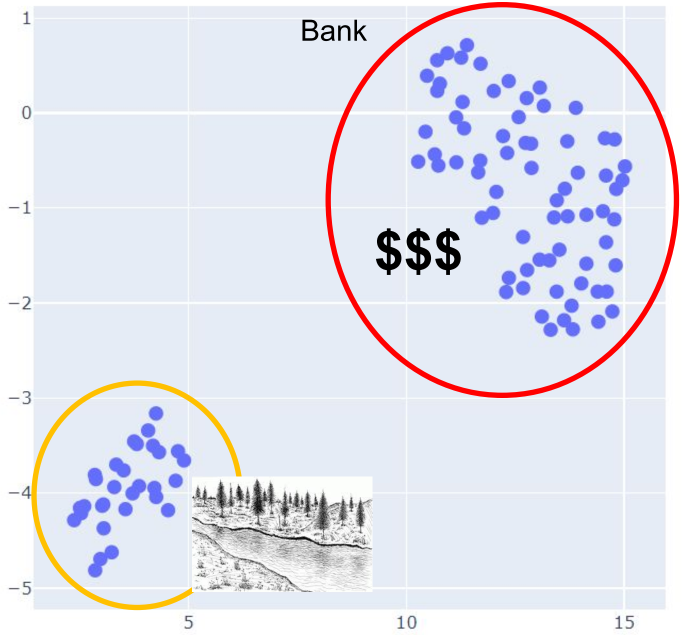
Static+count vs. contextualized models

PPMI



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

Contextualized models



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

Take homes

- Semantic change has wider implications, linguistic research is <u>only one</u> of them

- Many models of (computational) meaning exist
  - **All** models (that were described) are based on the DH
    - DH has its limitations
    - There are other types of models
  - The most recent (best?) model is not necessary the most suitable one
    - This largely depends on the research hypothesis and domain of work
    - "Old" models have their advantageous

See you tomorrow



Introduction

Evaluation

Models of change

Research applications

LSC in your research

Some familiarity with NLP

Worked with text before

Interest in semantic change

Have your ideas for research applications

We are here for **YOU**: approachable anytime in ESSLLI

Homework!

- Please register to DuREL for tomorrow's class at: https://durel.ims.uni-stuttgart.de/register