

Semantic change

Kristian Ristic

Research topics in natural language processing
TU Wien

What is (lexical) semantic change?

- Lexical semantics (or lexicosemantics) is the study of word meanings.
- Semantic change: how the meaning(s) of a word change(s)

Semantic change is also known as semantic shift, semantic progression, semantic development, or semantic drift

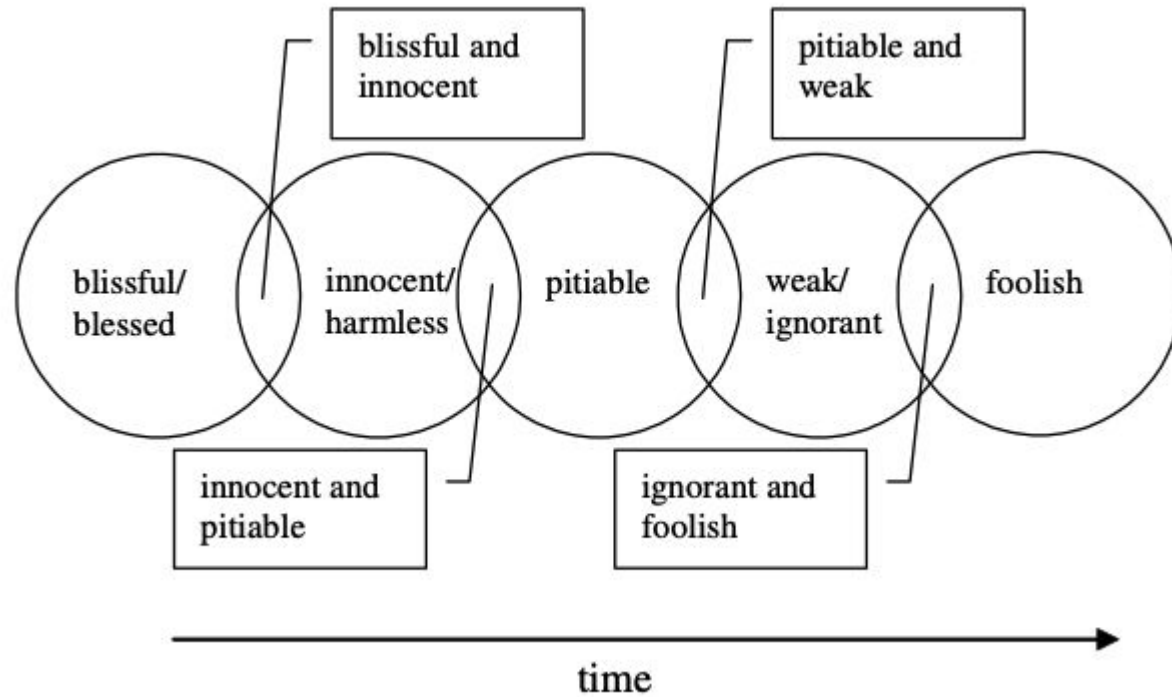
“Traditional” classification of change:

1. More positive (*melioration*) or negative (*pejoration*) meaning
2. Broader (*generalization*) or more narrow (*specialization*) meaning
3. As result of *metaphor* or *metonymy*
 - a. Example for metaphor: “**foot** of a mountain”
 - b. Examples for metonymy:
 - i. “The **White House** decided that ... “
 - ii. “**Liverpool** should hire Jose Mourinho”

“Traditional” classification of change:

1. More positive (*melioration*) or negative (*pejoration*) meaning
2. Broader (*generalization*) or more narrow (*specialization*) meaning
3. As result of *metaphor* or *metonymy*
 - a. Example for metaphor: “**foot** of a mountain”
 - b. Examples for metonymy:
 - i. “The **White House** decided that ... “
 - ii. “**Liverpool** should hire Jose Mourinho”

However, they cannot be applied to all changes and are not mutually exclusive!



Google Books Ngram Viewer

Q gay

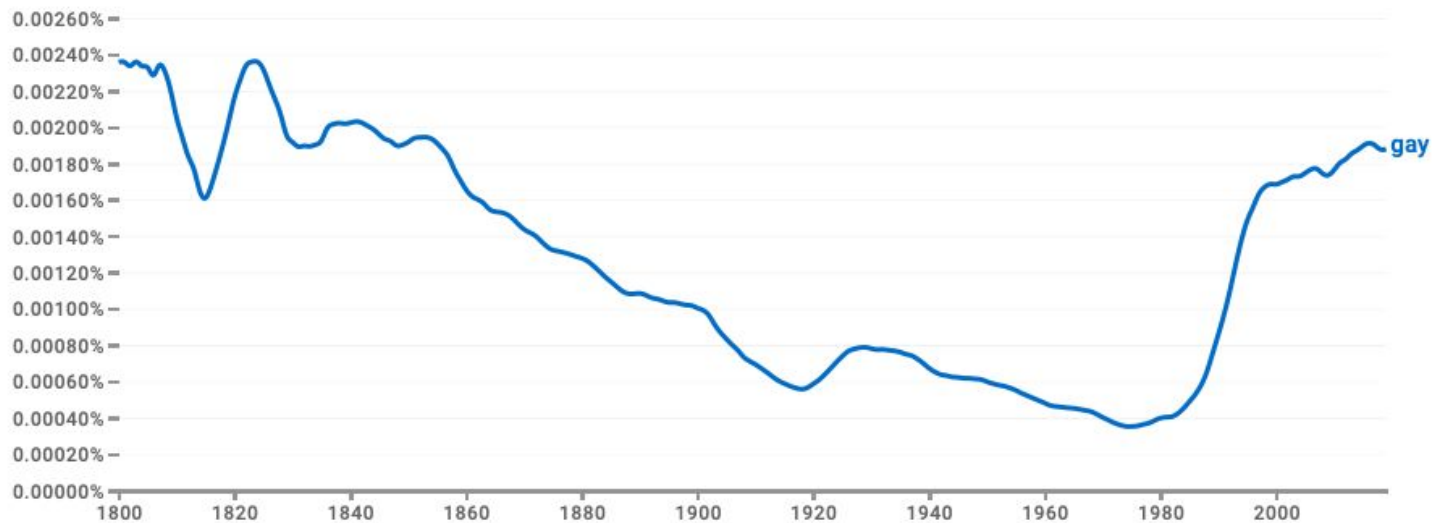


1800 - 2019

English (2019)

Case-Insensitive

Smoothing of 2



(click on line/label for focus)

Motivations for semantic change:

Motivations = reasons why change occurs. Mechanisms = how change occurs

Motivations are often not directly discoverable from the historical record

Some ideas:

- Law of differentiation (1897, 1917): near-synonyms tend to diverge in meaning
 - Inefficient to have multiple forms with similar meaning
- Pragmatic inferencing
 - Perception of speech and negotiation of meaning among interlocutors
 - Distinction between pragmatics (meaning beyond what is said) and semantics (coded meaning)

List of motivations for semantic change by Blank (1997)

- Linguistic forces
- Psychological forces
- Sociocultural forces
- Cultural/encyclopedic forces

How to detect/analyze semantic changes with computational models?

No out-of-the box method :(

Adapt existing methods such

- Word embeddings from distributional semantic models
- Contextual word embeddings (usually pre-trained)
- Semantic role labeling
- Sense clustering

Embeddings algorithms

- PPMI (positive point-wise mutual information)
- SVD on ppmi matrix
- SGNS (Skip-gram with negative sampling)

Positive point-wise mutual information (PPMI)

As *association metric* between the word w and the (context) word c

$$\text{PMI}(w, c) = \log \frac{\hat{p}(w, c)}{\hat{p}(w)\hat{p}(c)} = \log \frac{\#(w, c) \cdot |D|}{\#(w) \cdot \#(c)}$$

$$\text{PPMI}(w, c) = \max \left(\log \frac{\hat{p}(w, c)}{\hat{p}(w)\hat{p}(c)}, 0 \right)$$

$$M_{i,j} = \max \left(\log \frac{\hat{p}(w_i, c_j)}{\hat{p}(w_i)\hat{p}(c_j)} - \alpha, 0 \right)$$

SVD on ppmi matrix

From sparse to dense by reducing dimensionality

1. Take PPMI matrix (M) as input
2. Do eigen-value decomposition with $k < d$. (Usually $100 \leq k \leq 300$)

$$M_k = U_k \cdot \Sigma_k \cdot V_k$$

The dot-products between the rows of $W = U_k \Sigma_k$ are equal to the dot-products between rows of M_k .

$$W^{SVD} = U_k \cdot \Sigma_k \quad C^{SVD} = V_k$$

Traditional Skip-gram

$$P(w_{t+j} = o \mid w_t = c) = P(o \mid c) = \frac{\exp(\mathbf{u}_o^\top \mathbf{v}_c)}{\sum_{k=1} \exp(\mathbf{u}_{w_k}^\top \mathbf{v}_c)}$$

Skip-gram with negative sampling (SGNS)

(it becomes a binary logistic regression)

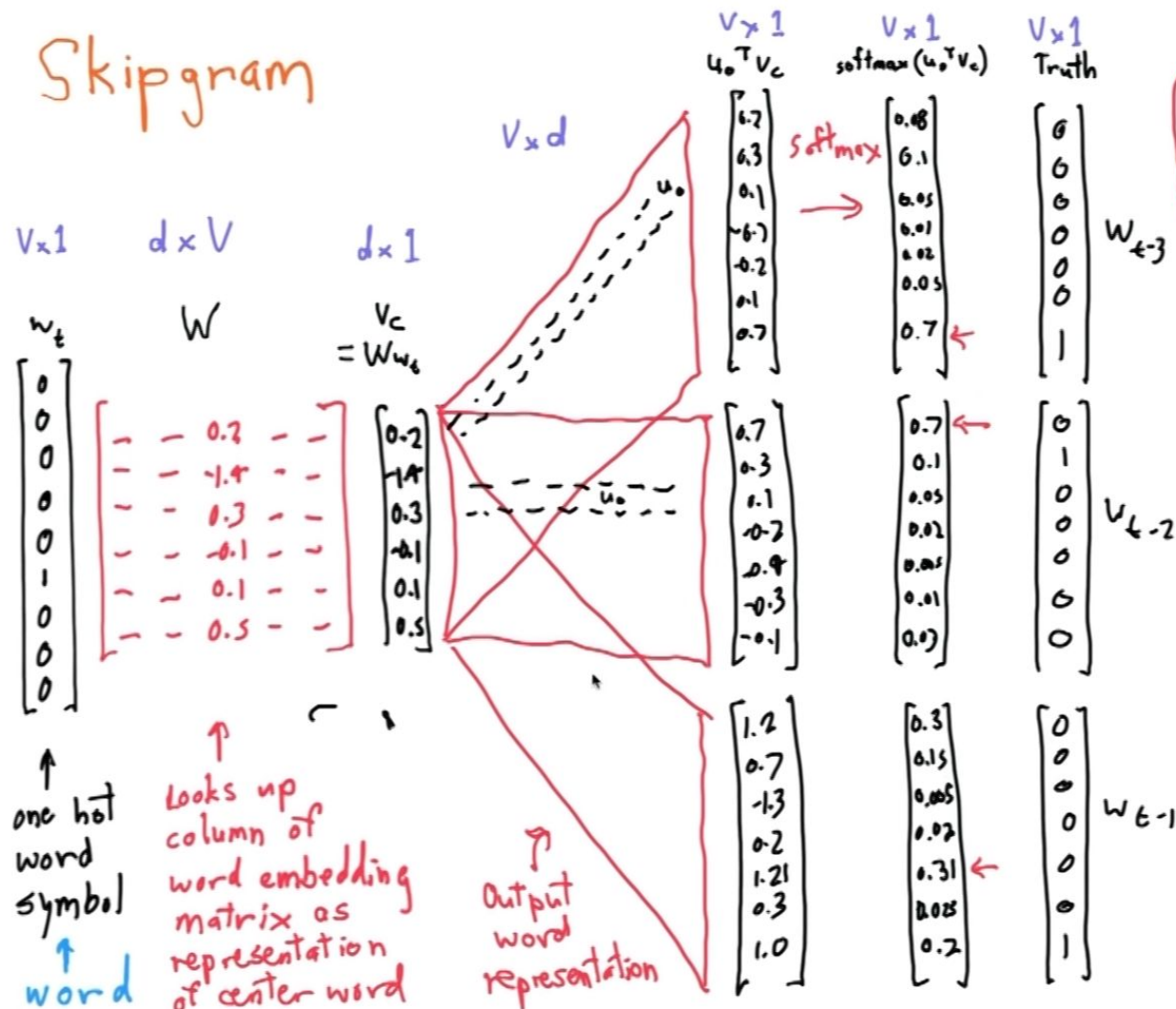
$$P(D = 1 \mid o, c) = \sigma(\mathbf{u}_o^\top \mathbf{v}_c) \quad P(D = 0 \mid w_k, c) = \sigma(-\mathbf{u}_{w_k}^\top \mathbf{v}_c)$$

$$L(o, c) = - \left(\log \sigma(\mathbf{u}_o^\top \mathbf{v}_c) + \sum_{k=1}^K \log \sigma(-\mathbf{u}_{w_k}^\top \mathbf{v}_c) \right)$$

Skip-gram as a neural network

Idea: learn embeddings as task of word prediction

Skipgram



Softmax

$$p_i = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

Connection between PMI and Skip-gram model

If we multiply WW'^T , we get a $|V| \times |V|$ matrix M ,
each entry m_{ij} corresponding to some association between input word i
and output word j

$$WW'^T = PMI - \log k$$

... back to semantic change

instead of only two big matrices (W and W') $\rightarrow$ we have matrices
(=word embeddings) for every time-slice we want to analyze

Alignment problem

- Many loss functions used for training low-dimensional embeddings (SVD) and SGNS are invariant to rotations.
- Magnitude and direction of the vectors are inconsistent across models.
Embeddings from different models are not comparable :(

-> Vectors must be aligned to the same coordinate axes!

Orthogonal Procrustes

Given two matrices A and B , find orthogonal matrix R that:

$$\min_{\mathbf{R}} \|\mathbf{A}\mathbf{R} - \mathbf{B}\|_F$$

1. Do eigen-value decomposition on the covariance matrix M between A and B

$$M = B^T A$$

$$M = U \Sigma V^T$$

2. Compute the optimal orthogonal matrix R

$$R = V U^T$$

3. Apply transformation

$$A_{\text{alig}} = AR$$

Joint optimization problem

$$\min_{U(1), \dots, U(T)} \left[\frac{1}{T} \sum_{t=1}^T \|Y(t) - U(t)U(t)^T\|_F^2 + \frac{\lambda}{2} \sum_{t=1}^T \|U(t)\|_F^2 + \frac{\tau}{2} \sum_{t=2}^T \|U(t) - U(t-1)\|_F^2 \right]$$

Embeddings $U(t)$ to reconstruct PPMI matrices $Y(t)$

Penalty term: enforces embedding to be of low-rank

Smoothing term: embeddings should change gradually, If $\tau = 0$ no change (needs to be >0)

This approach enforces alignment across all time slice!

Word embeddings are aligned... what now?

1. Pair-wise similarity time-series
2. Measuring semantic displacement
3. Evaluate synchronic accuracy: how to measure word similarity within individual time-periods
4. Evaluate diachronic validity: how to quantify semantic changes over time
 - a. Detecting known shifts
 - b. Discovering shifts from data
5. Evaluate statistical laws of semantic change?

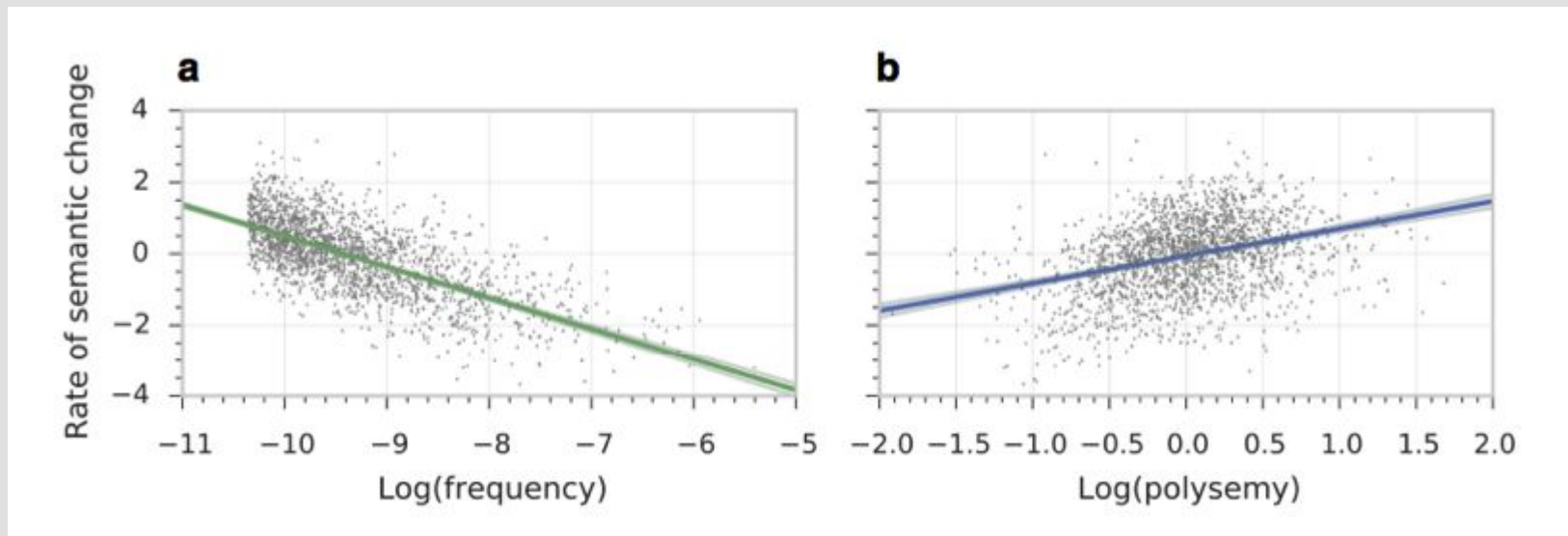
Detecting shifts from data

- Set of attested historical shifts (# 10) used to evaluate the methods

Method	Corpus	% Correct	%Sig.
PPMI	ENGALL	77.1	51.9
	COHA	85.7	52.4
SVD	ENGALL	92.6	81.5
	COHA	95.8	62.5
SGNS	ENGALL	100.0	88.9
	COHA	87.5	50.0

Statistical laws of semantic change

- How a word's rate of semantic change depends on its frequency and polysemy?



Law of conformity

Frequently used words change at slower rates

The logarithm of a word's frequency has a significant and substantial negative effect on rates of semantic change.

Law of innovation

Polysemous words change at faster rates

- Measuring word polysemy is tricky.
 - Trick used: measuring a word's contextual diversity as a proxy for its polysemousness
1. Construct empirical co-occurrence networks for the top-10,000 non-stop words of each language using the PPMI measure
 2. Define a “clustering coefficient” (low clust. coeff -> high polysemy score)
- > Logarithm of polysemy score exhibits a strong positive effect on rates of semantic change

What about contextualised word representations?

First time in 2020: “Analysing Lexical Semantic Change with Contextualised Word Representations”

- For word of interest, and a context of occurrence, extract activations of all of BERT’s hidden layers.
- Build “usage representations” (matrix U) by summing the activations dimension-wise. Matrix contains a 128-token window around the target word as well as a temporal label
- For each word, do k-means
- Quantify semantic change based on some similarity/distance metrics

Limitations

Static embeddings

- one embedding per word
- based just on co-occurrence statistics

Contextualized embeddings:

- pre-training bias
- token embedding captures changes but not proper semantic shifts (acquiring or losing a lexicographic sense). Problematic cases identified:
 1. Words of very context-dependent ‘fluid’ meaning, used in all sorts of contexts.
 2. Words frequently used in a very specific context in a particular time bin (‘data bursts’)
 3. Words used as a proper name in a particular time bin only.
 3. Words being used in significantly different syntactic roles.

Challenges

- Need for larger and more robust gold standard test sets of semantic shifts
- Identify the motivations and/or source of such a change
- Identify groups of words that shift together
- Build robust methods for small datasets
- From diachronic to synchronic

Conclusion

- Static embeddings are still very solid (and efficient).
- Contextualized architectures are promising, but their usage is not straight-forward. Numerous fine-tuning tasks are required.
- Contextual variance != semantic change ??
- Semantic drift = historical semantic change ??

Application for semantic change analysis

<https://sem.kristiam.xyz>