# Drifting Distributions?

## Possibilities and Risk of Using Distributional Semantics for studying Concept Drift

Drift-a-LOD, Amsterdam 11 September 2017

Antske Fokkens

# Drifting Distributions

- What is concept drift?
- What is distributional semantics?
- How can distributional semantics be used?
- How can it not be used?

# Concept Drift

- Aspects of a concept (Wang et al. 2011):
  - the *intension*
  - the *extension*
  - the *label*

# Intension

- Frege's *sense*: the sense provides a function that takes you to the extension of the concept and a perspective on the denoted concept

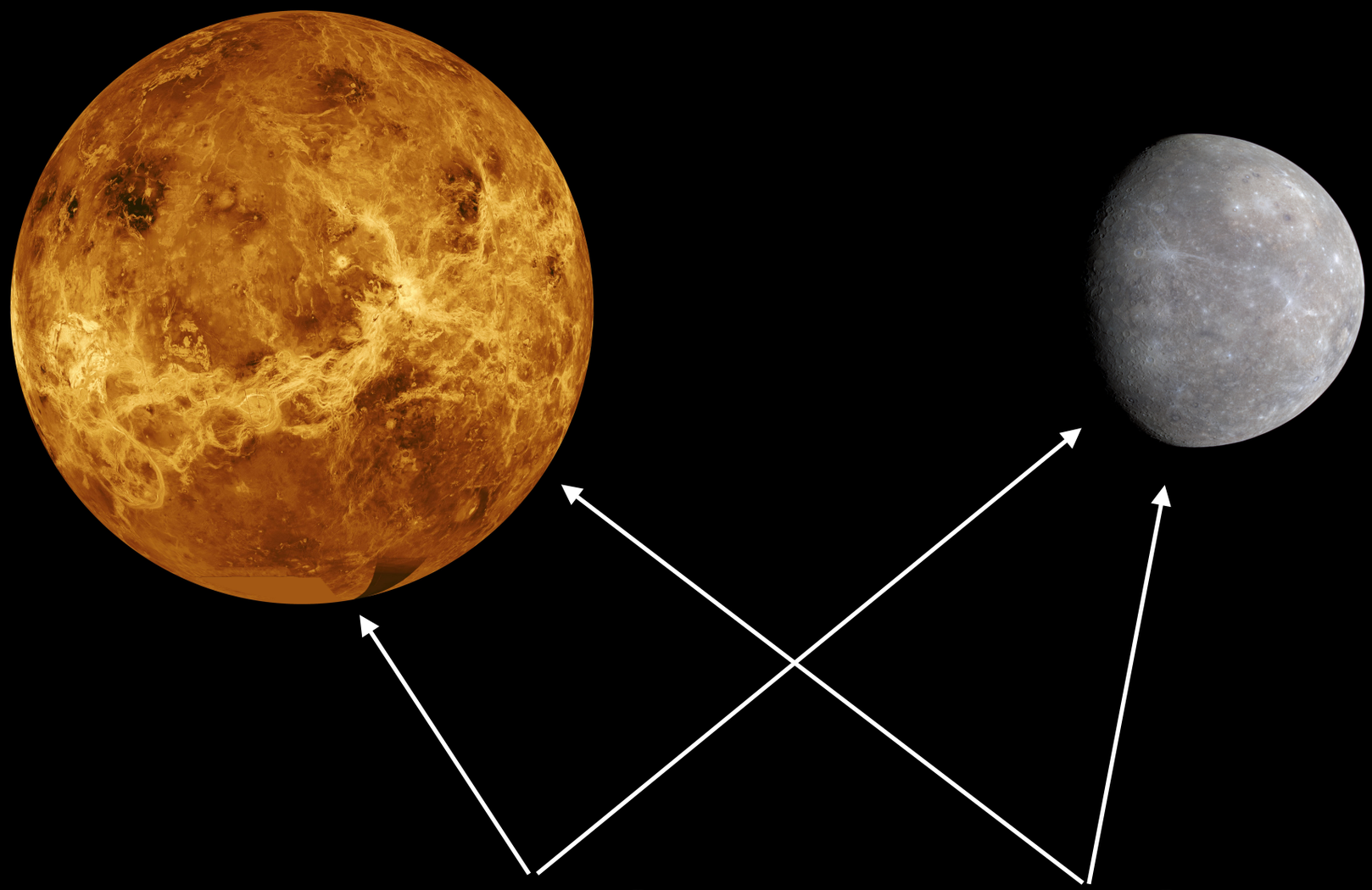
# Extension

- The set of things that are defined by the intension in the world (what is being denoted)

# Labels

- The words that are used to refer to something

# Meaning



# Concept Drift

- Concept drift for scholars (following Fokkens et al. 2016):
  - *typically*, changes in intension (perspective), where the core meaning stays the same (Kuukanen, 2008)
  - changes in extension can be relevant for specific concepts (e.g. EU) or in extreme cases

# Distributional Semantics

The meaning of a word is determined by its usage (Wittgenstein)

=> words used in a similar context will have similar meaning (Harris 1954; Firth 1957)

# Distributional Semantics

- A bottle of tesguïno is on the table.
- Everybody likes tesguïno.
- Tesguïno makes you drunk.
- We make tesguïno out of corn.

(Jurafsky and Martin, 2015)

# *Vector Semantics*

- Distributional semantics approaches generally represent words as *vectors*

# Co-occurrence vectors

| | aardvark | ... | computer | data | pinch | result | sugar | ... |
|---|---|---|---|---|---|---|---|---|
| **apricot** | 0 | ... | 0 | 0 | 1 | 0 | 1 | |
| **pineapple** | 0 | ... | 0 | 0 | 1 | 0 | 1 | |
| **digital** | 0 | ... | 2 | 1 | 0 | 1 | 0 | |
| **information** | 0 | ... | 1 | 6 | 0 | 4 | 0 | |

Jurafsky and Matrin (2015)

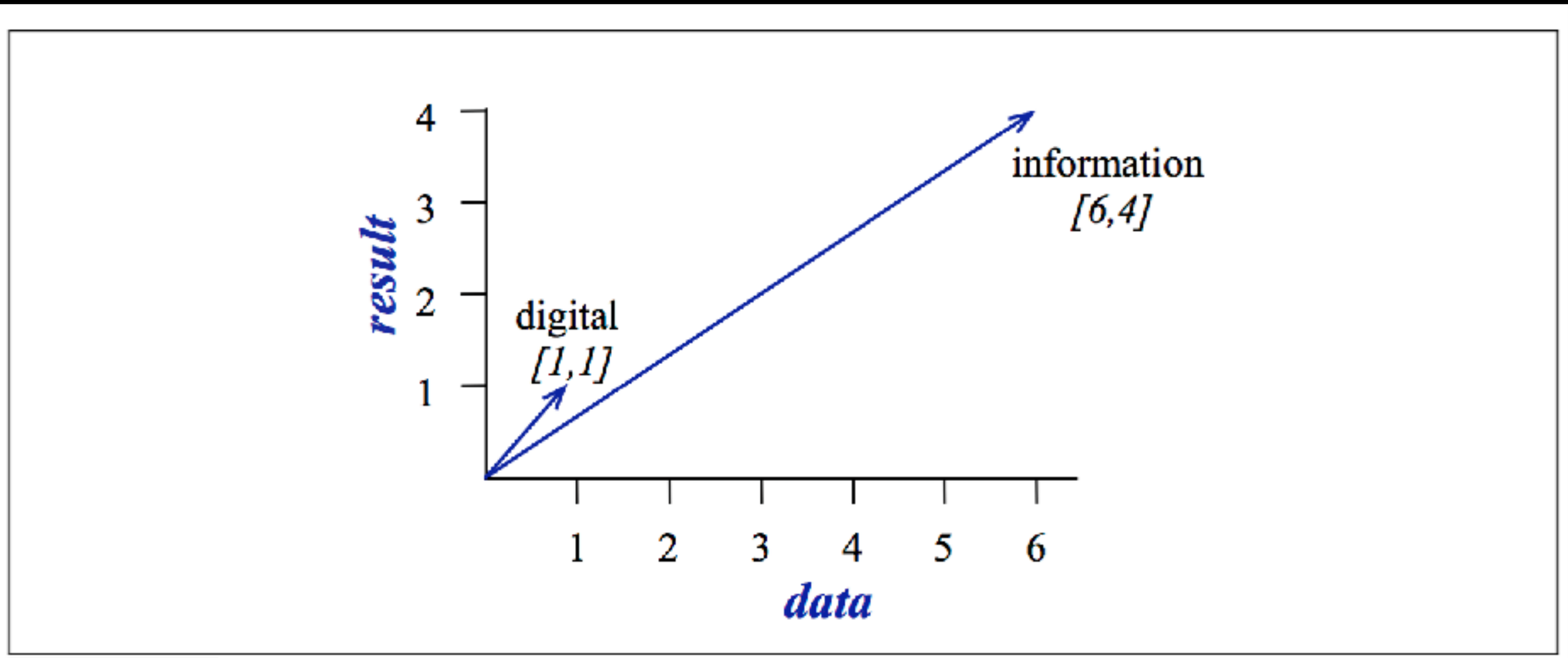
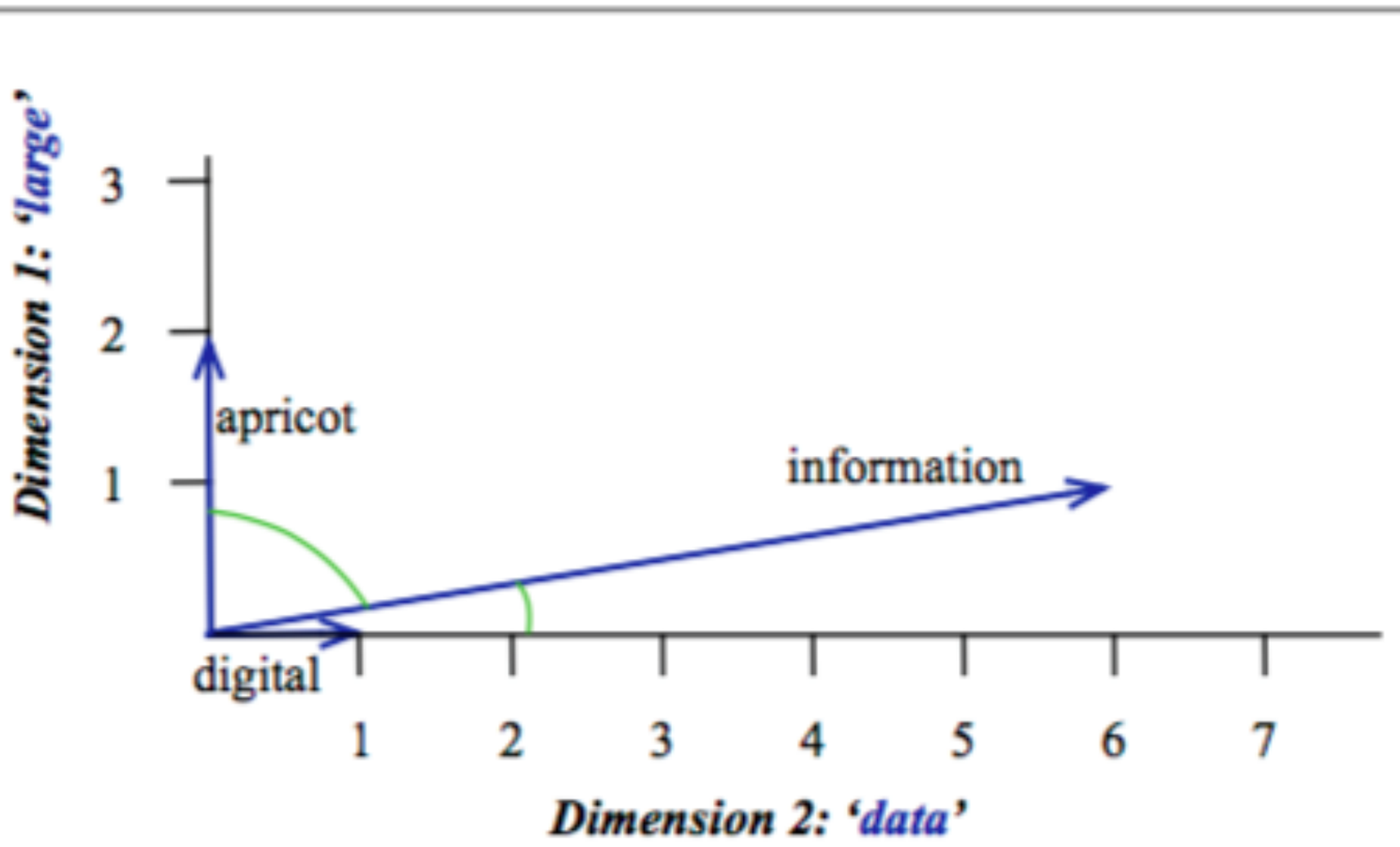



**Figure 19.5** A spatial visualization of word vectors for *digital* and *information*, showing just two of the dimensions, corresponding to the words *data* and *result*.

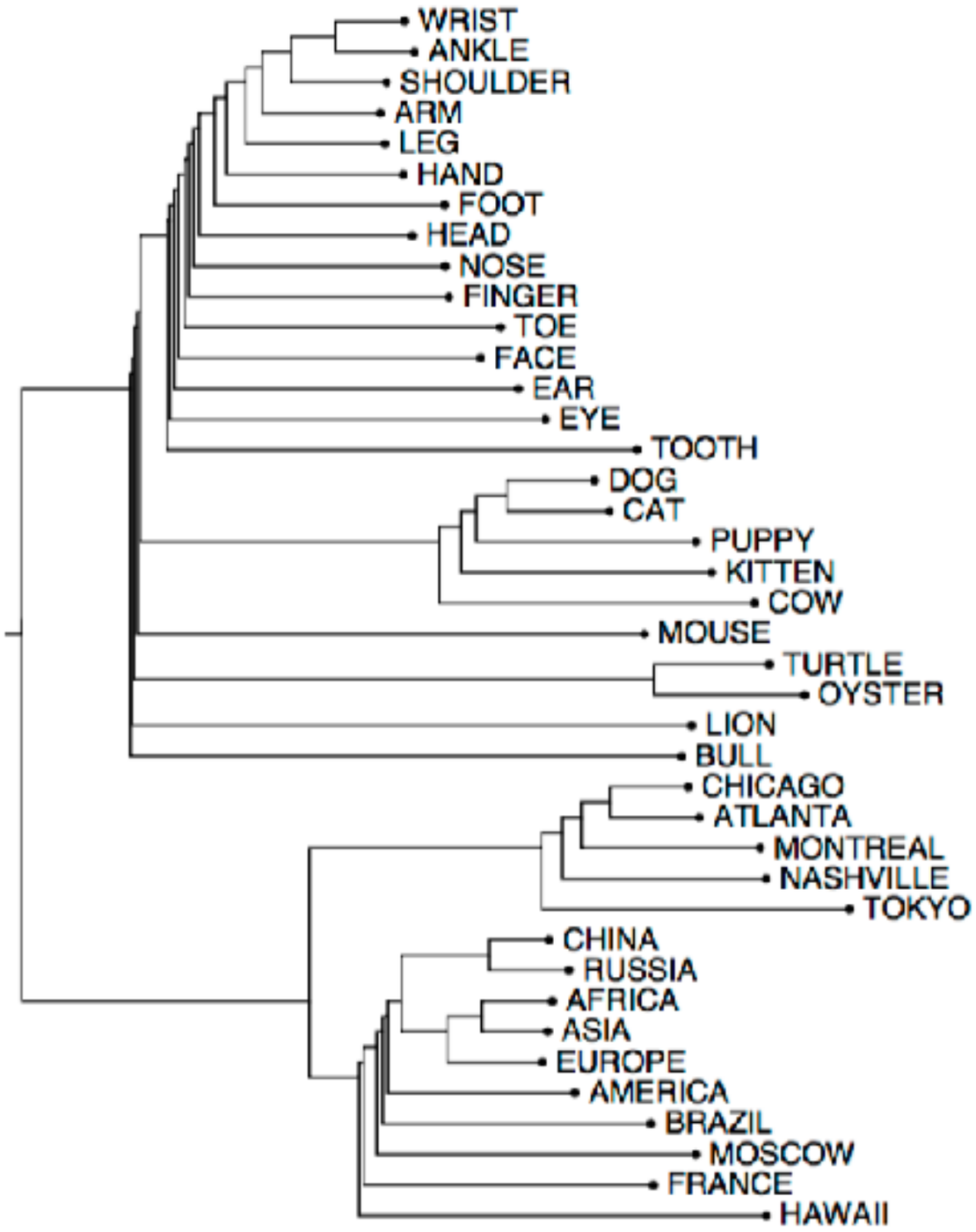
Jurafsky and Matrin (2015)



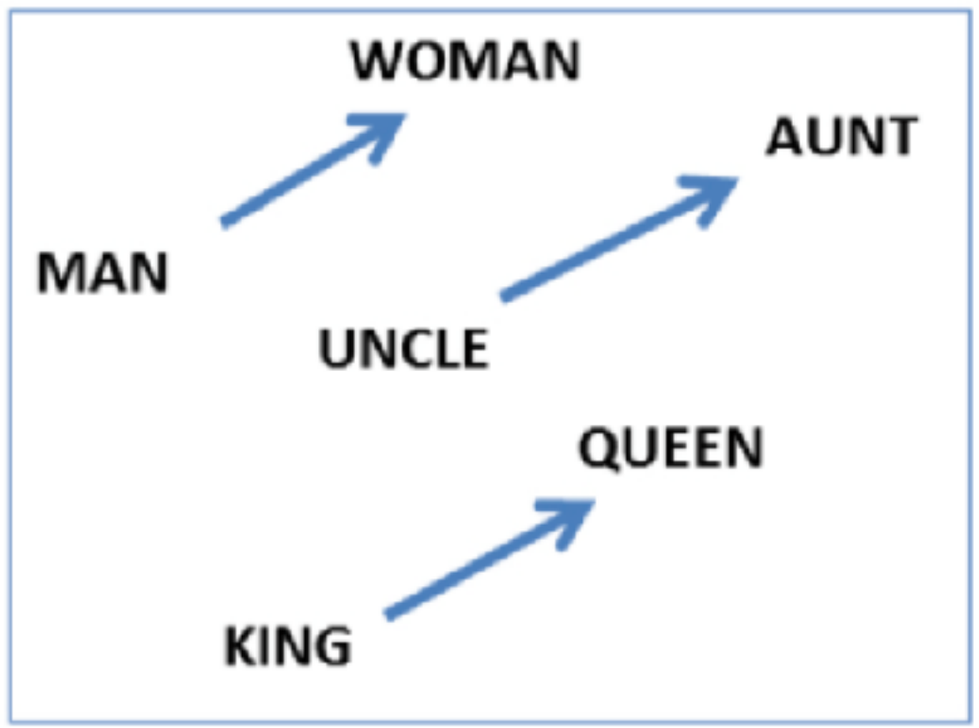

Jurafsky and Matrin (2015)

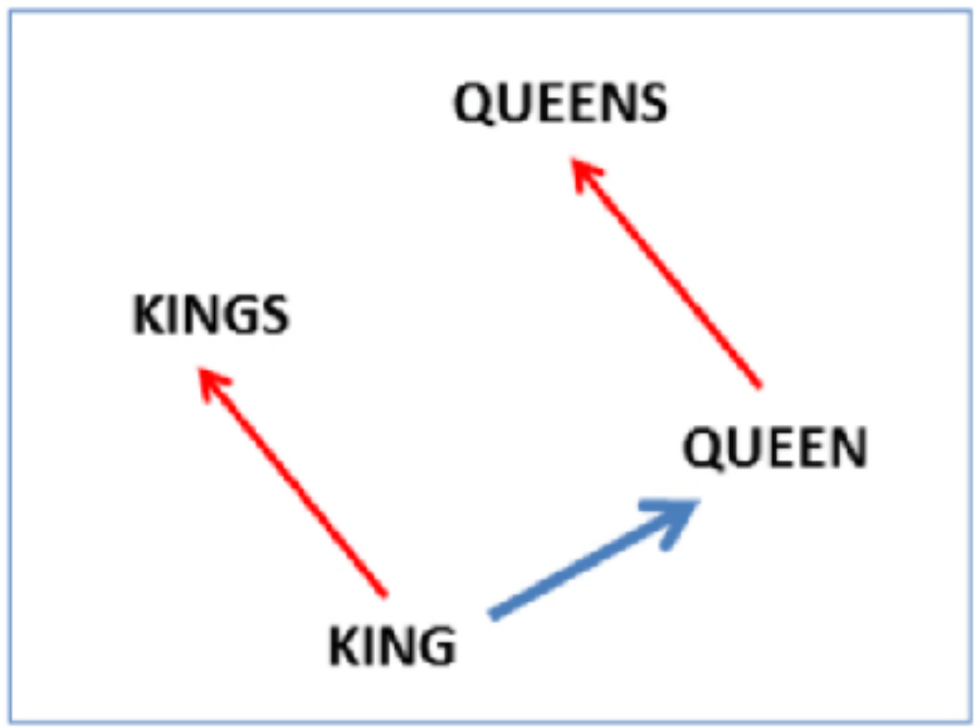
# Beyond counting co-occurrence

- Relatedness vs similarity:
  - text/paragraph or sentence as context => words related to the same topic
  - small window of close terms as context => less general relatedness, more similarity
- Relevance:
  - frequent co-occurrence with *the* or *wicket*




Jurafsky and Matrin (2015)




Jurafsky and Matrin (2015)

# Distributional Semantics for concept drift

- Distributional semantics can provide insight into the relation between *label* and *intension*
- Used for detecting change in meaning (*sense shift*)
- Can this also be used for detecting concept drift?

# 2 open questions

- How to go beyond *sense shift*?
- What is the reliability of the method?

# Concept systems

- Betti and Van den Berg (2014): concepts should not be examined in isolation
- Geeraerts (p.c.): change in the concept itself is best examined by investigating related concepts

# Reliability

- How reliable or indicative are measures of change?
- How reliable are distributional models?

# Word Embeddings

- Used for detecting diachronic change. Common approach: changes in n-nearest neighbors

# word2vec

- Hellrich & Hahn (2016): n-nearest neighbors change when running word2vec on the same corpus

# Count vs Predict

- Baroni et al (2014): Predictive models work better than count models
- Levy et al (2015): if you use same hyperparameters: count is better for similarity, predict for analogy

*How does this hold up to the (in)stability of word2vec?*

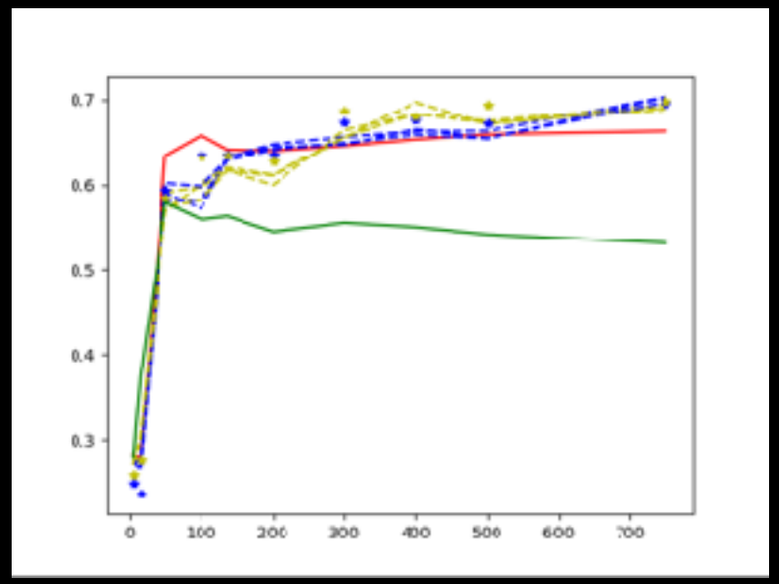
# Models tested

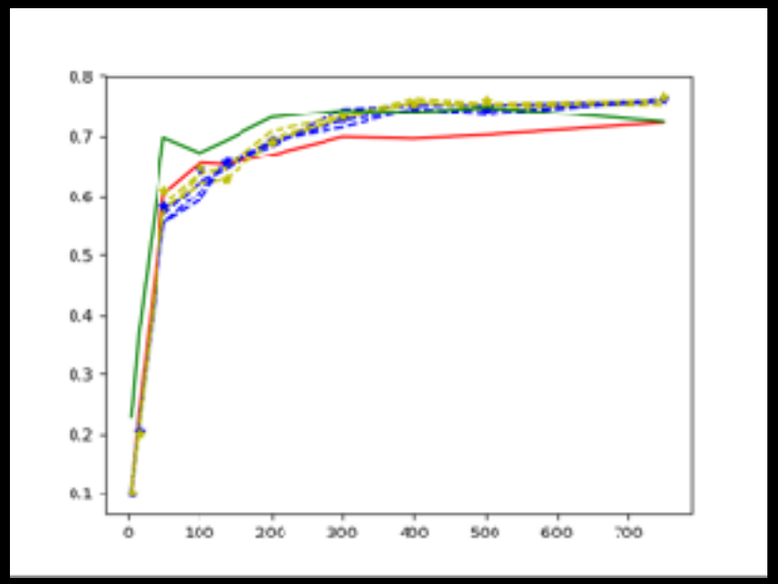
- Optimal settings from Levy et al. for models:
  - PPMI
  - SVD
- word2vec:
  - 3 random initiations
  - svd initiation
  - look at examples in different order (beginning to end & end to beginning)

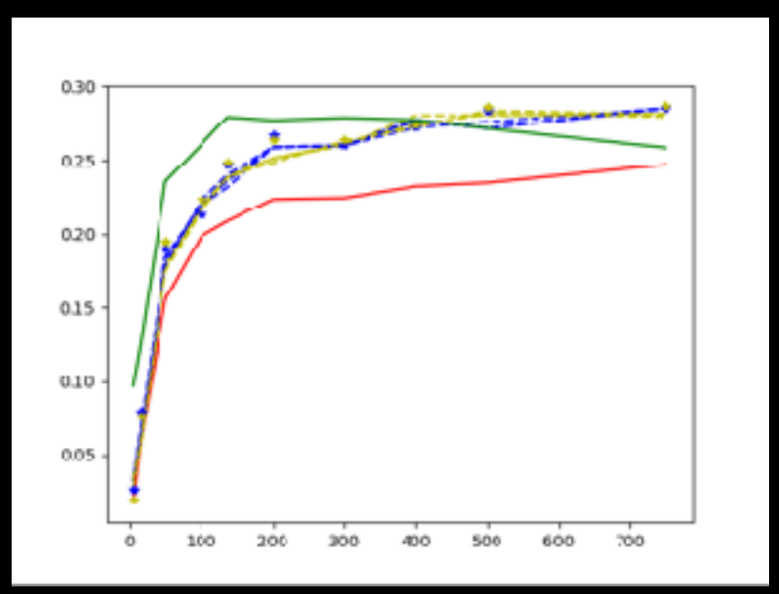
# Experimental Setup

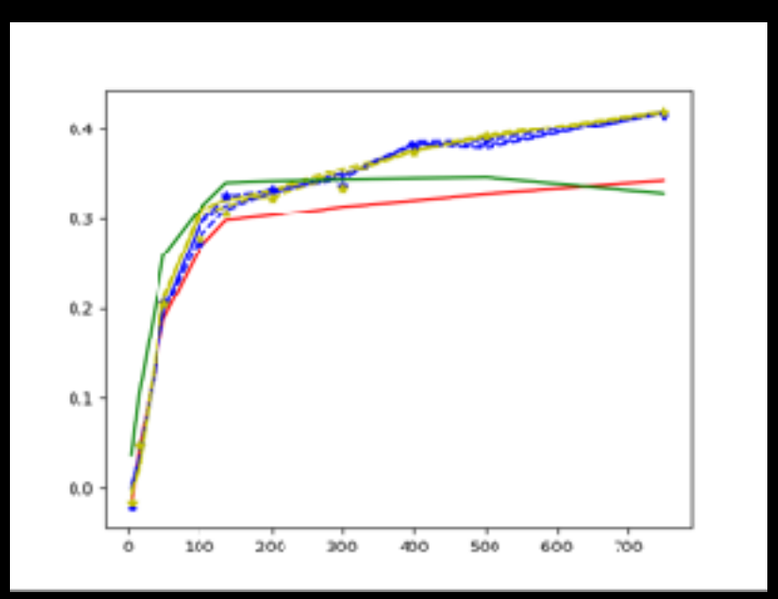
- Wikipedia dump Jan 2017:
  - 1.8 Billion words, randomized
  - Subset (by taking head and tail) of 0.12M, 5M, 15M, 50M, 100M, 200M, 300M, 400M, 500M, 750M words
- Standard evaluation sets:
  - Similarity/relatedness pairs
  - Analogy evaluation

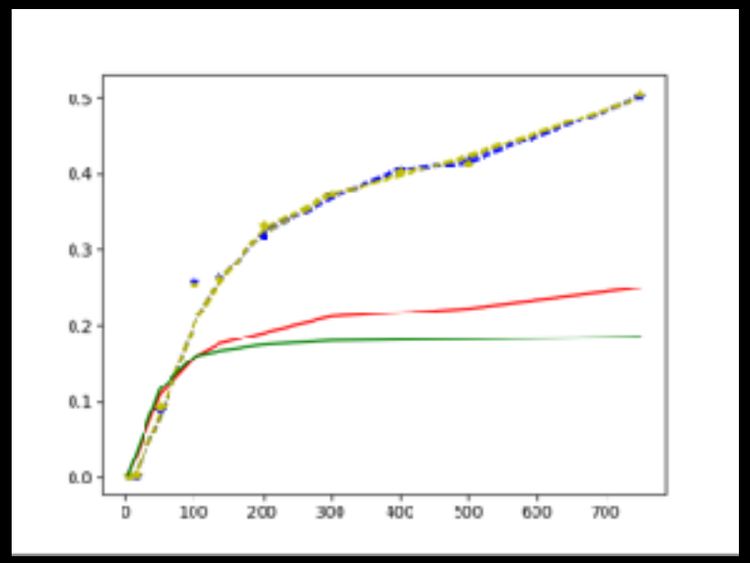
# Similarity evaluation

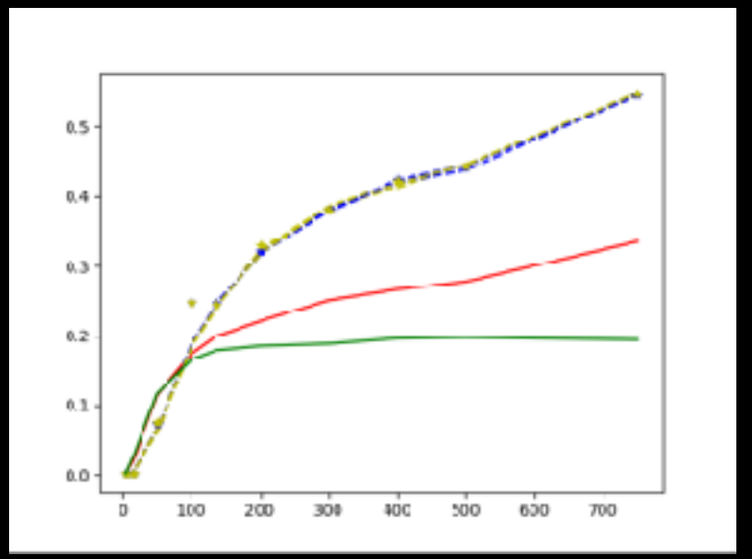


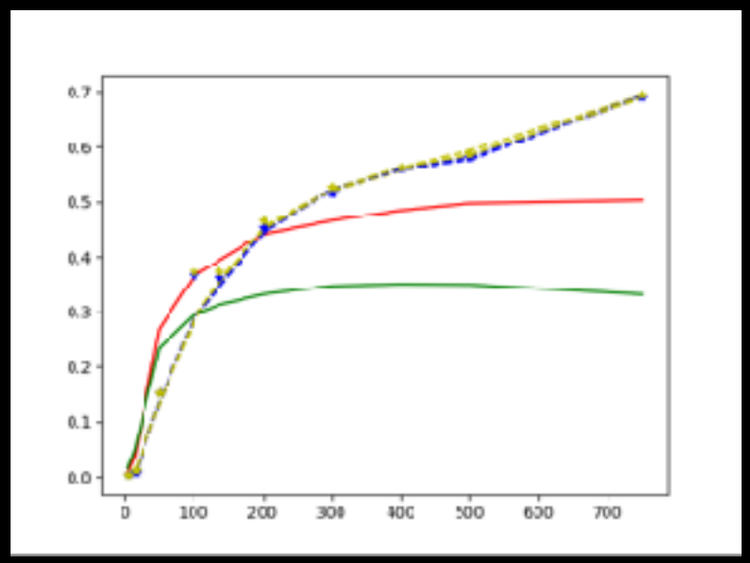



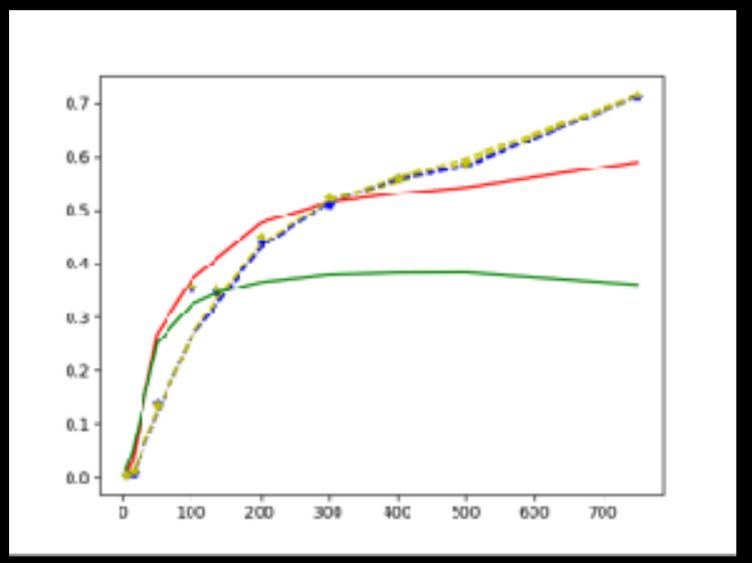




0.5
0.4
0.3
0.2
0.1
0.0
0
100
200
300
400
500
600
700


0.5
0.4
0.3
0.2
0.1
0.0
0
100
200
300
400
500
600
700


0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
100
200
300
400
500
600
700


0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
100
200
300
400
500
600
700

# Use case:

- Can distributional semantics provide insight into the ways in which *racism* changed in the 20th century? (Sommerauer 2017)

# Theory

- Sociology, social psychology and anthropology:

  Classical open racism had declined towards the end of the 20th century and replaced by a more subtle form of discrimination

  a cultural, ethnical explanation rather than biological

# Method

- Concepts to study:
  - Core concepts: *race, ethnicity, culture* (*racial, ethnic, cultural*)
  - Subconcepts: *language, nationality, religion* (*linguistic, national, religious*)
  - Instances: *blacks, whites, foreigners, immigrants, Jews, Arabs, Turks*

# Related concepts

- How do these relate to:
  - *difference, conflict, superiority*
  - *history/historic, genetics/genetic*
  - *relation, relationship, marriage*
  - *value, belief, attitude*

# Corpora

| Corpus | Composition | Corpus size |
|---|---|---|
| COHA | genre-balanced | 22.5 M - 27.9 M words<br>Average: 24.5 M words |
| Google N-grams ALL | Google books of all genres, not evenly balanced | 11.6 B - 82.5 B words<br>Average: 29 B words |
| Google N-gram Fiction | Google books fiction | 925 M - 11.3 B words<br>Average: 3.0 B words |

# Nearest Neighbor Comparison

- How do the nearest neighbors of a concept change from one decade to another?
- How does the overlap in nearest neighbors change between two related concepts?
  - within a decade?
  - between decade 1 and decade 2

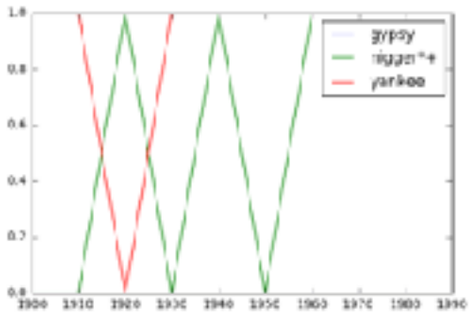
# Changes in relations

- How closely are various concepts related?
- How does this change over time?

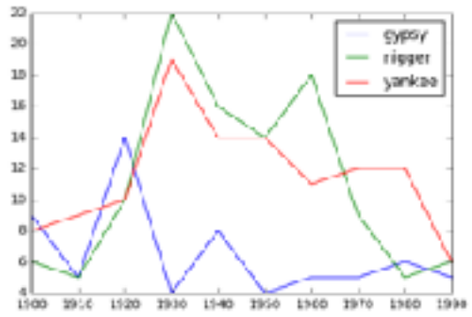
# Reliability check

- Compare different measurements
- Compare effect on target words to effects on control words

# Changes in frequency



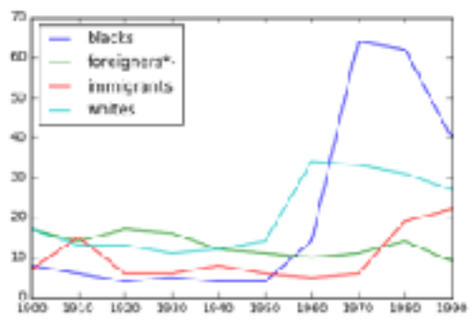

(A) English n-grams.




(B) COHA.

FIGURE 4.4: Changes in the frequencies of ethnic slurs representing instances of *RACE*.

(A) English n-grams.

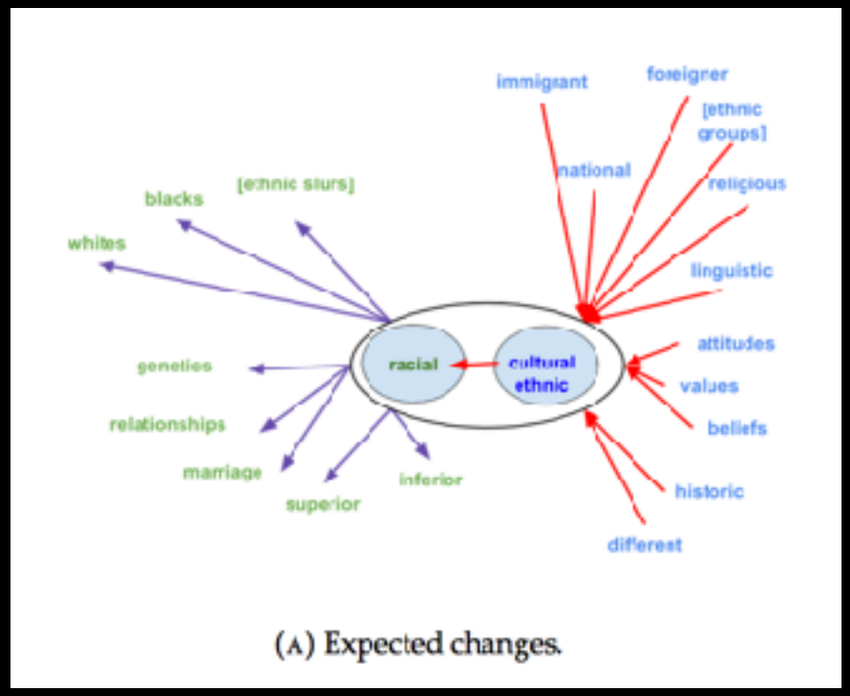



(B) COHA.

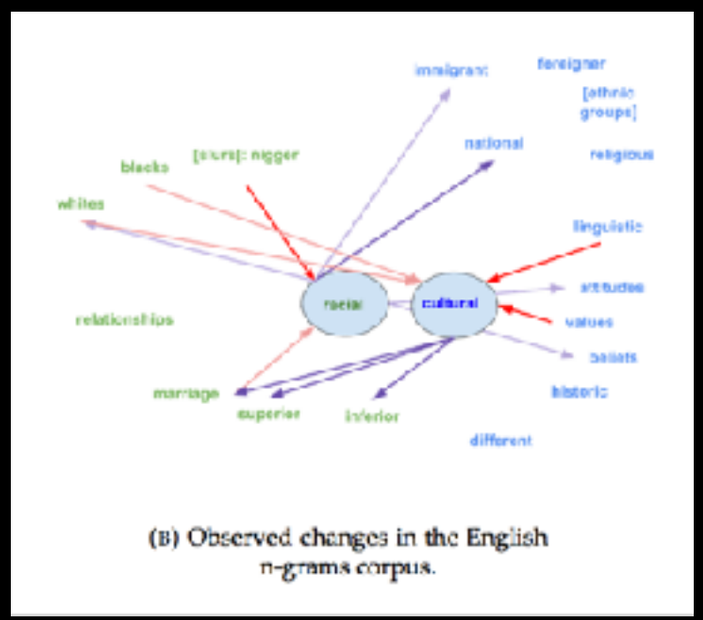
FIGURE 4.5: Changes in the frequencies of added labels representing instances of *RACE*.

- Sommerauer (2017), p. 66

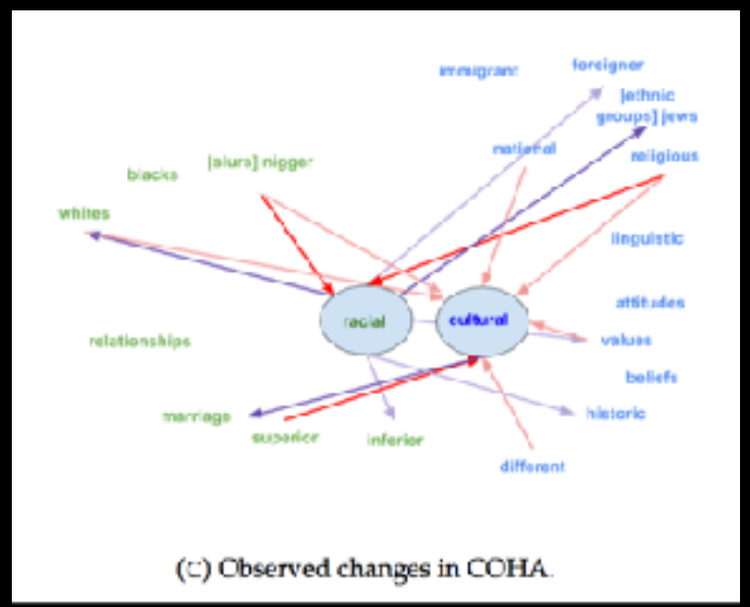
# Changing relations




(A) Expected changes.




(B) Observed changes in the English n-grams corpus.




(C) Observed changes in COHA.

# N-gram results

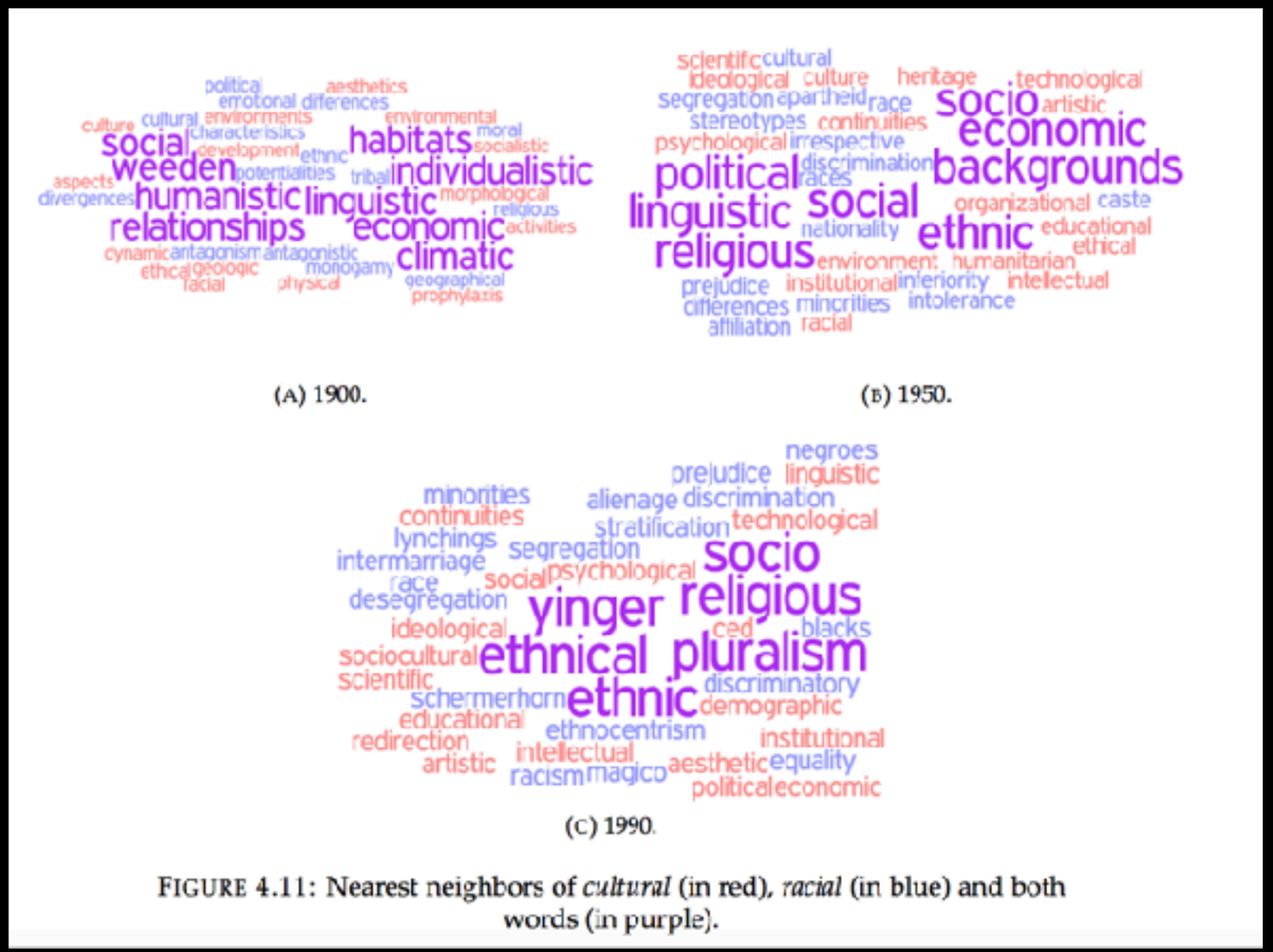


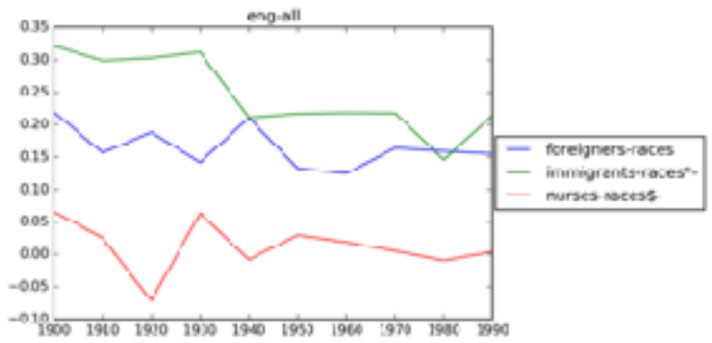

FIGURE 4.11: Nearest neighbors of *cultural* (in red), *racial* (in blue) and both words (in purple).

- Sommerauer (2017), p. 71

# Conclusions

- Relations between concepts partially changed as expected
- Nearest neighbor show clear confirmation of *racial* shifting to a concept mainly associated with discrimination
- *ethnic* and *religious* move from only `racial' to `racial&cultural'
- *political* moved from `racial'

# Instances & control word



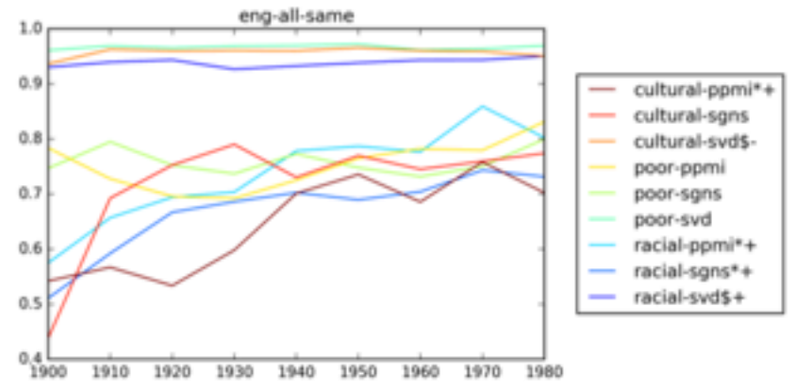

(A) Social group labels in relation to *races*.

(B) Social group labels in relation to *cultures*.

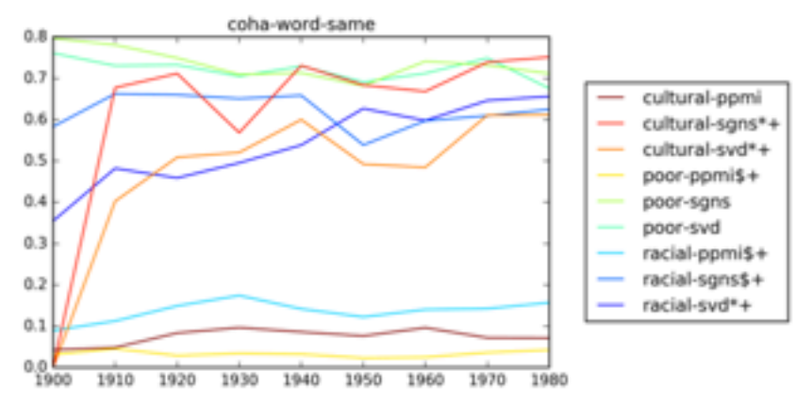
FIGURE 4.16: Similar patterns in the relations beteen social group labels and *RACE* in COHA.

- Sommerauer (2017), p. 83

# Relational variations based on alternative models




(A) English n-grams.




(B) COHA.

FIGURE 4.17: Variations in the changes in the words representing the core concepts of *RACE* measured by direct vector comparison in all models

# Nearest Neighbors depending on various models

(A) English n-grams

FIGURE 4.20: Shared nearest neighbors of *racial* among the three different models in the English n-grams corpus in 1900.

# Conclusions (revised)

- Control words show that changes in relations between core concepts did not yield reliable insights in this study
- Fluctuation in results depending on the model that was used shows that care must be taken when interpreting relational results
- One insights seem to hold across variation:
  - racism moved from something seen as similar to `ethnic' and `cultural' to something mainly associated with discrimination

# Using distributional semantics for concept drift?

- Possible, but:
  - translating the concept under investigation to measurable tokens is non-trivial
  - careful for artifacts of (small) data
  - use a solid methodological setup:
    - control terms
    - alternative measures for testing hypotheses
    - alternative models for representing data

# References


- Baroni, Marco, Georgina Fina and German Kruzewski (2014) Don't count, predict! A systematic comparison of context-counting vs. concept-predicting semantics vectors. In: *Proceedings of ACL,* Baltimore, USA.
- Betti, Arianna and Hein van den Berg (2014) Modelling the history of ideas. In: *British Journal for the History of Philosophy, 22(4),* P. 812-835.
- Firth, John R. (1957) A synopsis of linguistic theory 1930-1955. *Studies in linguistic analysis.*
- Fokkens, Antske, Serge ter Braake, Isa Maks and Davide Colin (2016). On the Semantics of Concept Drift: Towards Formal Definitions of Semantic Change. In: *Proceedings of Drift-a-LOD.* Bologna, Italy.
- Harris, Zellig S (1954). Distributional Structure. *Word* 10(2-3). p 146-162.
- Hellrich, Johannes and Udo Hahn (2016) Bad Company-Neighborhoods in Neural Embedding Spaces Considered Harmful. In: *Proceedings of COLING,* Osaka, Japan.
- Jurafsky, Daniel and James Martin (2015) Speech and language processing. *Preprint of third edition.* https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf
- Kuukkanen, Jouni (2008) Making sense of conceptual change. In: *History and theory* 47.3 p. 351-372.
- Levy, Omer, Yoav Goldberg and Ido Dagan (2015). Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics,* 3, p.211-225.
- Sommerauer, Pia (2017) *From old to new racism? Investigating known dangers in distributional semantic approaches to conceptual change*. MA thesis. Vrije Universiteit, Amsterdam
- Wang, Shenghui, Stefan Schlobach and Michel Klein (2011): Concept drift and how to identify it. *Web Semantics: Science Services and Agents on the World Wide Web* 9.3 (2011) p 247-265.