Distributional Semantic Models
studying linguistic shift

LOT Winterschool 2019, Day 4

Antske Fokkens

Studying linguistic shift

- Distributional semantic models indicate how words are used in a specific corpus
  - potential **bug** for creating good semantic models: quality of the model highly
  - potential **feature** for studying differences in language use

Language change

- The meaning of words can change over time, moderately or drastically:
    - a specific aspect of meaning changes (more specific, or general) (e.g. *to detail*)
    - a new meaning/use of the word is added next to its original meaning (e.g. *cell*)
    - a new meaning of a word becomes dominant (e.g. *gay*)
    - the overall meaning of a word shifts (e.g. *awful*)

Studying linguistic change

- Investigate how words are used in various corpora
- Distributional semantic technologies provide a way to do so on a large scale:
    - models of the full vocabulary in a given point in time
    - semantic representation of each word can be compared

From Hamilton et al. (2016)a

Corpora

- Google N-grams: 1505 - 2009
    - 850 Billion words
    - Different biases across decades
- Google literature: 1579 - 2009
    - 75 Billion words
    - One genre only
- COHA: 1810 - 2010
    - 420 Million words
    - Carefully balanced across genres

More details

Decade	Google All	Google Fiction	COHA
1900	13.47B	1.11B	22.5M
1910	12.56B	0.93B	22.7M
1920	11.99B	1.14B	25.6M
1930	11.65B	1.08B	24.4M
1940	11.68B	1.04B	24.1M
1950	17.68B	1.64B	24.4M
1960	31.95B	2.27B	23.9M
1970	41.98B	3.13B	23.8M
1980	54.32B	6.37B	25.2M
1990	82.49B	11.3B	27.9M

Based on Sommerauer (2017)

Hamilton et al. 2016a

- Embeddings from Google corpora & COHA
- Created using:
  - PPMI
  - SVD
  - word2vec: SGNS negative sampling
    => next decade initialized by model from previous decade

Hamilton et al. 2016a

- *Law of conformity*

  => inverse power-law between word frequency and change

- *Law of innovation*

  => higher rate of semantic change for polysemous words

Methods

- Pairwise comparison:
  - did the distance between v_1 and v_2 change from time *t* to time *t*+1?
- Individual comparison:
  - when comparing v_1 in the model for time *t* to v_1 in the model for time *t+1*, how much did the vector change?

Individual comparison

- PPMI: keep dimensions stable

- SGNS & SVD: align the models minimizing differences between cosine similarities

Test words

Word	Moving towards	Moving away	Shift start	Source
gay	homosexual, lesbian	happy, showy	ca 1950	(Kulkarni et al., 2014)
fatal	illness, lethal	fate, inevitable	<1800	(Jatowt and Duh, 2014)
awful	disgusting, mess	impressive, majestic	<1800	(Simpson et al., 1989)
nice	pleasant, lovely	refined, dainty	ca 1890	(Wijaya and Yeniterzi, 2011)
broadcast	transmit, radio	scatter, seed	ca 1920	(Jeffers and Lehiste, 1979)
monitor	display, screen	—	ca 1930	(Simpson et al., 1989)
record	tape, album	—	ca 1920	(Kulkarni et al., 2014)
guy	fellow, man	—	ca 1850	(Wijaya and Yeniterzi, 2011)
call	phone, message	—	ca 1890	(Simpson et al., 1989)

From Hamilton et al. (2016)a

Test words & results

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

Word	Moving towards	Moving away	Shift start	Source
gay	homosexual, lesbian	happy, showy	ca 1950	(Kulkarni et al., 2014)
fatal	illness, lethal	fate, inevitable	<1800	(Jatowt and Duh, 2014)
awful	disgusting, mess	impressive, majestic	<1800	(Simpson et al., 1989)
nice	pleasant, lovely	refined, dainty	ca 1890	(Wijaya and Yeniterzi, 2011)
broadcast	transmit, radio	scatter, seed	ca 1920	(Jeffers and Lehiste, 1979)
monitor	display, screen	—	ca 1930	(Simpson et al., 1989)
record	tape, album	—	ca 1920	(Kulkarni et al., 2014)
guy	fellow, man	—	ca 1850	(Wijaya and Yeniterzi, 2011)
call	phone, message	—	ca 1890	(Simpson et al., 1989)

From Hamilton et al. (2016)a

Identified shifts

Method	Top-10 words that changed from 1900s to 1990s
PPMI	know, got, would, decided, think, stop, remember, **started**, must, wanted
SVD	harry, **headed**, **calls**, **gay**, wherever, male, **actually**, special, cover, naturally
SGNS	**wanting**, **gay**, **check**, **starting**, **major**, **actually**, touching, harry, **headed**, romance

Word	Language	Nearest-neighbors in 1900s	Nearest-neighbors in 1990s
wanting	English	lacking, deficient, lacked, lack, needed	wanted, something, wishing, anything, anybody
asile	French	refuge, asiles, hospice, vieillards, infirmerie	demandeurs, refuge, hospice, visas, admission
widerstand	German	scheiterte, volt, stromstärke, leisten, brechen	opposition, verfolgung, nationalsozialistische, nationalsozialismus, kollaboration

From Hamilton et al. (2016)a

conformity & innovation

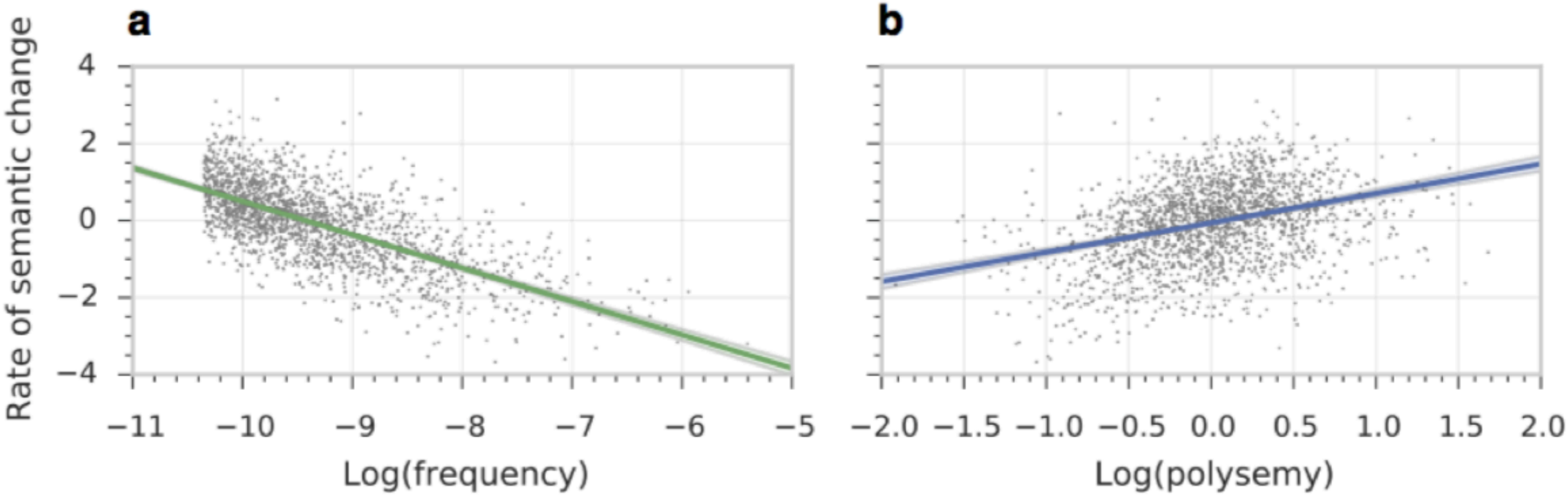
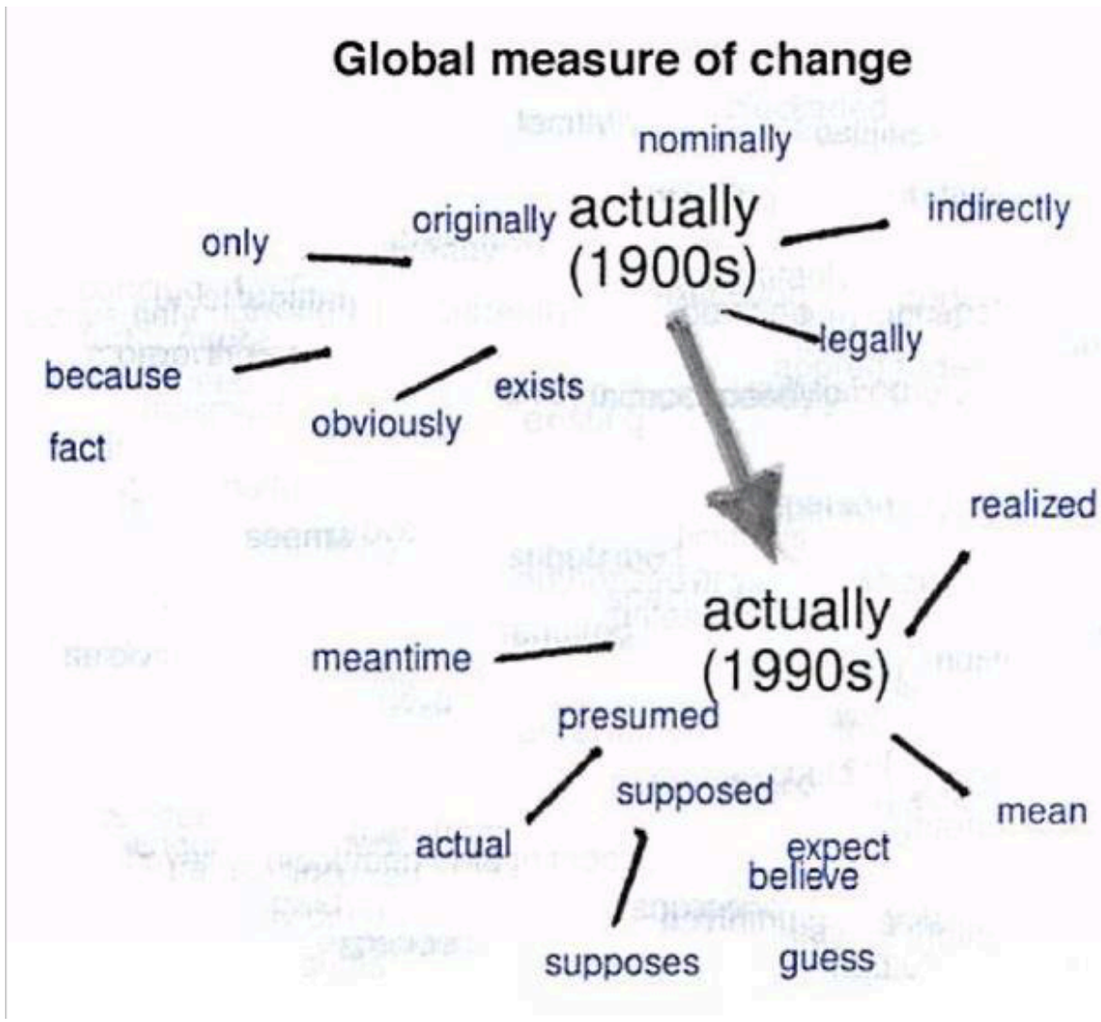
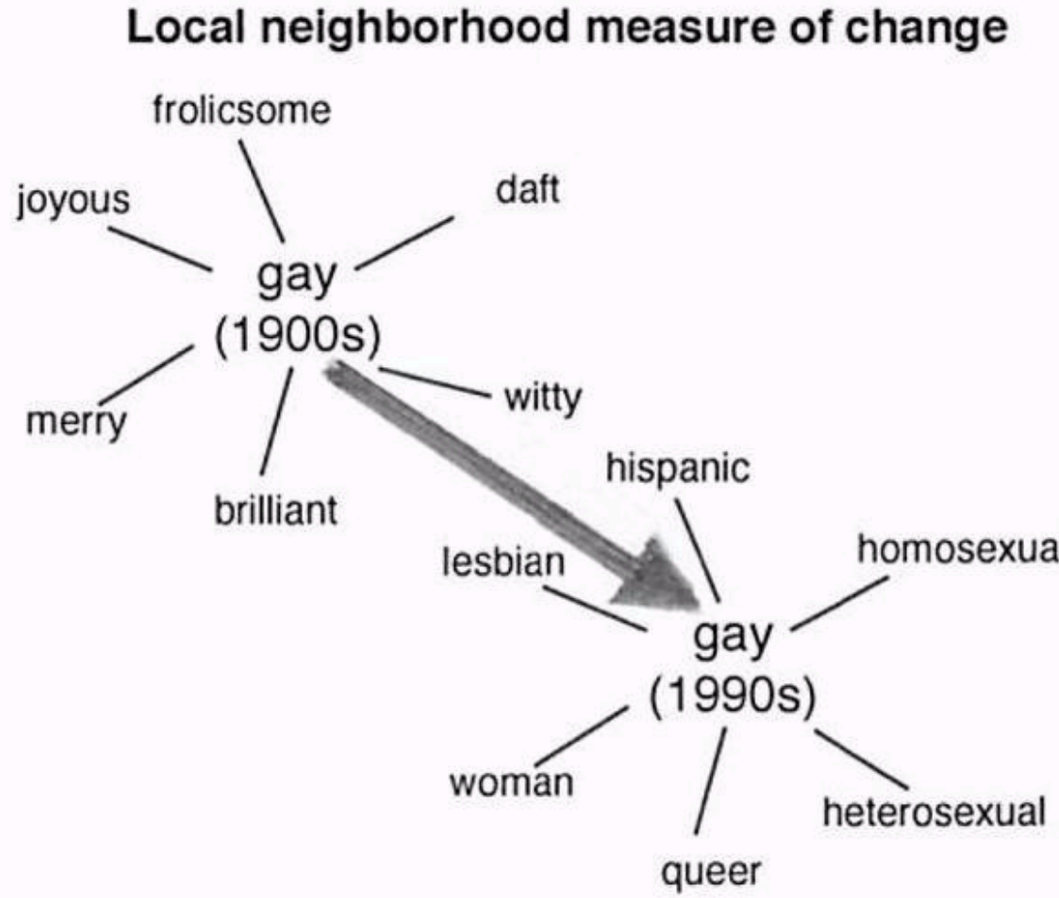


Figure 2: Higher frequency words have lower rates of change (**a**), while polysemous words have higher rates of change (**b**). The plots show robust linear regression fits (Huber, 2011) with 95% CIs on the 2000s decade of the COHA lemma data.

From Hamilton et al. (2016)a

Global vs Local Comparison





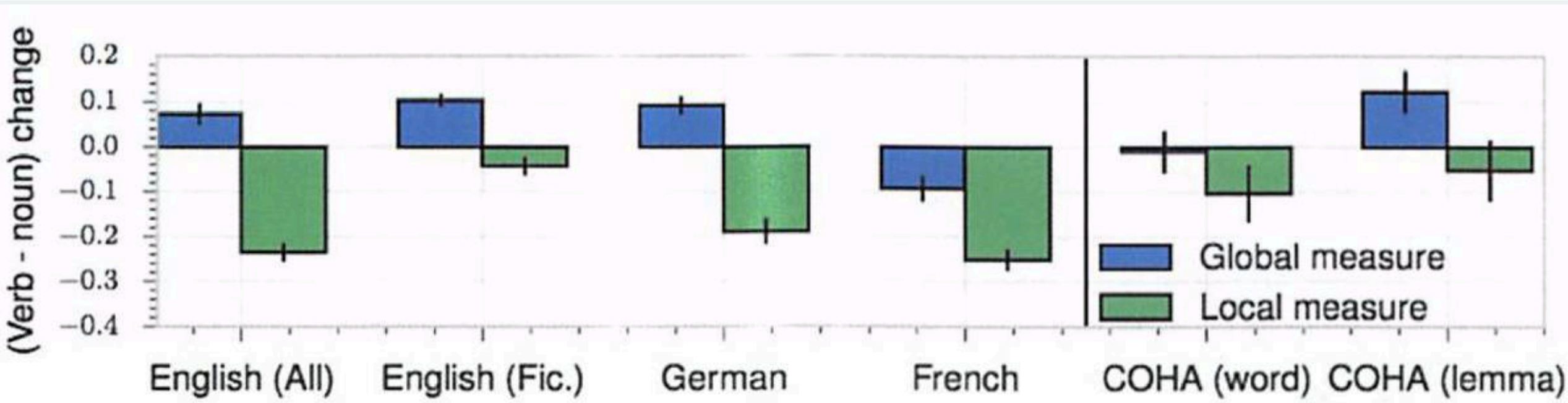
From Hamilton et al. (2016)b

Method

- Global change: as before (shift of vector itself from one to the next)

- Local change: secondary based on cosine similarity of the words 25-nearest-neighbors

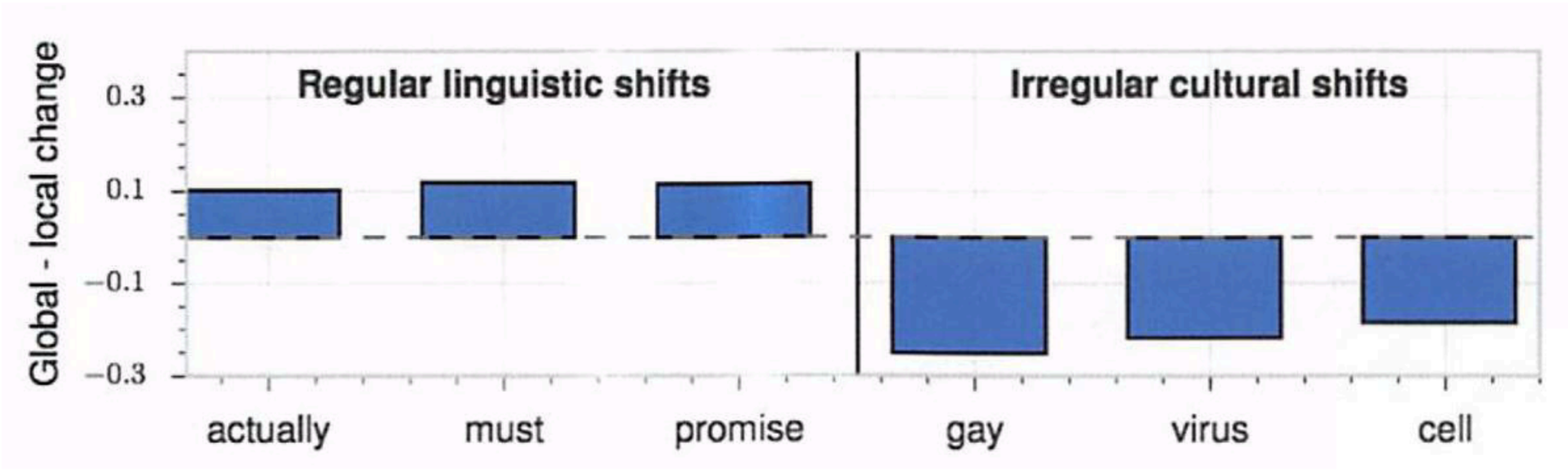
Nouns vs Verbs



From Hamilton et al. (2016)b

Case study

Word	1850s context	1990s context
actually	"…dinners which you have actually eaten."	"With that, I actually agree."
must	"O, George, we must have faith."	"Which you must have heard ten years ago…"
promise	"I promise to pay you…'	"…the day promised to be lovely."
gay	"Gay bridals and other merry-makings of men."	"…the result of gay rights demonstrations."
virus	"This young man is…infected with the virus."	"…a rapidly spreading computer virus."
cell	"The door of a gloomy cell…"	"They really need their cell phones."



From Hamilton et al. (2016)b

Global vs Local Change

- Why this difference?
- What does it mean?
- Polysemy/frequency effect?

Studying Shifts using Embeddings

- What problems might arise with these methods?

- How can these be addressed?

References

- Hamilton, W.L., Leskovec, J. and Jurafsky, D., 2016a. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (Vol. 1, pp. 1489-1501).
- Hamilton, William L., Jure Leskovec, and Dan Jurafsky. 2016b. Cultural shift or linguistic drift? comparing two computational measures of semantic change. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, p. 2116. NIH Public Access, 2016.
- Hellrich, Johannes, and Udo Hahn. 2016. Bad company—neighborhoods in neural embedding spaces considered harmful. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp. 2785-2796.
- Sommerauer, Pia. 2017. *From Old to New Racism? Investigating known dangers in distributional semantic approaches to conceptual change*. MA thesis, VU University.