

Words with consistent  
diachronic usage patterns  
are learned earlier.

A computational analysis using temporally  
aligned word embeddings



Cassani, Bianchi & Marelli

AMLaP 2020

# The hypothesis

*General:* There's a tight relation between how languages change over centuries and how people learn them during acquisition.

*Specific:* There's a unique relation between how stable the usage patterns over time of a word are and how early children learn it.

# How we addressed it

- CoHA + AoA norms + concreteness norms + SUBTLEX-US
- Temporally aligned Word Embeddings with a Compass (**TWEC**)
- Measures of diachronic coherence in language use:
  - Vector coherence (VC)
  - Local neighborhood coherence (LNC)
  - Jaccard (J) coherence

# Stability of usage patterns

We used the **Temporally-aligned Word Embeddings with a Compass** (TWEC, Di Carlo & al, 2019): for each time slice in a corpus, TWEC generates an embedding for each word.

These embeddings are aligned: words with stable meaning will have similar vectors along all the time-periods.

TWEC modifies the training process of CBoW (Mikolov & al, 2013):

- first creates a general embedding (the **compass**)
- then aligns time-specific embeddings using the compass as a constraint
- after training, the time-specific embeddings are aligned

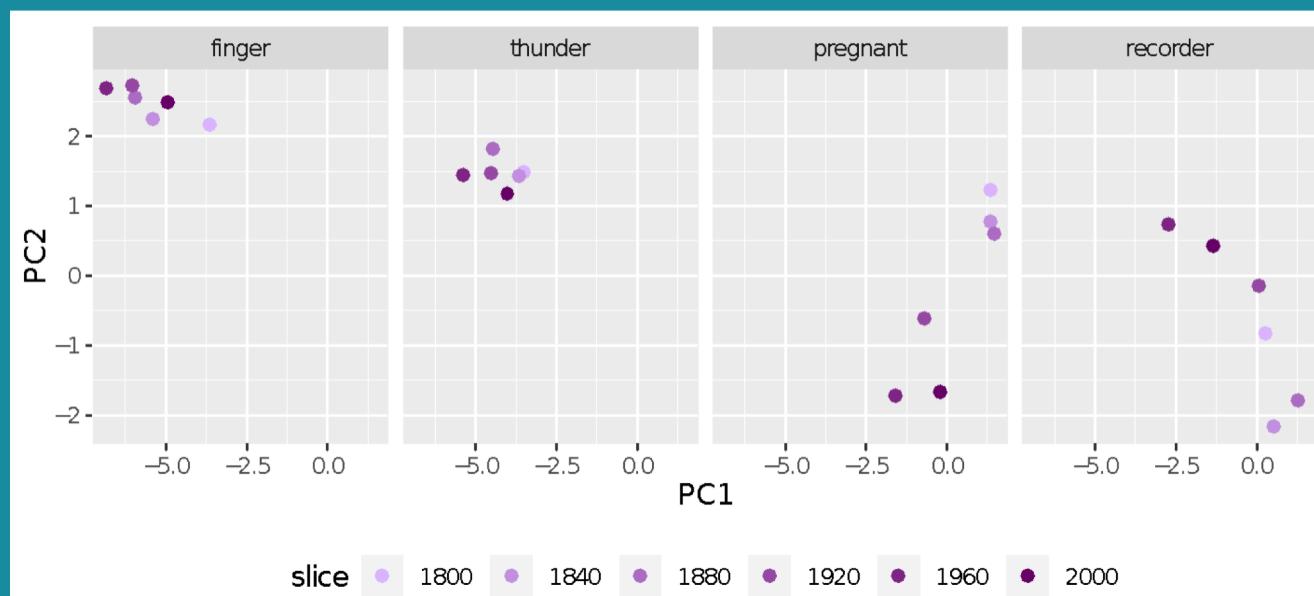
# Measure of diachronic stability

- **Vector Coherence (VC)**: how close to each other are the word embeddings of a same word in different time slices?
- **Local Neighborhood Coherence (LNC)**: how similar is the relation between a word embedding and its nearest neighbors' embeddings in different time slices?
- **Jaccard (J) Coherence**: how coherent are the nearest neighbors of a word embedding in different time slices?
- **Random baselines** for each (rVC, rLNC, rJ) to make sure any relation with AoA is not due to other properties of the embedding spaces.

# Upgrades

- State of the art methods in NLP
- Scalable, data-driven operationalization of language change
- Puts language use front and center
- AoA norms as a dependent variable
  - no variable acting as a proxy for unobserved constructs
  - more epistemologically sound to predict human intuitions from data

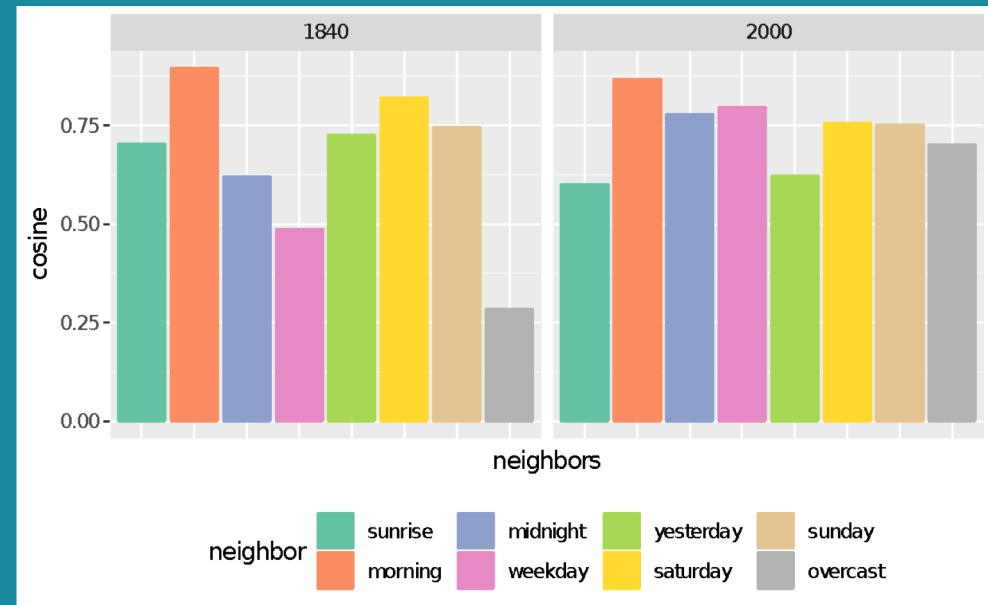
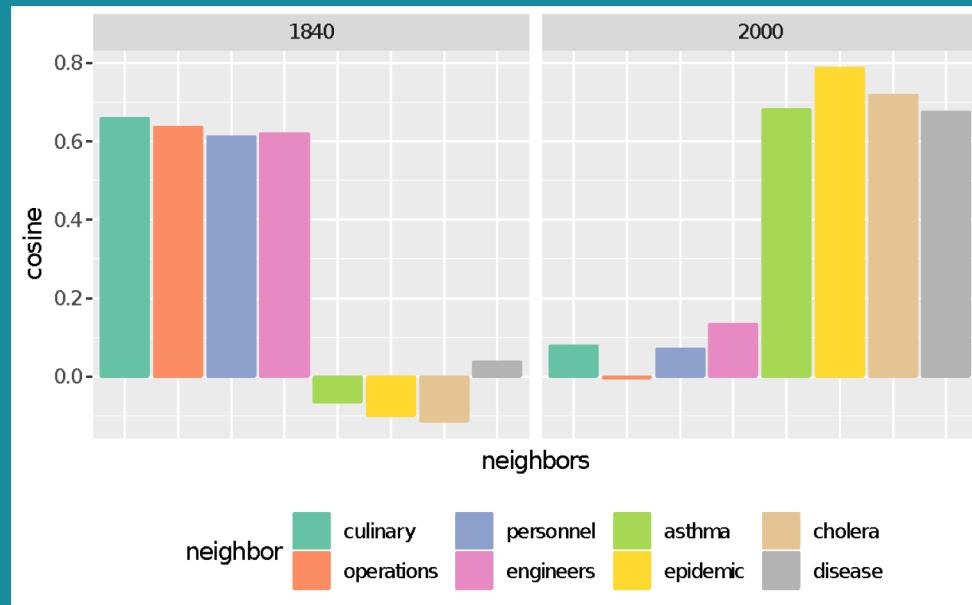
# Vector coherence



**High:** body parts, basic level concepts, time expressions, specialized meanings

*Low:* technological changes, cultural shifts, sexual sphere

# Local Neighborhood Coherence



High: (less prototypical) body parts and time expressions, animals, fruit, non-arbitrary form-meaning mappings, suffering

Low: technological changes, cultural shifts, sexual sphere

# Jaccard Coherence

Target	Neighbors (1840)	Neighbors (2000)	J
fifth	sixth, fourth, eighth, ninth, tenth, seventh, twentieth, <b>twelfth</b> , thirtieth, fourteenth, <b>second</b> , fortieth, inclusive, fiftieth, <b>eleventh</b> , thirteenth, <b>fifteenth</b> , sixteenth, thirty, seven	fourth, seventh, sixth, ninth, eighth, tenth, 12th, 13th, <b>eleventh</b> , 11th, 14th, 33rd, 10th, <b>fifteenth</b> , consecutive, <b>twelfth</b> , 34th, 15th, 16th, <b>second</b>	5.39
aids	engineering, auxiliary, superintendence, tactics, desideratum, adjunct, culinary, housewifery, procurement, ception, ose, operations, commissariat, offi, reconnaissance, portant, ciety, engineers, cult, department	hiv, std, epidemic, trachoma, syphilis, malaria, hodgkin, cholera, communicable, polio, cancer, alzheimer, lupu, ebola, lupus, asthma, ovarian, testicular, tuberculosis, colon	0.22

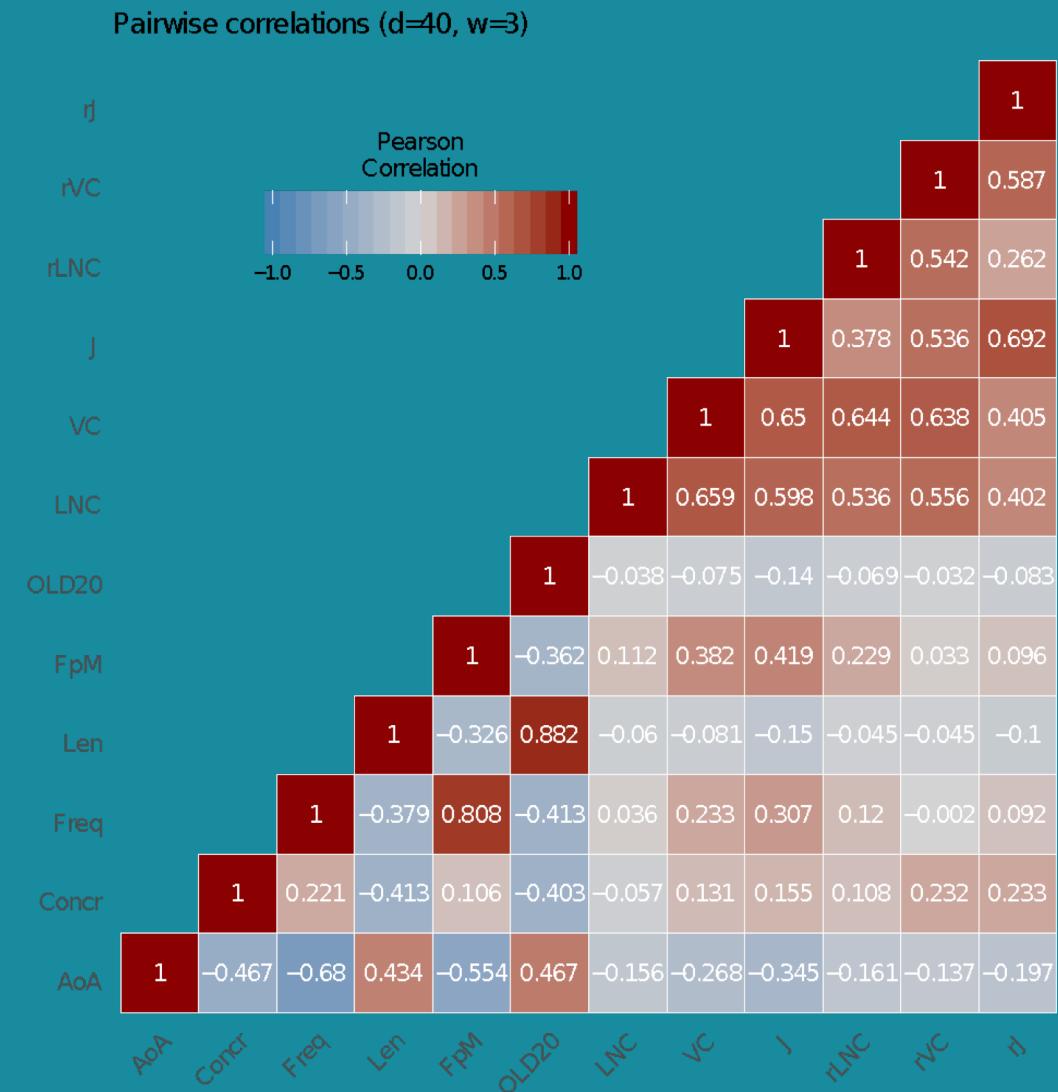
**High:** time expressions, numerals, kinship relations, suffering, fruit, basic level concepts, non-arbitrary form-meaning mappings

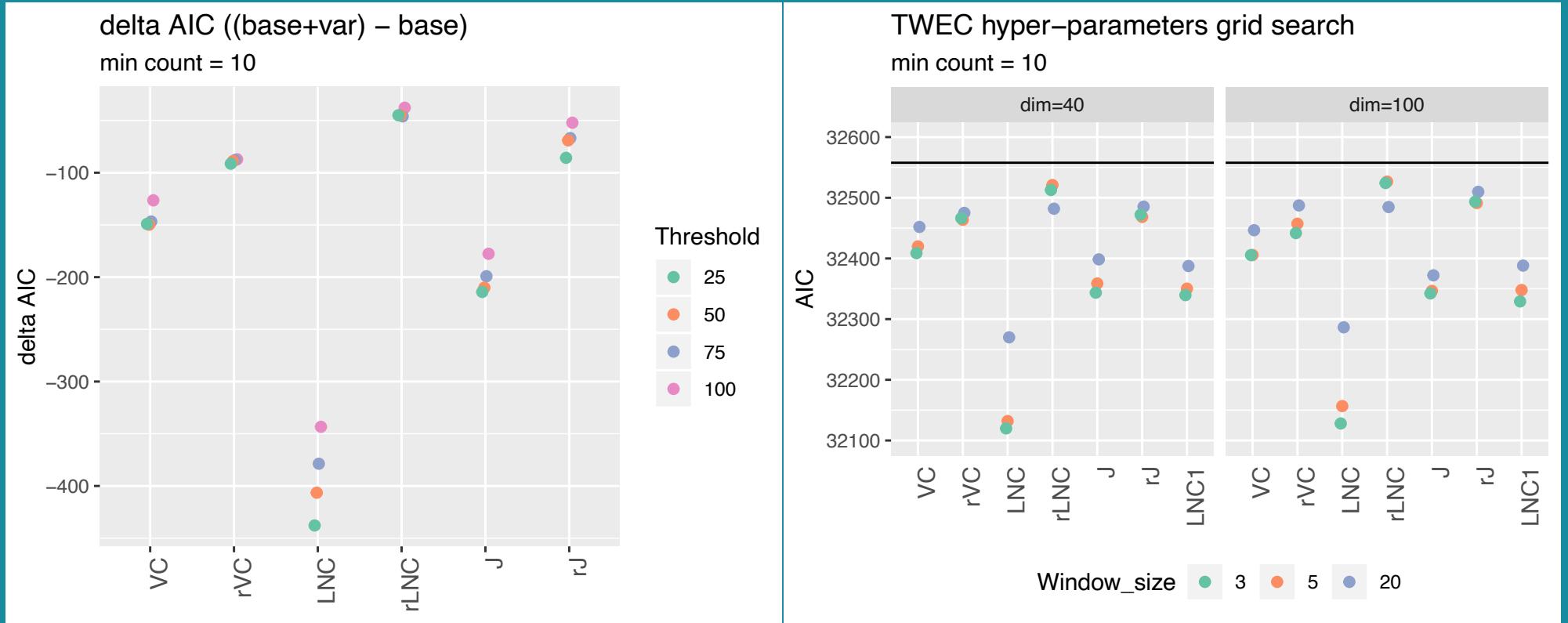
**Low:** technological changes, cultural shifts, sexual sphere

# Pairwise correlations

Measures of language change correlate but don't capture the same thing.

They all have non-zero correlations with AoA, but VC and J have stronger correlations with frequency and concreteness than LNC.





## A unique effect

All measures of language change explain more variance than a baseline model including frequency, concreteness, phonological neighborhood density, length and dominant PoS tag, with LNC bringing the highest improvement.

We conclude that there's a unique relation between language learning and change: **words with more stable usage patterns over time are learned earlier.**

# Thanks!

[g.cassani@tilburguniversity.edu](mailto:g.cassani@tilburguniversity.edu)

[f.bianchi@unibocconi.it](mailto:f.bianchi@unibocconi.it)

[marco.marelli@unimib.it](mailto:marco.marelli@unimib.it)