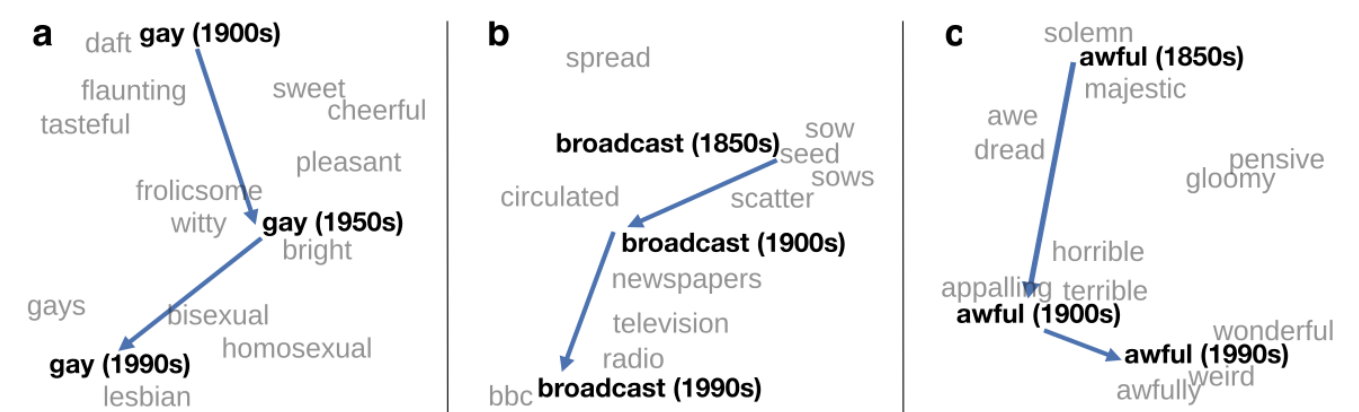


Statistical Mechanics of Semantic Evolution

Bogdán Asztalos¹, Gergely Palla², Dániel Czégel³

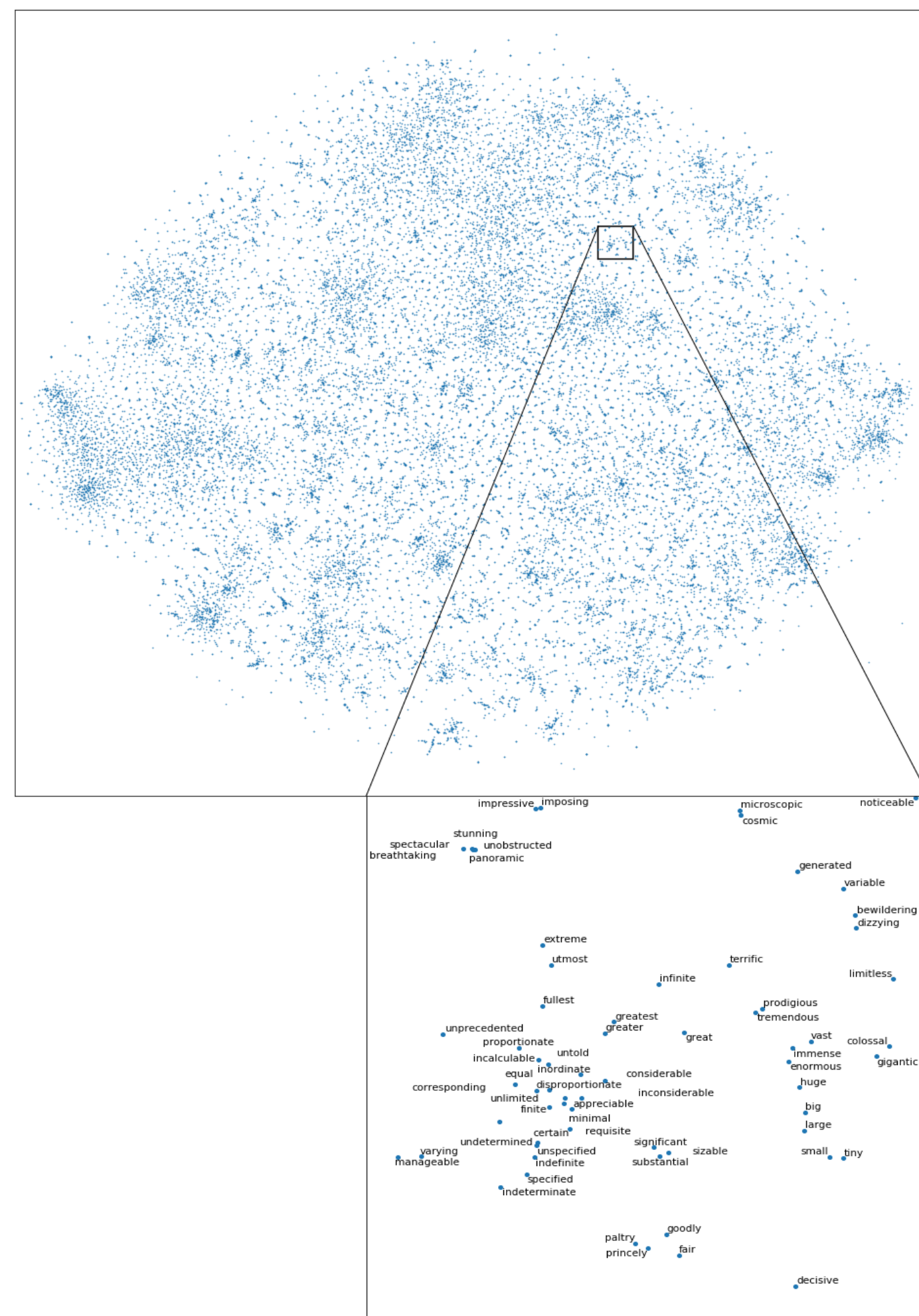
1. Department of Biological Physics, Eötvös University, Hungary;
2. MTA-ELTE Statistical and Biological Physics Research Group, Hungary;
3. Institute of Evolution, Centre for Ecological Research, Hungary

Semantic change and context change



Source of figure: [1].

A set of embedded words



INTRODUCTION

Discovering statistical laws of semantic change might point at mechanisms underlying our collective cumulative cultural evolution, the very feature that distinguishes us humans from the rest of the animal world.

Distributional hypothesis states that the meaning of a word can be characterized by the context in which it occurs, so the change of meaning can be observed through the statistical change of the context. This property allows us to study semantic dynamics with the tools of statistics and statistical physics.

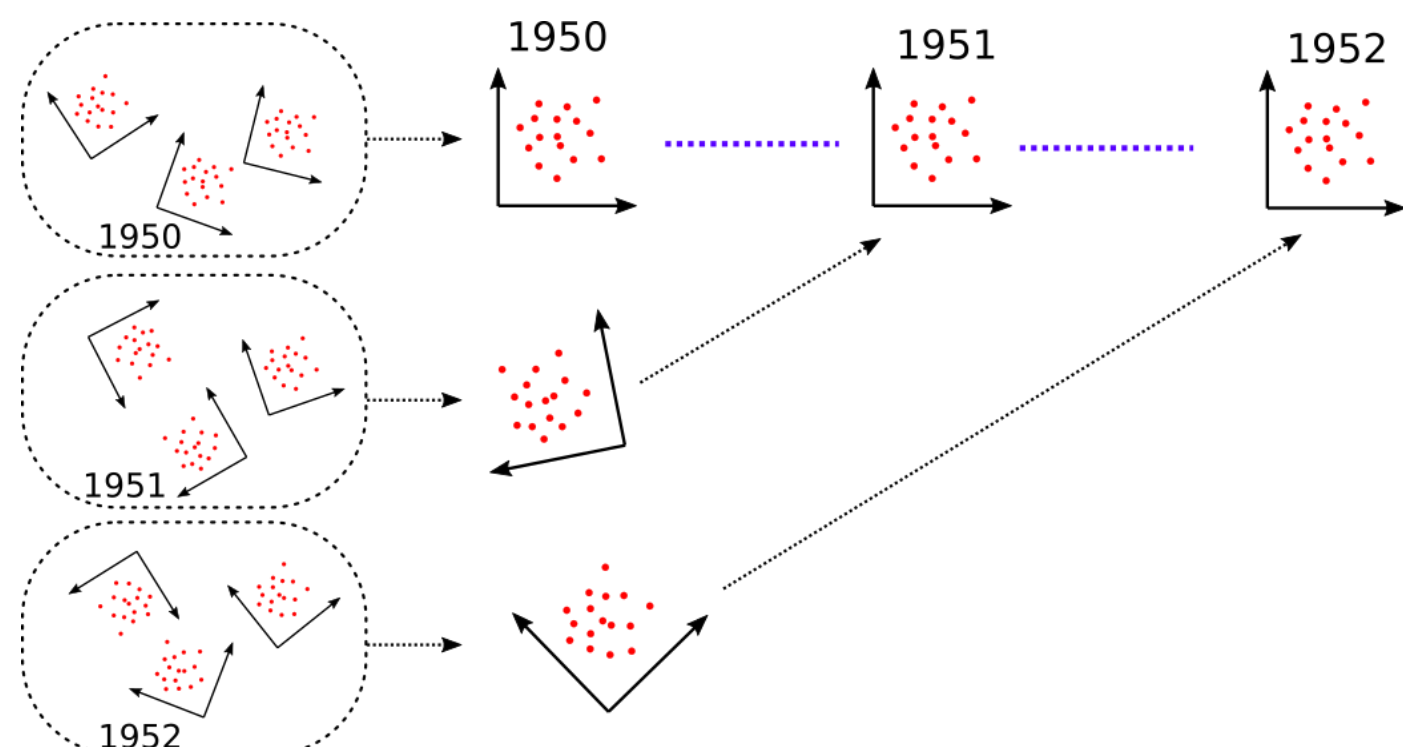
WORD EMBEDDING

Word Embedding is a tool of natural language processing where words are represented as vectors of real numbers. Word2vec's skip-gram model is an efficient neural network based algorithm that embeds words in a several hundred-dimensional space while approximately keeping the semantic distance between words: The more similar the meaning of two words are, the closer they are placed in the embedded space. [2-3]

ALIGNING DIACHRONIC EMBEDDINGS

Time-dependent behavior of words can be studied by construct word embedding from data of different time-periods. Embedding from different data results different sets of points and due to the stochastic nature of Word2vec, axes of embedding space are just meaningless abstractions and they cannot be compared. Since two (close) states of language are similar the structure of the two embeddings must look alike so they can be aligned by finding the transformation that takes the same words closest to each other. The most general transformation that keeps the information of an embedding is orthogonal so aligning must be a rotation.

For specific studies we used 300 dimensional embeddings. To reduce measurement error we constructed embeddings 80 times from data of each time period, aligned them and took their average. Finally, we aligned the period-averages to each other getting the trajectory of words.

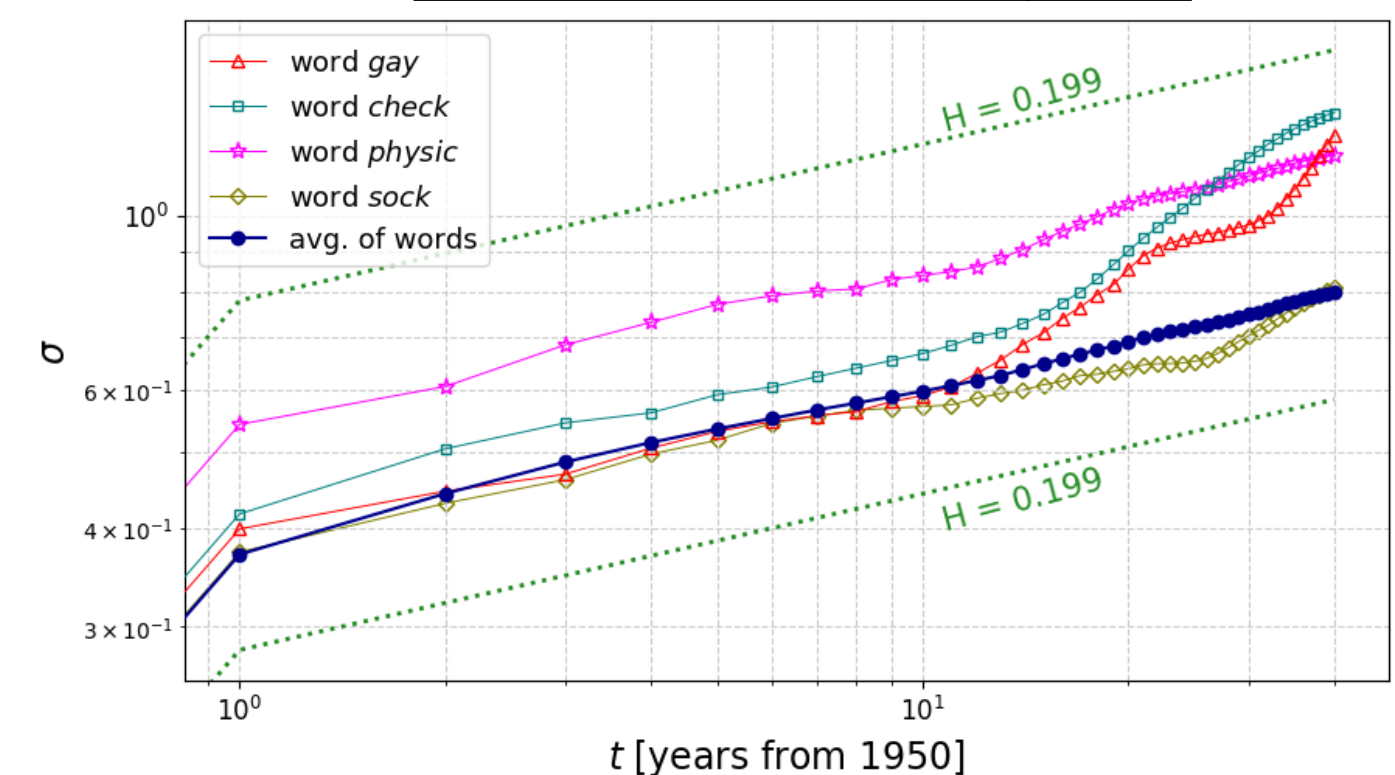


SUBDIFFUSION

Motion of most words can be described by subdiffusion with Hurst exponent $H \approx 0.2$, in contrast with diffusing particles that follow $H \approx 0.5$. To measure Hurst exponent we plotted the cumulative deviation (σ) of positions where the word was observed in the function of the time (t) pasted from the date of the first embedded data (1950). On log-log scale a linear function could be fitted on these points and its slope meant to be the Hurst exponent.

We generated random walks by randomize step-size and step-direction data of words. Totally random sizes and directions give back Brownian motion with $H \approx 0.5$, but the more property reproduced the more different H-value is got. This means that part of this deviation comes from non-independent directions, part of it comes from non-independent velocities at subsequent times.

Measurement of Hurst exponent



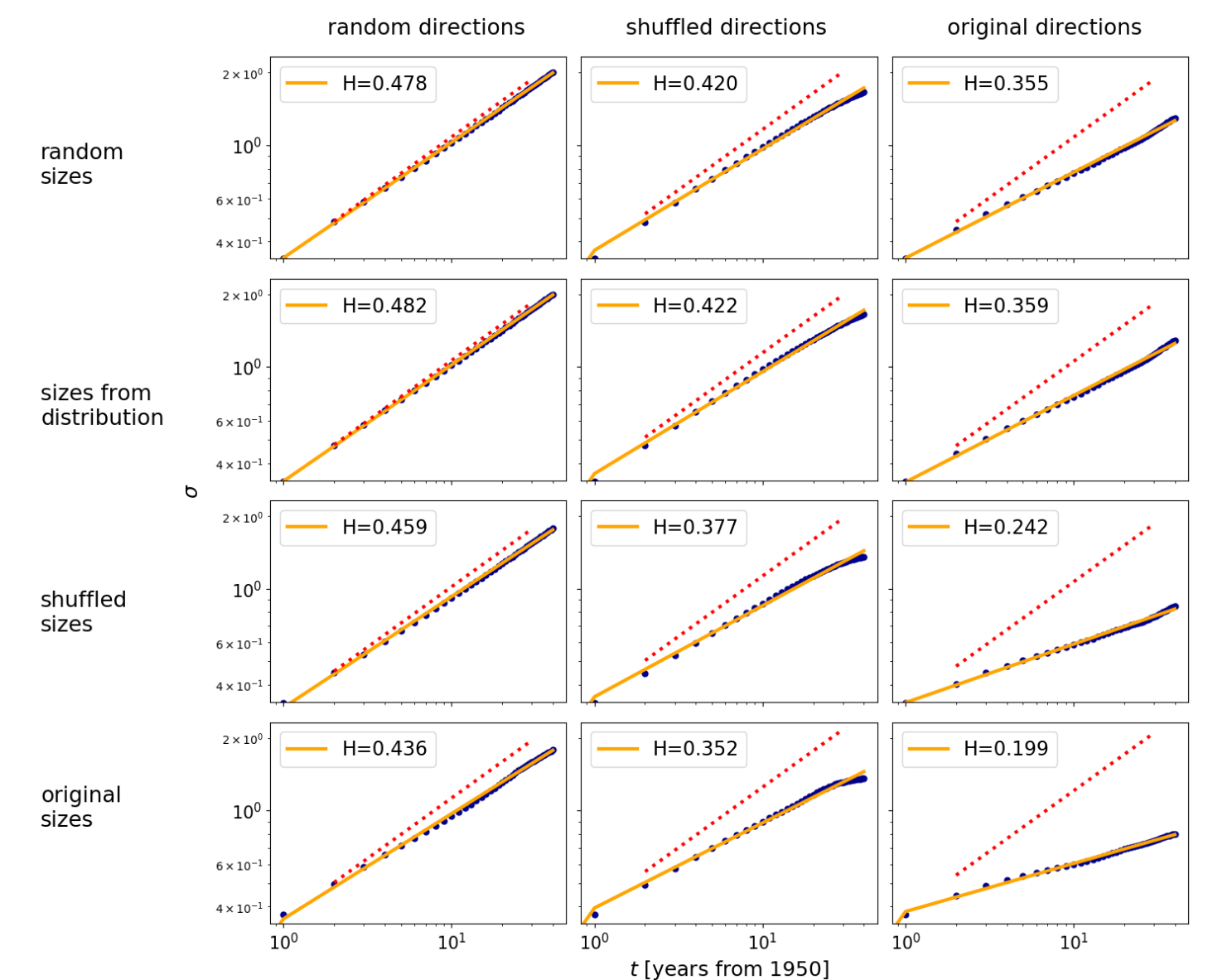
CONCLUSION

We showed that if an 'ensemble of trajectories' perspective is taken, there is one law that holds universally across languages: that semantic evolution is subdiffusive. Using a pipeline of diachronic distributional semantic embeddings we show that words follow a stochastic trajectories with average Hurst exponent $H \approx 0.2$. Yet neither observed step size distributions, nor memory in single trajectories indicate subdiffusion, which is therefore a more complicated, ensemble-level phenomenon.

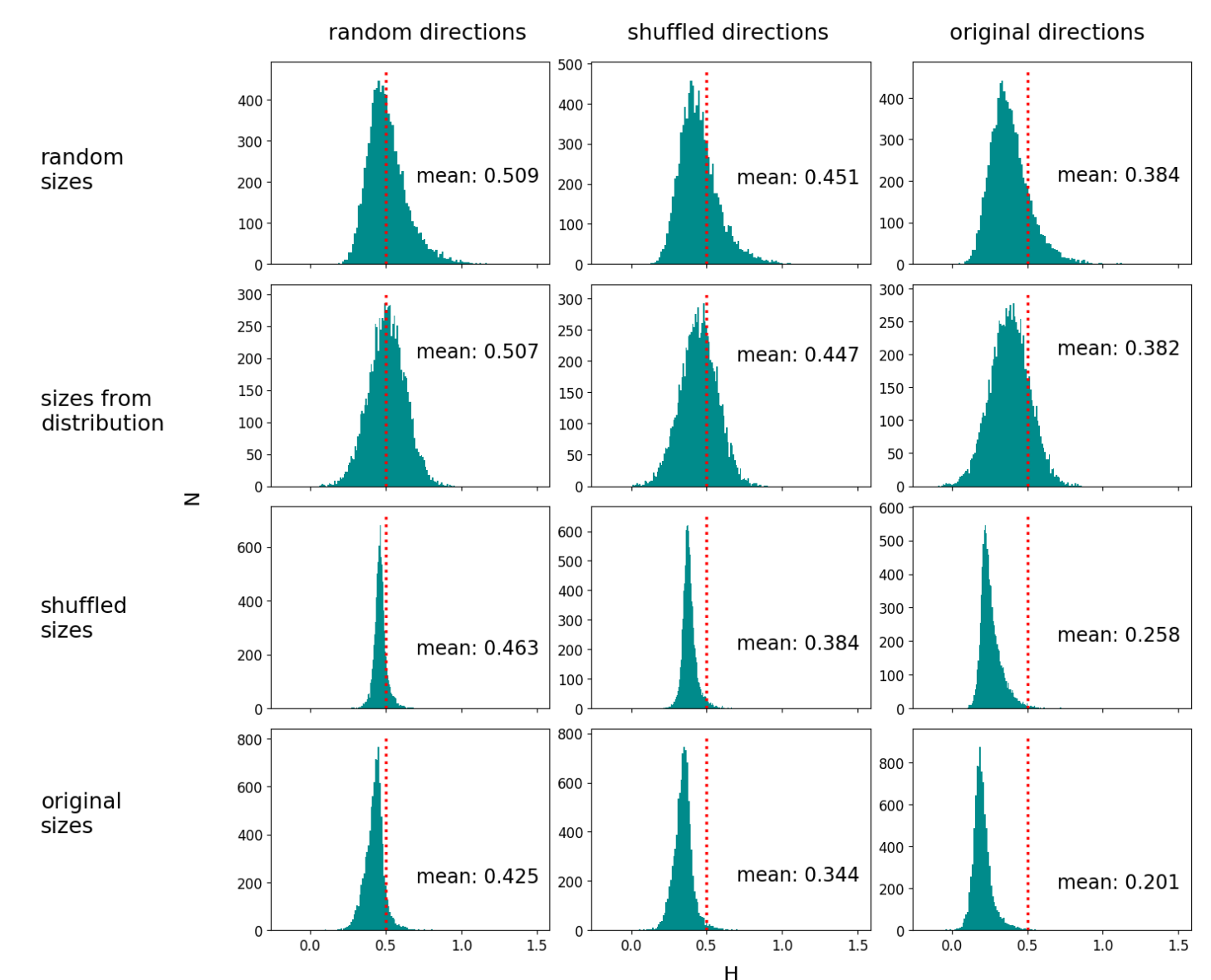
REFERENCES

- [1] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. „Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change.” *ACL* 2016.
- [2] Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems*. 2013.
- [3] Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013).

Hurst exponents in random walks



H-exp. distributions in random walks



THE RESEARCH WAS PARTIALLY SUPPORTED BY THE HUNGARIAN NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION OFFICE (GRANT NO. K 128780).