Frequent words are semantically more stable than rare ones: what computational modeling, corpus analysis, and psycholinguistic databases can tell us about lexico-semantic change

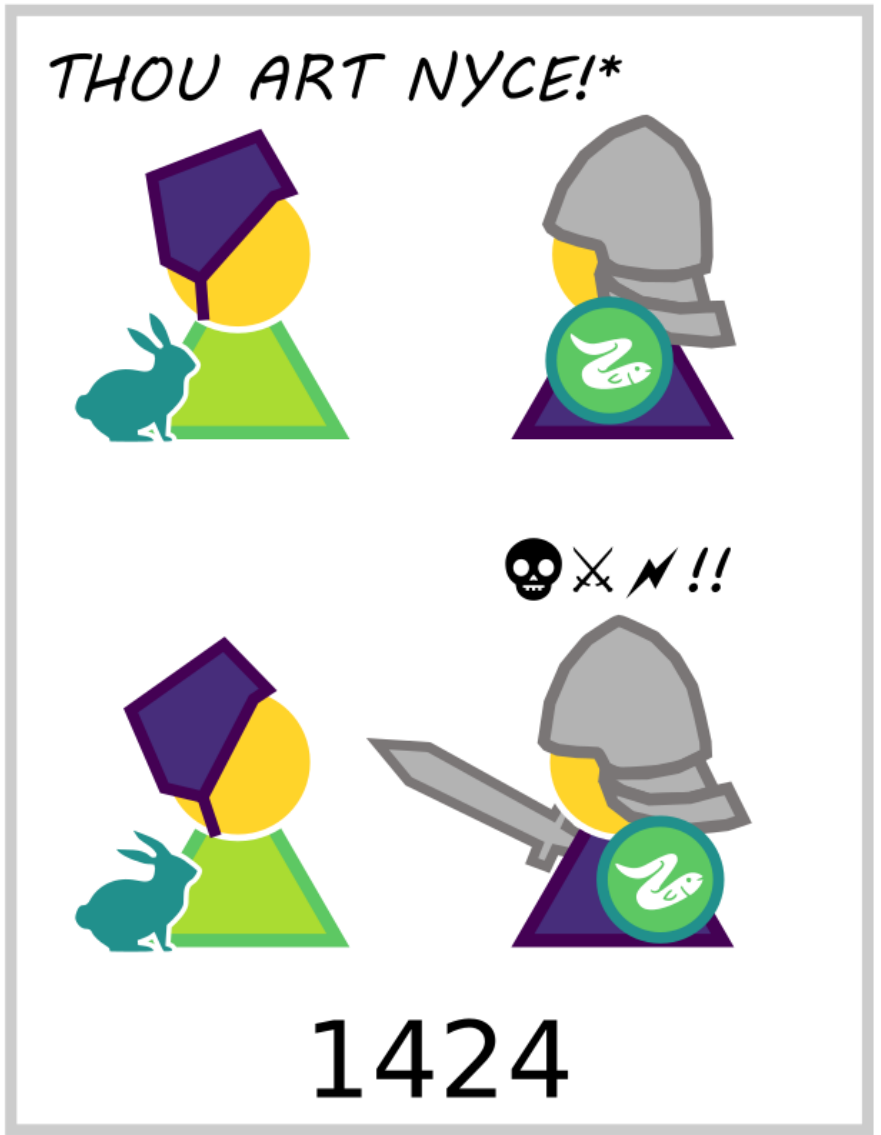
Andreas Baumann

University of Vienna, Department of German Studies & Research Network Data Science

https://www.germ.univie.ac.at/andreas-baumann

andreas.baumann@univie.ac.at


YOU ARE NICE!
THANK YOU!
2024

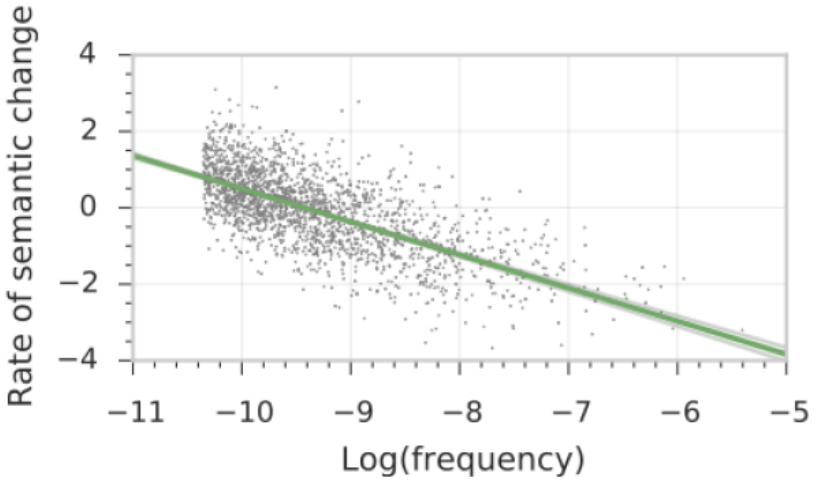


*Late Middle English: 'you are silly'

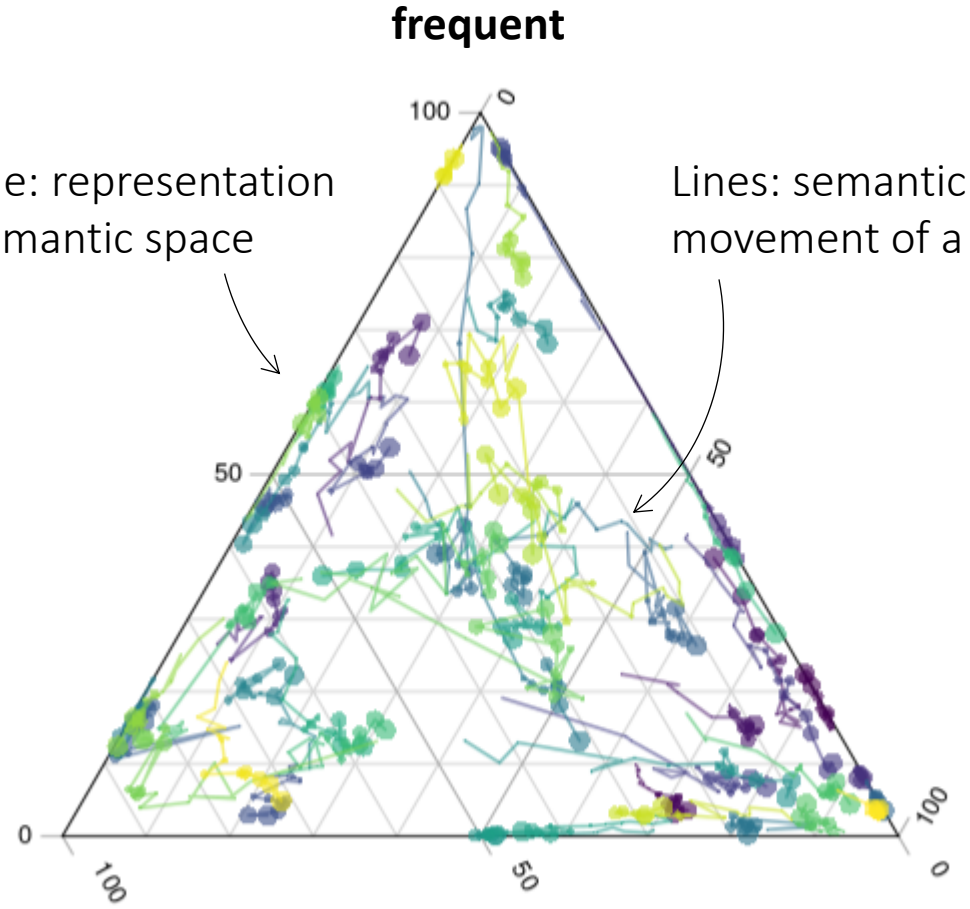
Words change their meaning

- Words **change** their **meaning** when used in novel contexts
  - Word semantics can **shift** (e.g., *nice*: negative → positive)
  - Words can obtain **additional senses** (e.g., *mess*: soup, cafeteria, disorder)
- Which **factors** drive semantic change?
  - Often examined: utterance **frequency**
  - Idea: entrenched words resist change more easily

Hamilton et al. (2016, *ACL*): rate of semantic change decreases with frequency



Words change their meaning

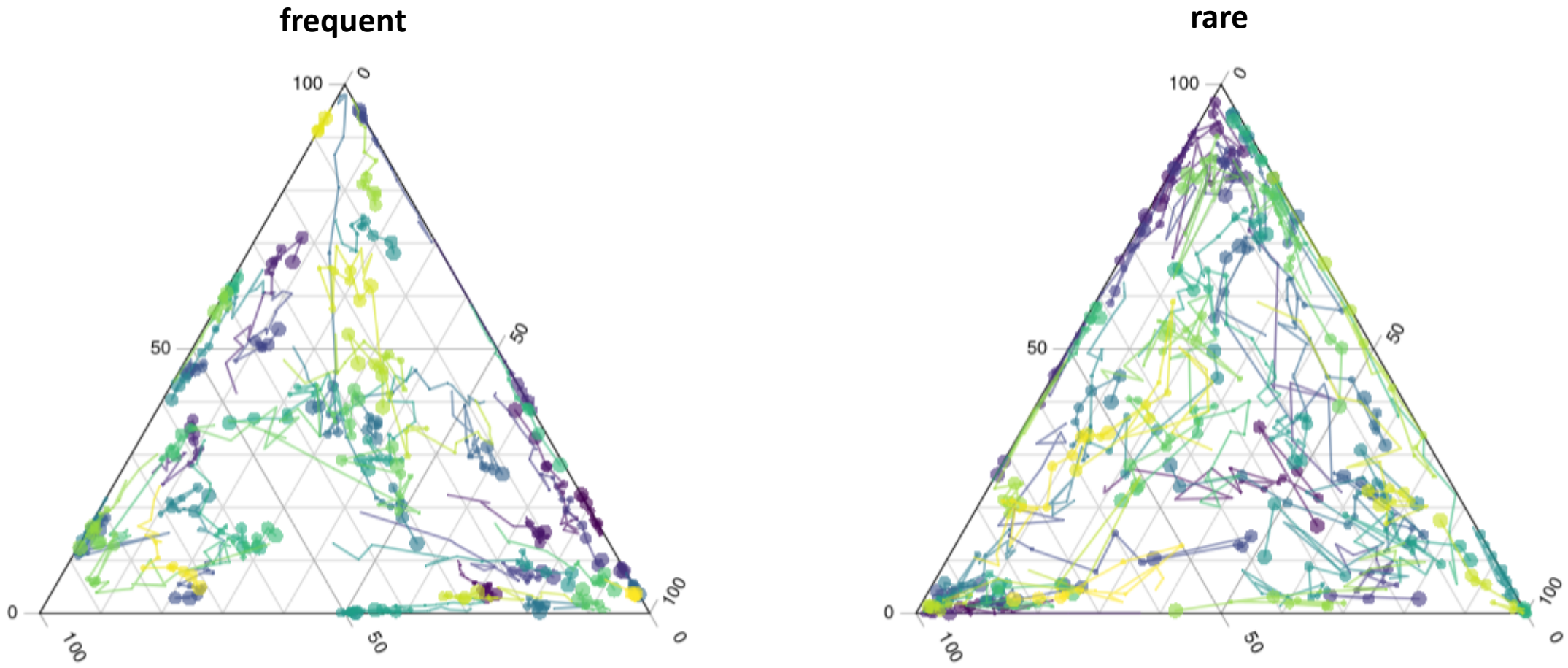


Triangle: representation of a semantic space

Lines: semantic movement of a word

Semantic movements of words in COHA with three senses according to OED (points correspond to probability distributions over 3 senses; each line corresponds to a single word; top/bottom 50 words displayed).

Words change their meaning



Semantic movements of words in COHA with three senses according to OED (points correspond to probability distributions over 3 senses; each line corresponds to a single word; top/bottom 50 words displayed).

Words change their meaning

- Here: three studies about the role of **frequency**
  - Study 1: frequency and semantic *variability*
  - Study 2: frequency and semantic *diversification*
  - Study 3: interaction of frequency and acquisition in semantic *variability*, *diversification*, and *displacement*
- **Data and methods**:
  - Digitized diachronic text data (eng/ger)
  - Acquisition data (eng)
  - Computational modeling of semantics
  - Population dynamics
  - Quantitative analysis of semantic change

Study 1: frequency and variability

- What is the effect of **frequency** on the semantic **variability** (vs. stability) of words?
- Hamilton et al. (2016): 200 years (eng)
- Do we see similar effects on a **shorter time scale** as well?

DE GRUYTER MOUTON Cognitive Linguistics 2023; 34(3–4): 533–568

Andreas Baumann, Klaus Hofmann*, Anna Marakasova, Julia Neidhardt and Tanja Wissik

Semantic micro-dynamics as a reflex of occurrence frequency: a semantic networks approach

https://doi.org/10.1515/cog-2022-0008
Received February 4, 2022; accepted September 8, 2023; published online October 19, 2023

Abstract: This article correlates fine-grained semantic variability and change with measures of occurrence frequency to investigate whether a word's degree of semantic change is sensitive to how often it is used. We show that this sensitivity can be detected within a short time span (i.e., 20 years), basing our analysis on a large corpus of German allowing for a high temporal resolution (i.e., per month). We measure semantic variability and change with the help of local semantic networks, combining elements of deep learning methodology and graph theory. Our micro-scale analysis complements previous macro-scale studies from the field of natural language processing, corroborating the finding that high token frequency has a negative effect on the degree of semantic change in a lexical item. We relate this relationship to the role of exemplars for establishing form–function pairings between words and their habitual usage contexts.

Keywords: semantics; diachronic linguistics; corpus linguistics; semantic networks; German

1 Introduction

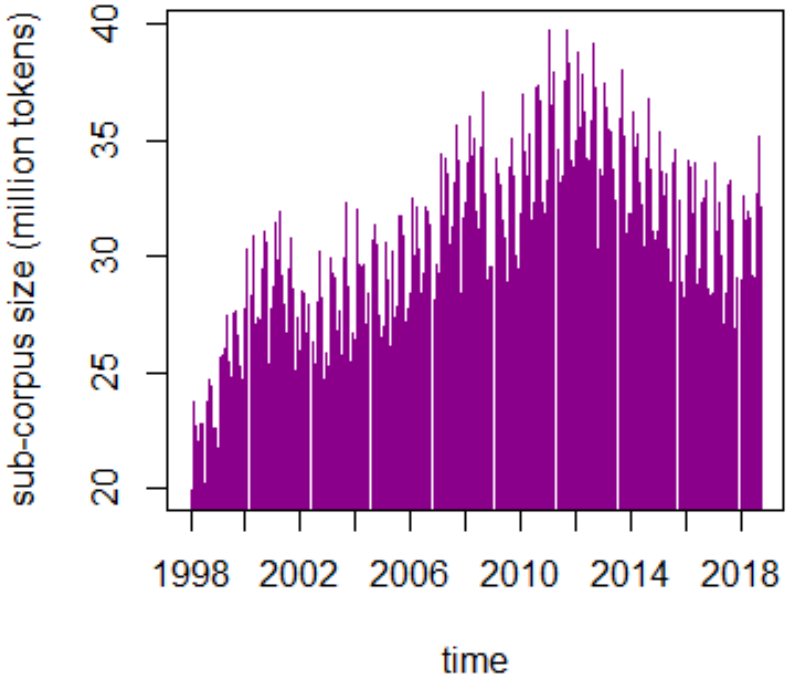
Semantic change is among the most conspicuous forms of diachronic variation found in language. When Romeo kills Tybald in the wake of a "nice […] quarrel" (*Romeo and Juliet*, Act 3, Scene 1), the average modern reader will invariably pause and

***Corresponding author: Klaus Hofmann**, Department of English and American Studies, University of Vienna, Vienna, Austria, E-mail: klaus.hofmann@univie.ac.at
Andreas Baumann, Department of European and Comparative Literature and Language Studies (EVSL), University of Vienna, Vienna, Austria
Anna Marakasova and Julia Neidhardt, Vienna University of Technology (TU Wien), Vienna, Austria
Tanja Wissik, Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH), Austrian Academy of Sciences, Vienna, Austria

Baumann, Hofmann, Marakasova, Neidhardt & Wissik, 2023, *Cognitive Linguistics*.

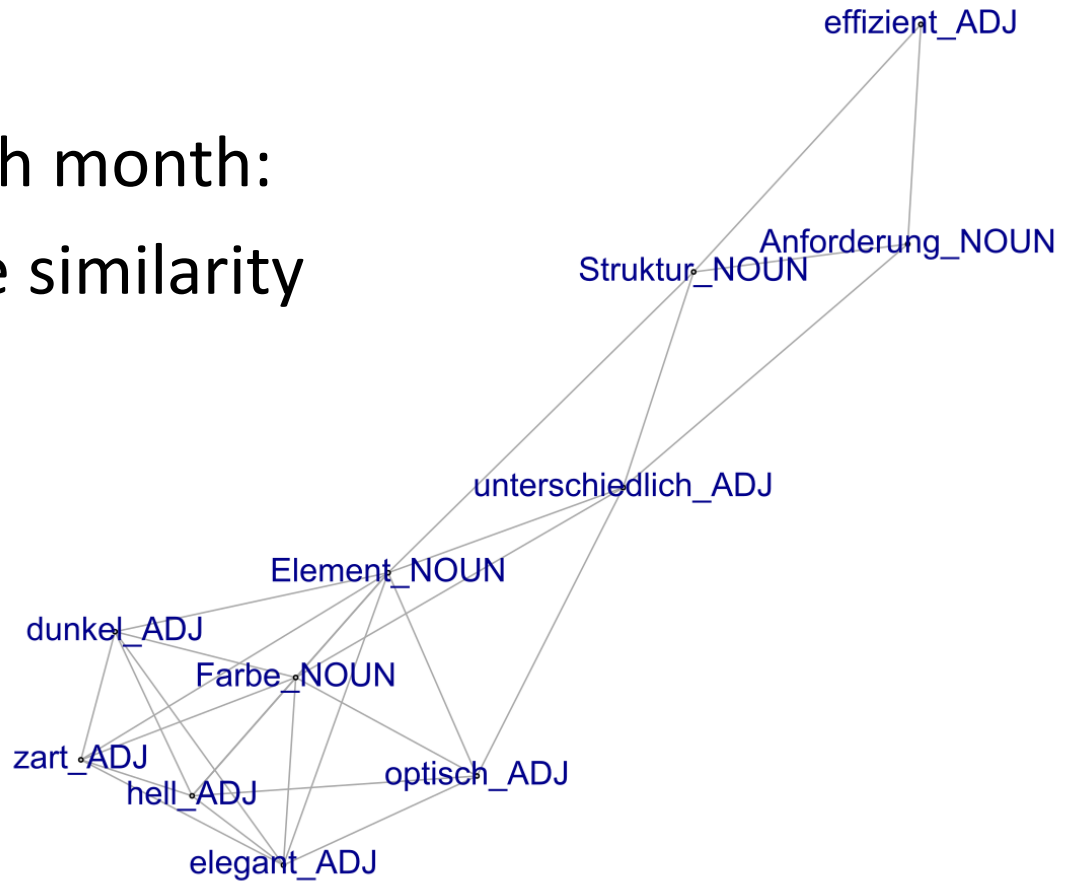
Study 1: data

- Austrian Media Corpus (Ransmayr et al. 2017)
  - Austrian German
  - 20 years (1997-2017)
  - 11 billion word tokens
  - PoS tagged
- Balanced sample of **~3000 target words**
  - wrt frequency, growth, fluctuation
- Trained skipgram embeddings for each month (Mikolov et al. 2013)
  - high temporal resolution
  - one vector per month per word



Study 1: lexical networks

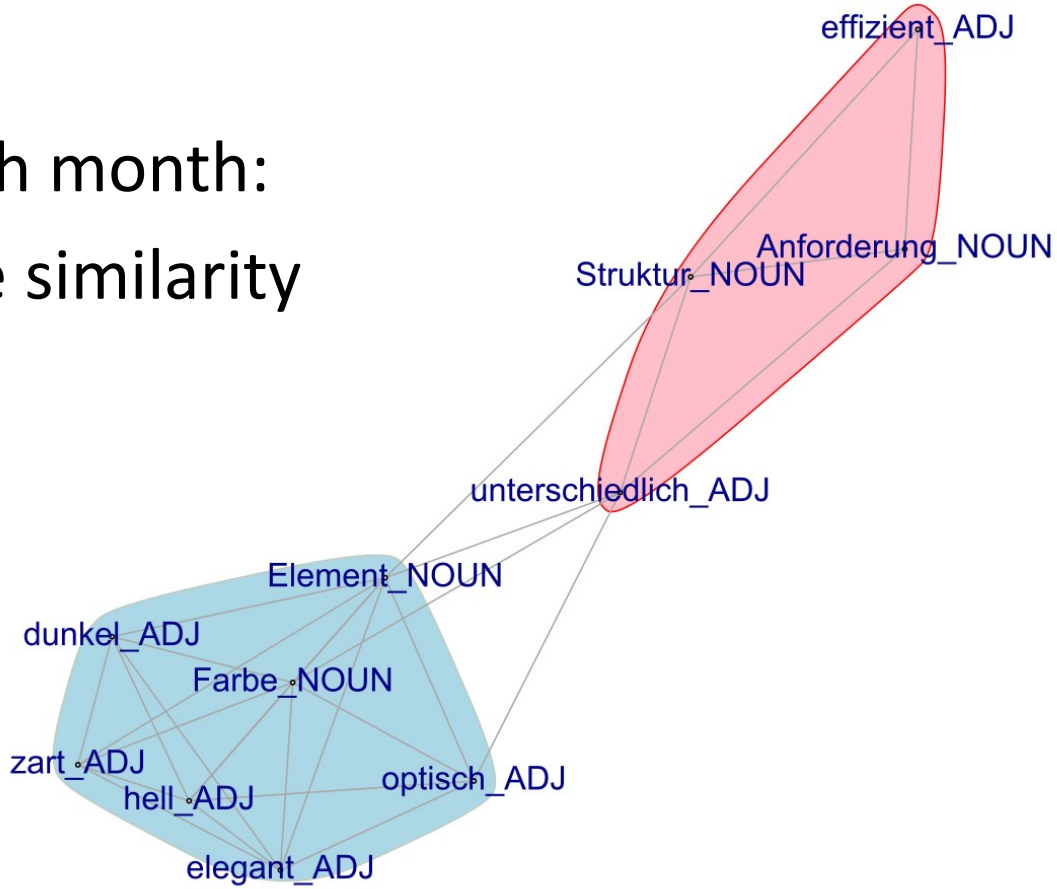
- For each target word and each month:
- **Ego-network** based on cosine similarity between vectors



Network of the word *transparent* in one month

Study 1: lexical networks

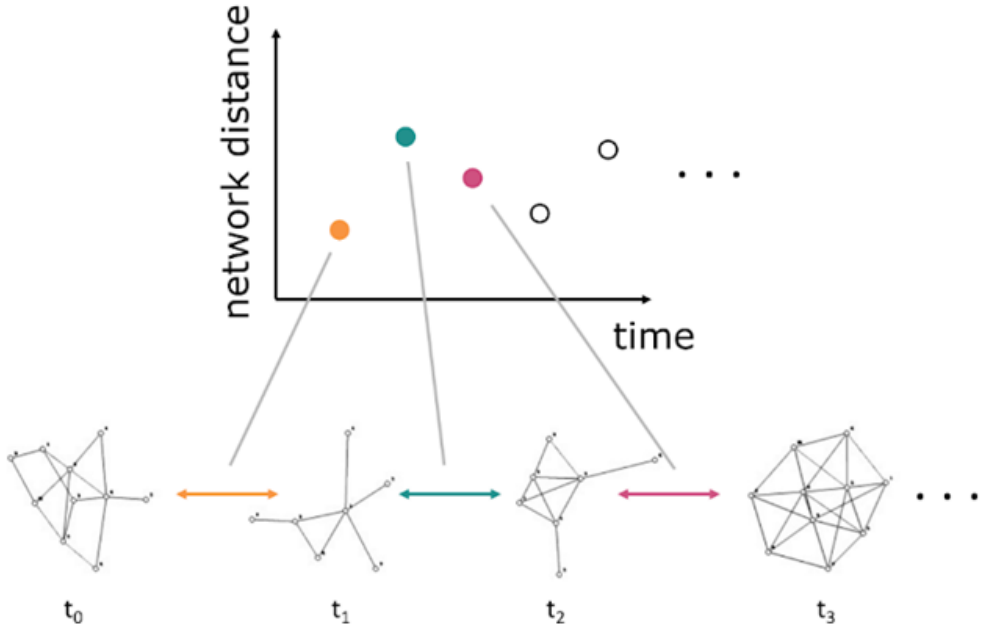
- For each target word and each month:
- **Ego-network** based on cosine similarity between vectors



Network of the word *transparent* in one month

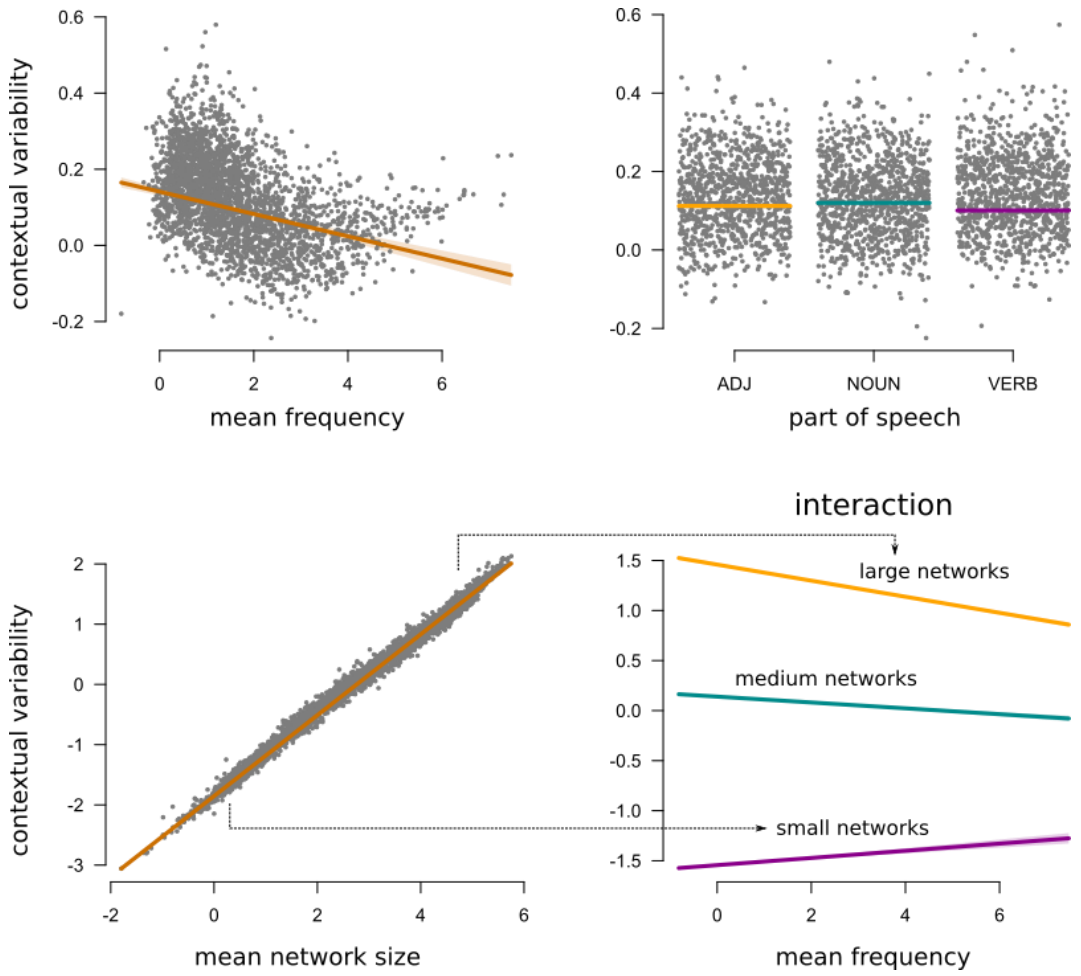
Study 1: measuring variability

- For each target word:
- Monthly distance between consecutive networks
- **Variability ~ average distance**



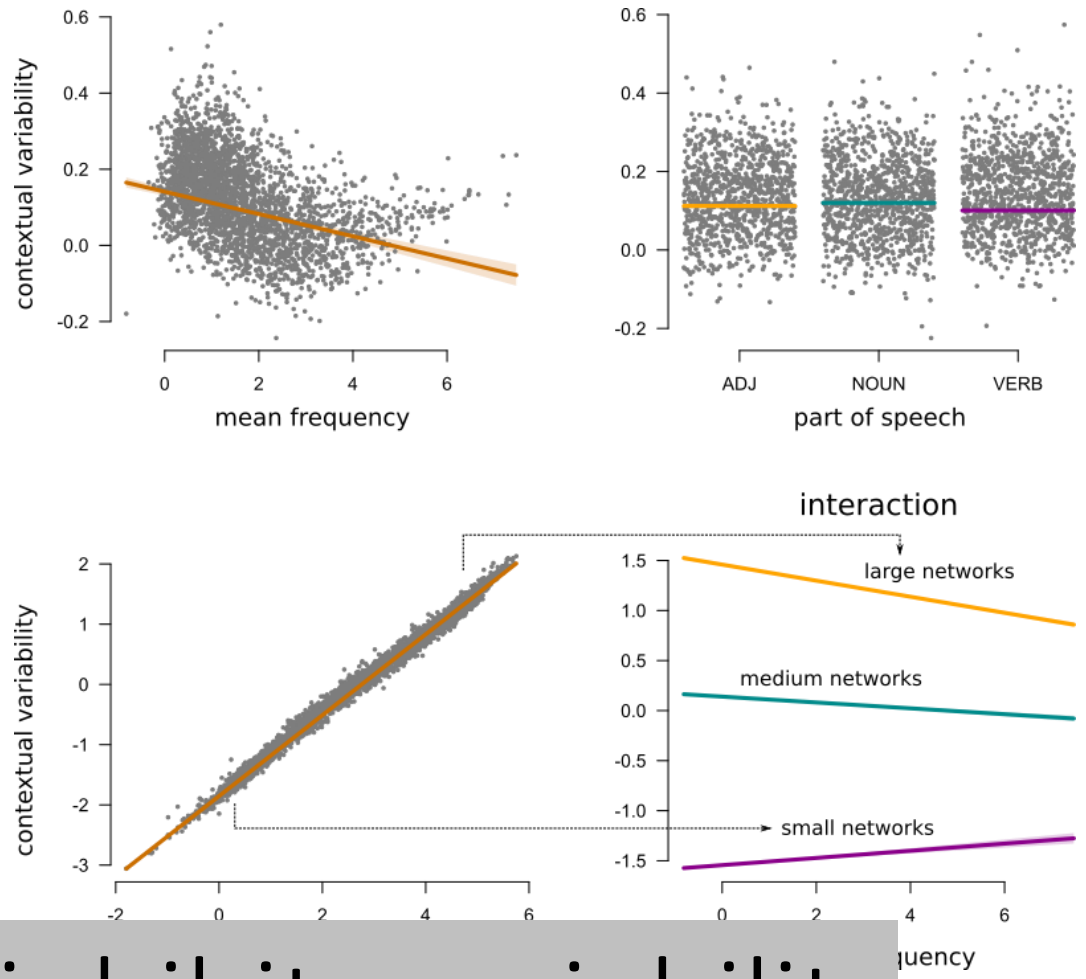
Study 1: results

- Linear model of variability
- controlling for network size and PoS
- **Result**: robust **negative effect** of frequency on semantic variability
- Even in short period of 20 yr!
- NB: effect reversed for words with small networks



Study 1: results

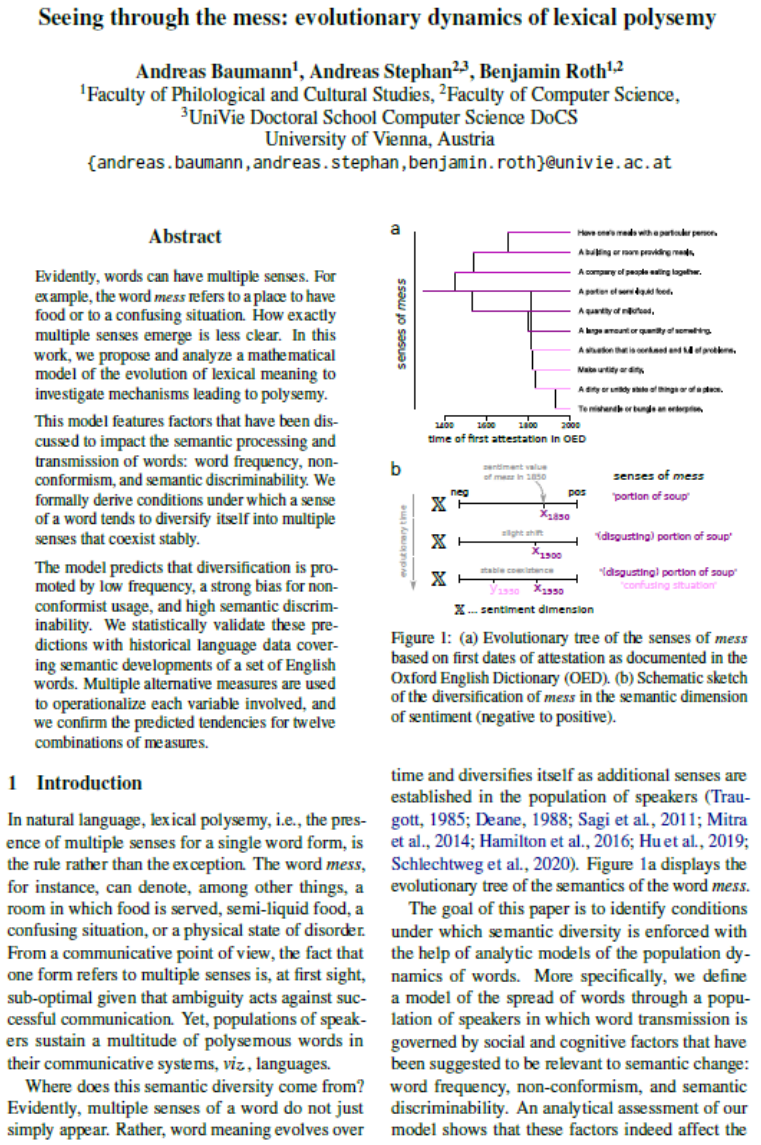
- Linear model of variability
- controlling for network size and PoS
- **Result**: robust **negative effect** of frequency on semantic variability
- Even in short period of 20 yr!
- NB: effect reversed for words with small networks



Résumé 1: frequency inhibits variability

Study 2: frequency and semantic diversification

- Words typically have **multiple senses**
- These senses have emerged over time
- E.g. the word *mess*



Seeing through the mess: evolutionary dynamics of lexical polysemy

Andreas Baumann[1], Andreas Stephan[2,3], Benjamin Roth[1,2]
[1]Faculty of Philological and Cultural Studies, [2]Faculty of Computer Science,
[3]UniVie Doctoral School Computer Science DoCS
University of Vienna, Austria
{andreas.baumann,andreas.stephan,benjamin.roth}@univie.ac.at

Abstract

Evidently, words can have multiple senses. For example, the word *mess* refers to a place to have food or to a confusing situation. How exactly multiple senses emerge is less clear. In this work, we propose and analyze a mathematical model of the evolution of lexical meaning to investigate mechanisms leading to polysemy.

This model features factors that have been discussed to impact the semantic processing and transmission of words: word frequency, non-conformism, and semantic discriminability. We formally derive conditions under which a sense of a word tends to diversify itself into multiple senses that coexist stably.

The model predicts that diversification is promoted by low frequency, a strong bias for non-conformist usage, and high semantic discriminability. We statistically validate these predictions with historical language data covering semantic developments of a set of English words. Multiple alternative measures are used to operationalize each variable involved, and we confirm the predicted tendencies for twelve combinations of measures.

Figure 1: (a) Evolutionary tree of the senses of *mess* based on first dates of attestation as documented in the Oxford English Dictionary (OED). (b) Schematic sketch of the diversification of *mess* in the semantic dimension of sentiment (negative to positive).

1 Introduction

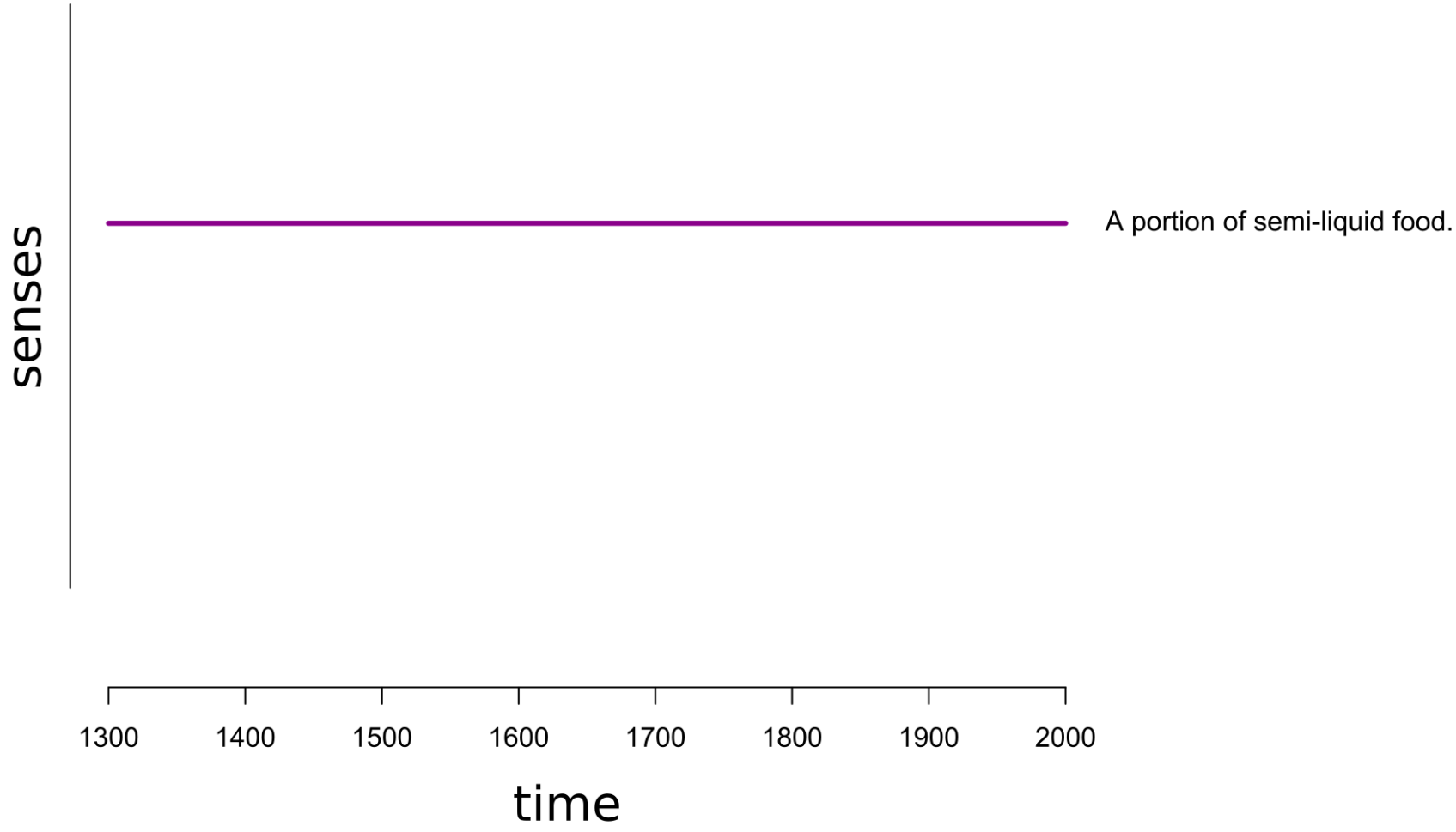
In natural language, lexical polysemy, i.e., the presence of multiple senses for a single word form, is the rule rather than the exception. The word *mess*, for instance, can denote, among other things, a room in which food is served, semi-liquid food, a confusing situation, or a physical state of disorder. From a communicative point of view, the fact that one form refers to multiple senses is, at first sight, sub-optimal given that ambiguity acts against successful communication. Yet, populations of speakers sustain a multitude of polysemous words in their communicative systems, *viz.*, languages.

Where does this semantic diversity come from? Evidently, multiple senses of a word do not just simply appear. Rather, word meaning evolves over time and diversifies itself as additional senses are established in the population of speakers (Traugott, 1985; Deane, 1988; Sagi et al., 2011; Mitra et al., 2014; Hamilton et al., 2016; Hu et al., 2019; Schlechtweg et al., 2020). Figure 1a displays the evolutionary tree of the semantics of the word *mess*.

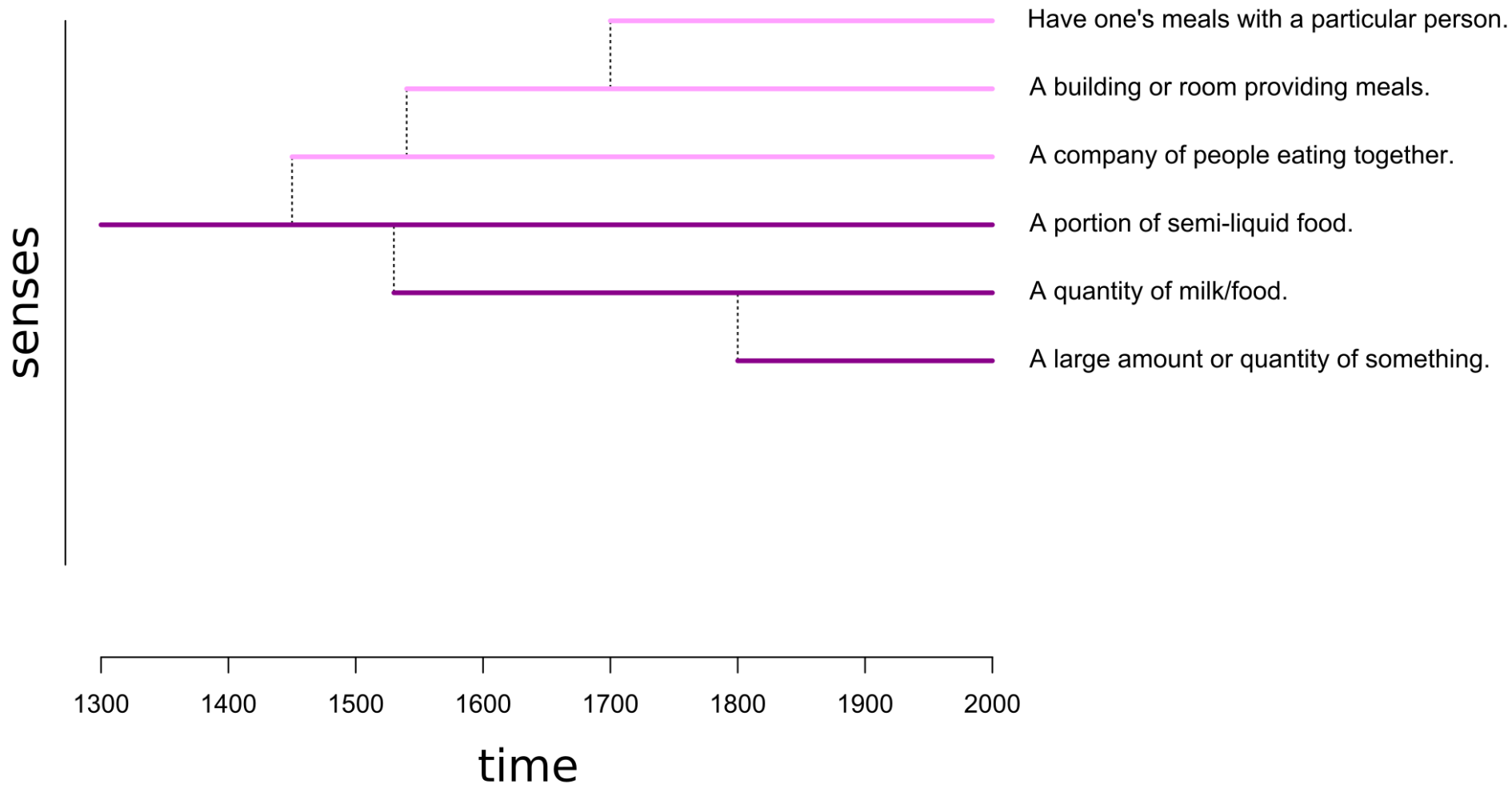
The goal of this paper is to identify conditions under which semantic diversity is enforced with the help of analytic models of the population dynamics of words. More specifically, we define a model of the spread of words through a population of speakers in which word transmission is governed by social and cognitive factors that have been suggested to be relevant to semantic change: word frequency, non-conformism, and semantic discriminability. An analytical assessment of our model shows that these factors indeed affect the

Baumann, Stephan, Roth, 2023, *EMNLP*.

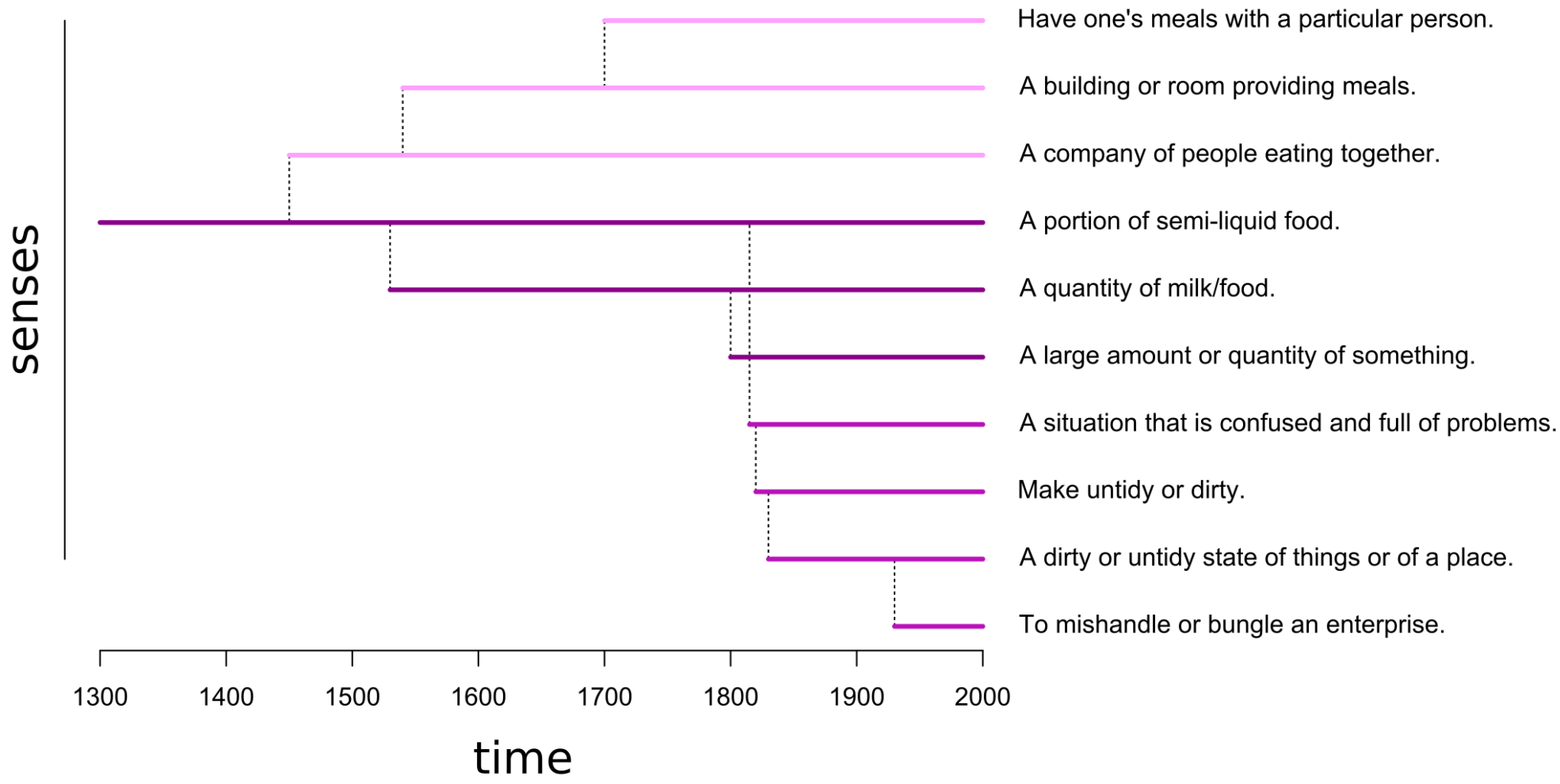
Study 2: semantic evolution of *mess*



Study 2: semantic evolution of *mess*



Study 2: semantic evolution of *mess*



Study 2: data

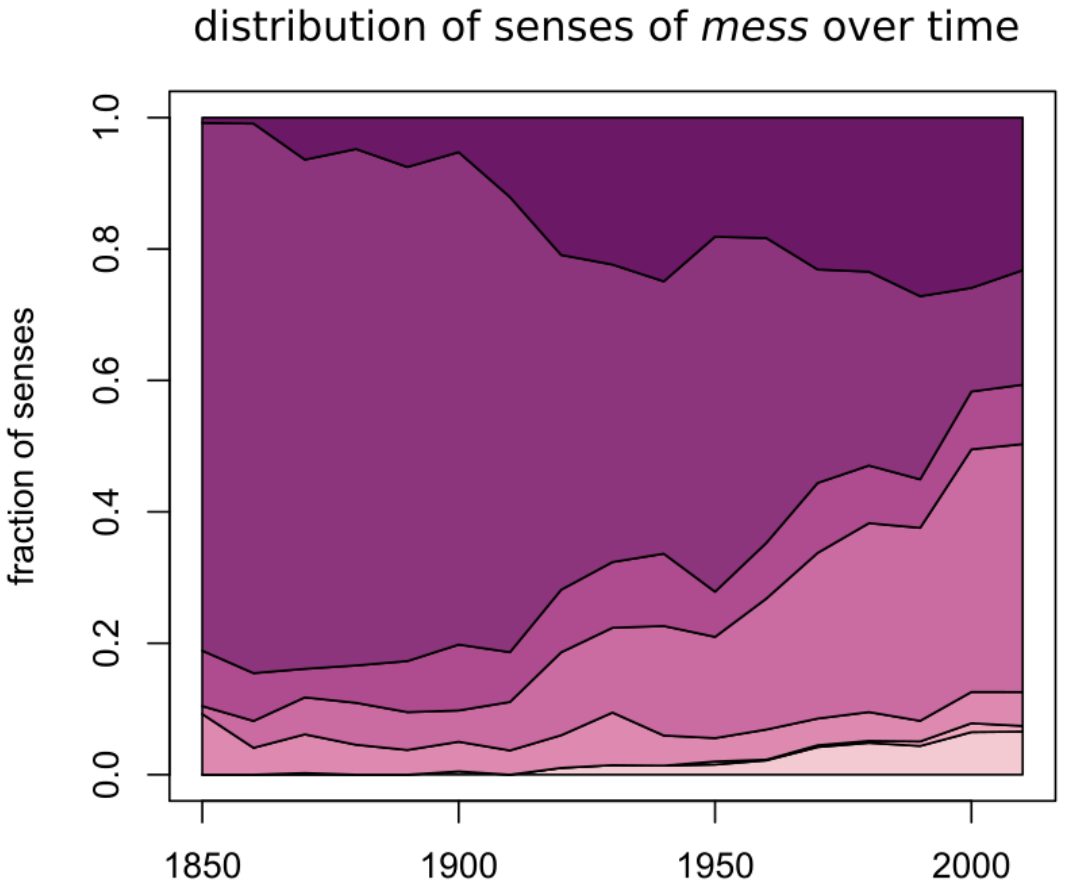
- Combination of **several data sets** (eng)
- **Sense-distributions** for ca. 3500 words based on Oxford English Dictionary and COHA (19th-20th century; Hu et al. 2019)
- Diachronically layered pre-trained word **embeddings** (HistWords)
- **Frequency** trajectories (COHA)

Study 2: tendency to diversify meaning

- To what extent did a word get more senses over time?
- Three options to measure polysemy:

A. Number of senses per decade (derived from Hu et al. 2019)

B. Diversity of sense distribution per decade (same data)

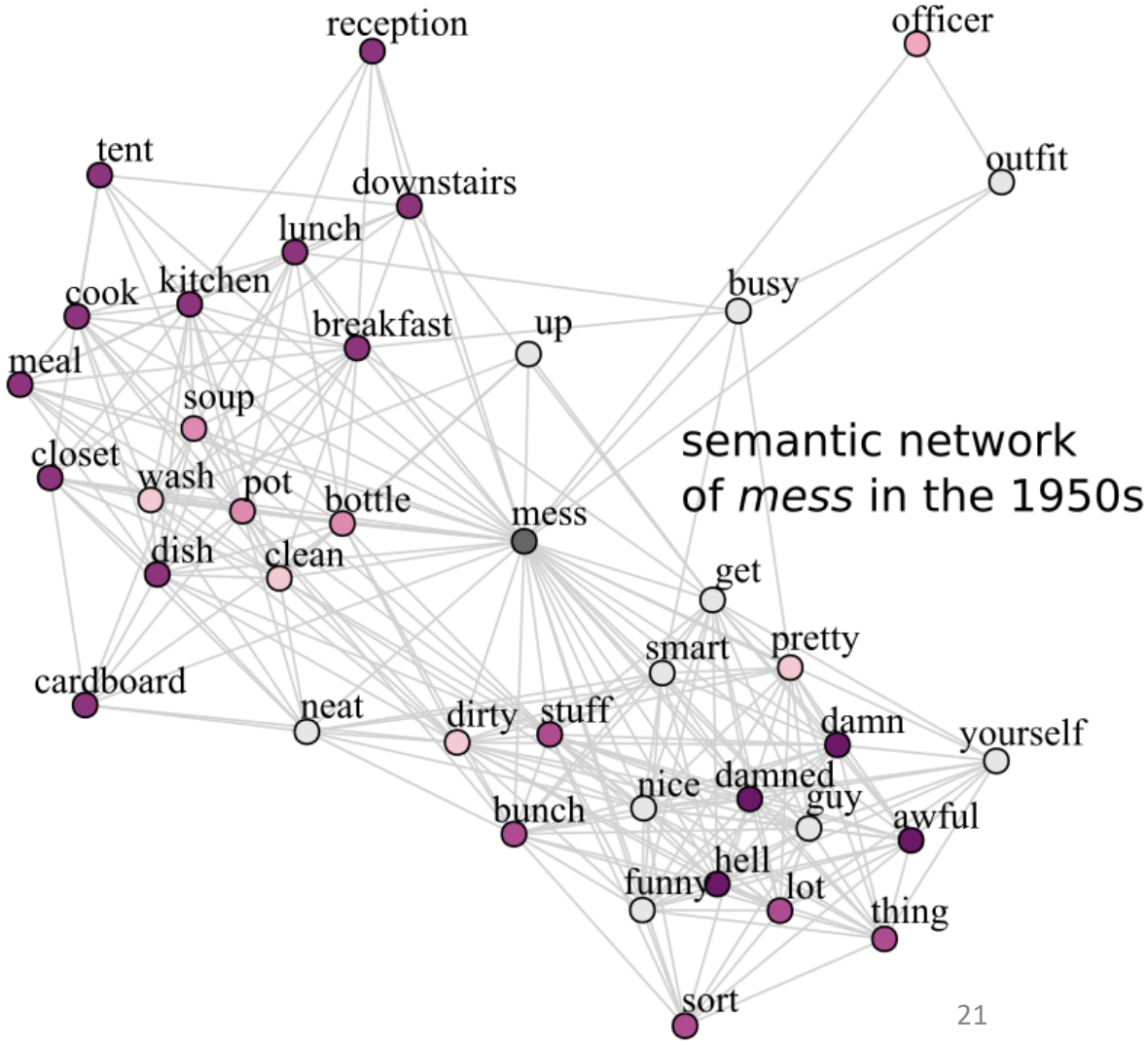


- A situation that is confused and full of problems.
- A building or room providing meals and recreational facilities for members of the armed forces.
- A large amount or quantity of.
- A dirty or untidy state of things or of a place.
- A portion of semi-liquid food.
- Have one's meals with a particular person, especially as a member of an armed forces' mess.
- Make untidy or dirty.

Study 2: tendency to diversify meaning

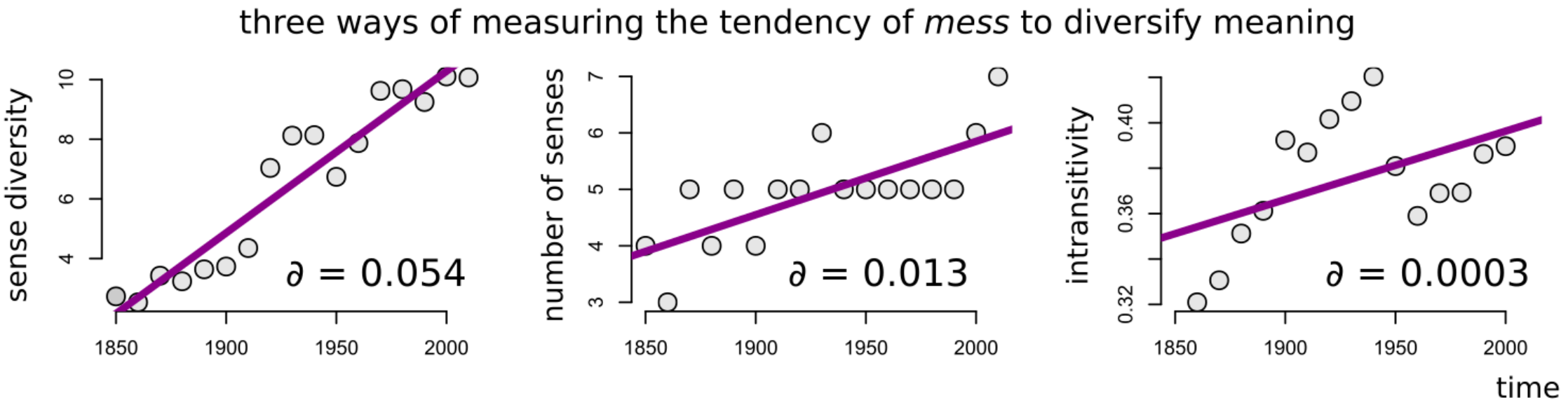
C. Intransitivity of lexical network per decade (Hamilton et al. 2016)

- Word embeddings for each word
- Construct network of semantic neighborhood
- Compute transitivity/clustering coefficient C
- $1 - C \leftrightarrow$ polysemy



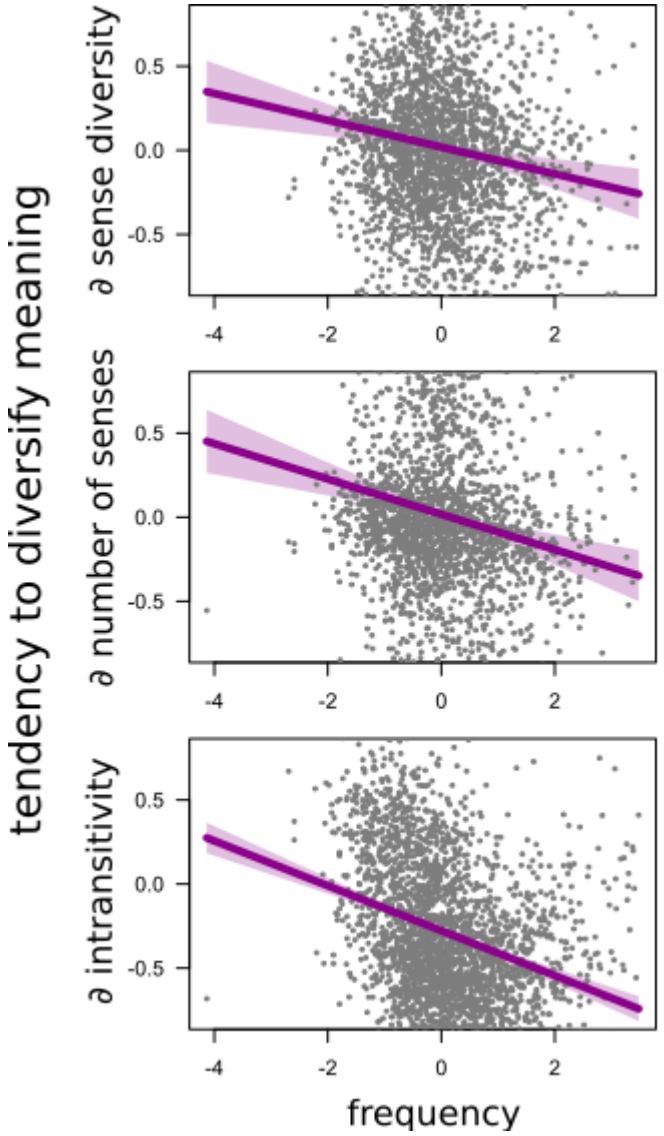
Study 2: tendency to diversify meaning

- For a single word: do this for each decade
- Take **slope** ∂ of each development as **tendency** to diversify meaning



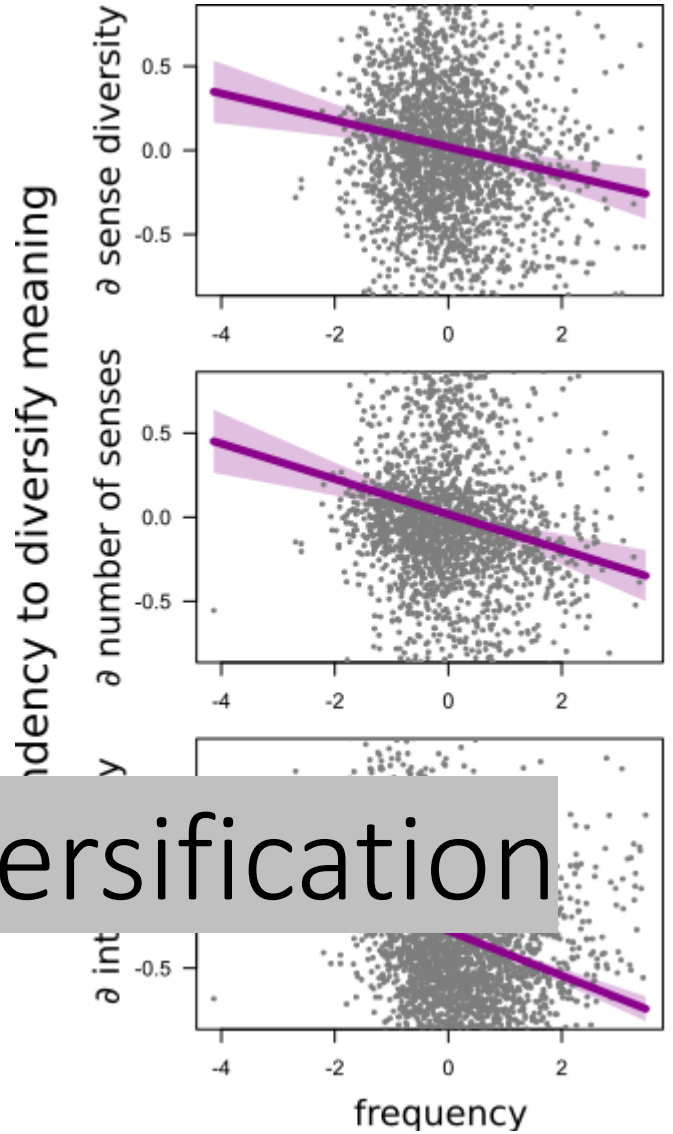
Study 2: effect of frequency on this tendency

- **Frequency**: measured via historical frequency trajectories
- Robust **negative effects** of (log) frequency on tendency of diversification ∂



Study 2: effect of frequency on this tendency

- **Frequency**: measured via historical frequency trajectories
- Robust **negative effects** of (log) frequency on tendency of diversification ∂



Résumé 2a: frequency inhibits diversification

…but why?

Study 2: pop-dyn modeling

- Set up **population dynamic model** (ODE) of the usage of a single variant depending on a semantic property x in a speaker population
  - e.g., *mess* = ‘portion of food’
  - semantic property x: ‘valence’

Shibuya Crossing (Tokyo, 09/2022)

Study 2: pop-dyn modeling

- Set up **population dynamic model** (ODE) of the usage of a single variant depending on a semantic property x in a speaker population
  - e.g., *mess* = ‘portion of food’
  - semantic property x: ‘valence’
- Allow for emergence of competing semantic **variants** x and y
  - e.g. *mess* = ‘disgusting portion of food’ with lower valence
- Analyze predictions about the resulting **long-term development** of the semantic property
  - e.g., development of average ‘valence’ of the word *mess*

Shibuya Crossing (Tokyo, 09/2022)

change in the number of users

frequency

adoption depending on x

positive frequency dependence

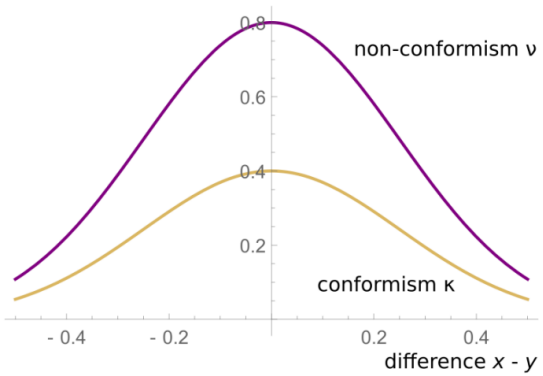
negative frequency dependence

$$\dot{U} = \varphi\,\alpha(x)UN + \kappa(\Delta = 0)UU - \nu(\Delta = 0)UU$$

conformism bias

non-conformism bias

difference between two variants



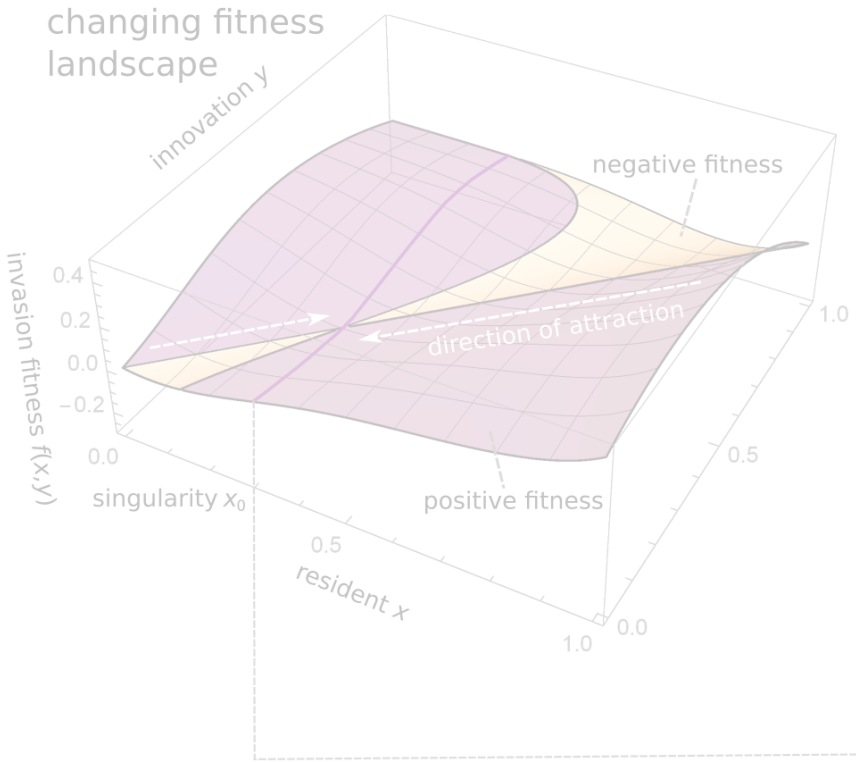
$$\kappa(\Delta) = \kappa_0 \exp(-1/2 \cdot \Delta^2\delta^2)$$

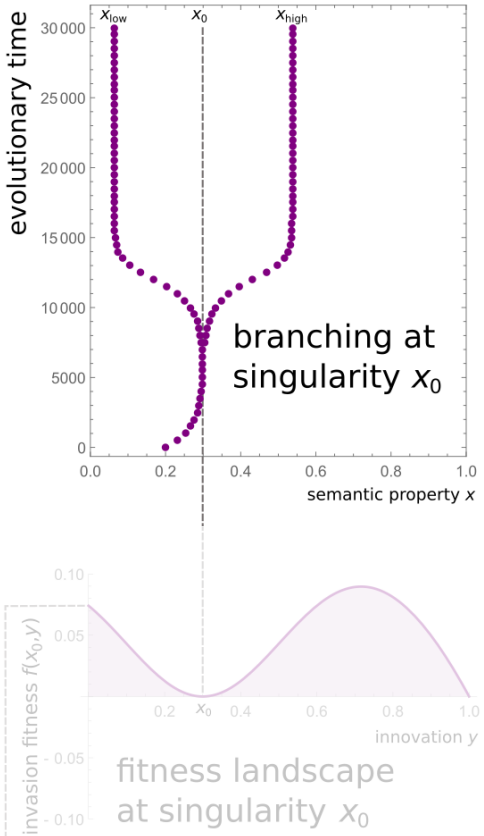
$$\nu(\Delta) = \nu_0 \exp(-1/2 \cdot \Delta^2\delta^2)$$

discriminability

Study 2: evolutionary invasion analysis

- Result*:* **frequency impedes** stable coexistence of multiple senses
- Reason:
  - **low** frequency yields **branching points**
  - **high** frequency yields **stable states** (evolutionary dead ends) in the meaning space that optimize transmission
  - only if there is a sufficiently high tendency to behave in a **non-conformist** way

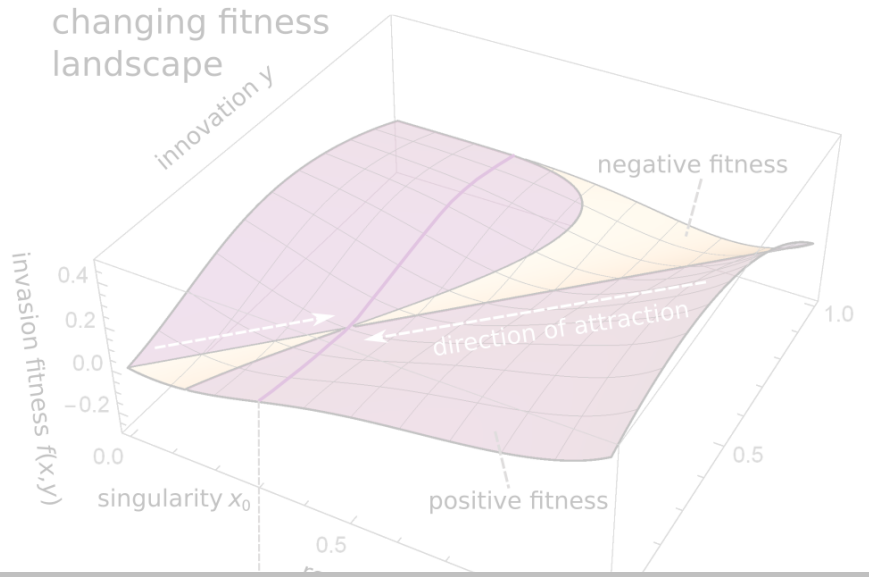


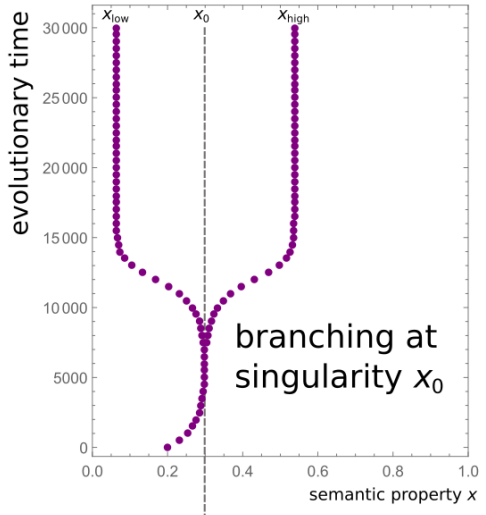


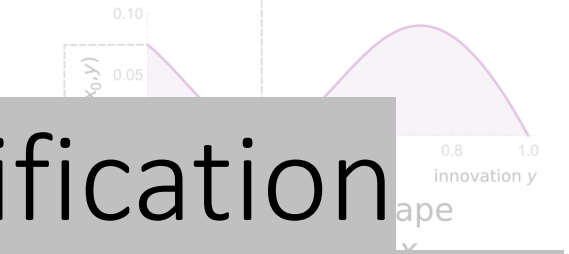
Geritz et al. (1996), Doebeli (2011)

Study 2: evolutionary invasion analysis

- Result*:* **frequency impedes** stable coexistence of multiple senses
- Reason:
  - **low** frequency yields **branching points**
  - **high** frequency yields **stable states** (evolutionary dead ends) in the meaning space that optimize transmission
  - only if there is a sufficiently high tendency to behave in a **non-**





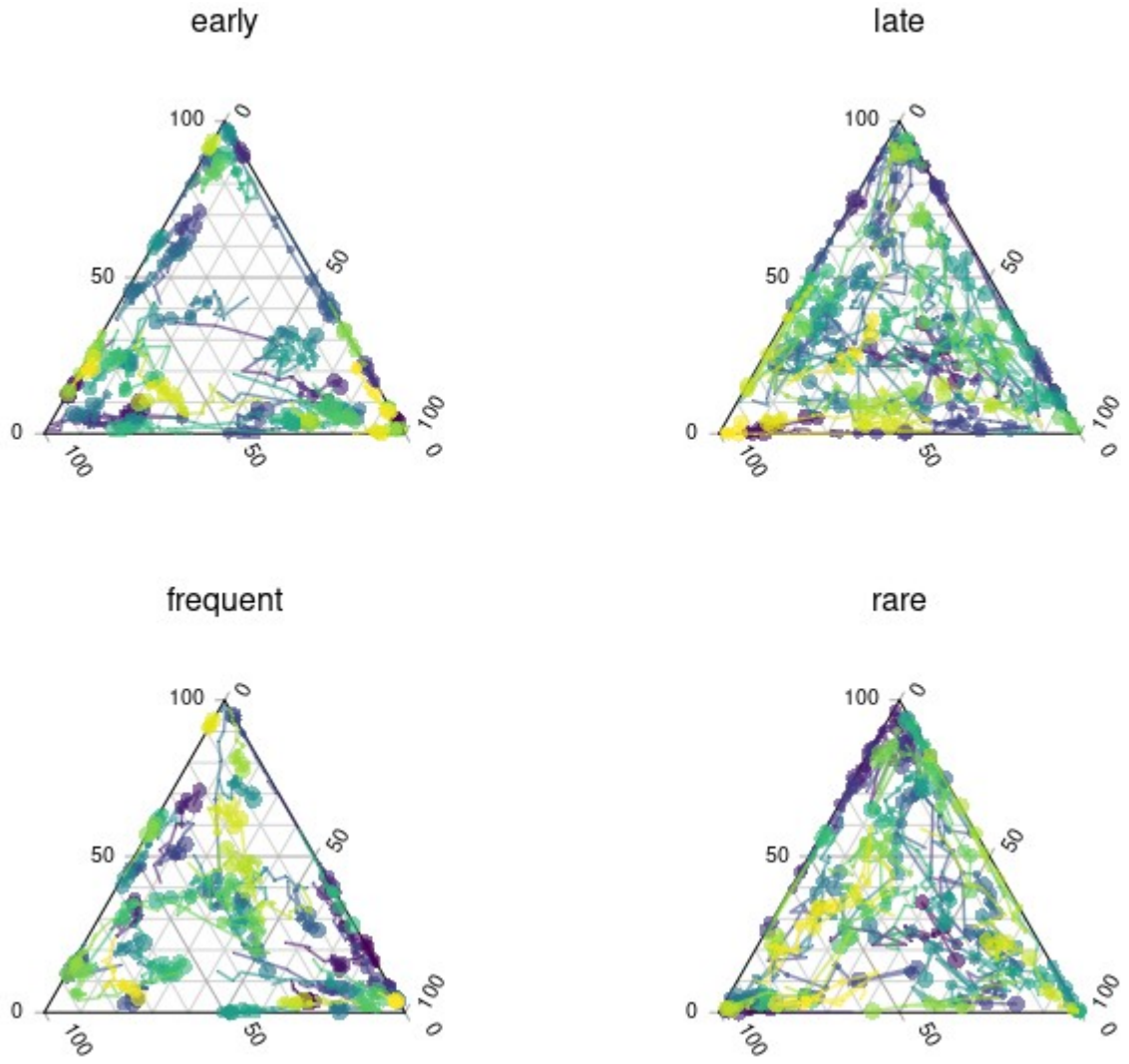


Résumé 2b: frequency inhibits diversification …because frequent words are transmitted easily

Geritz et al. (1996), Doebeli (2011)

Study 3: frequency and acquisition

- Lexical transmission also depends on acquisition
- How do **frequency and acquisition interact** in semantic dynamics?
- Is the effect of frequency modulated by ease of acquisition?
- Psycholinguistic data: subjective **age of acquisition** (AoA) ratings (Kuperman et al. 2012)

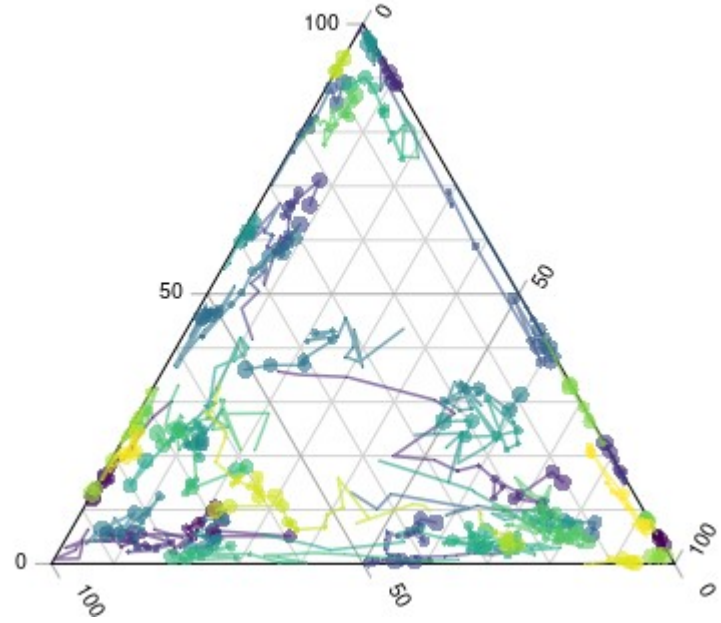


Baumann, Scheicher, Böhm & Hartmann, 2024, *Poster at CogSci*.

Study 3: frequency and acquisition

- Three ways of measuring semantic dynamics

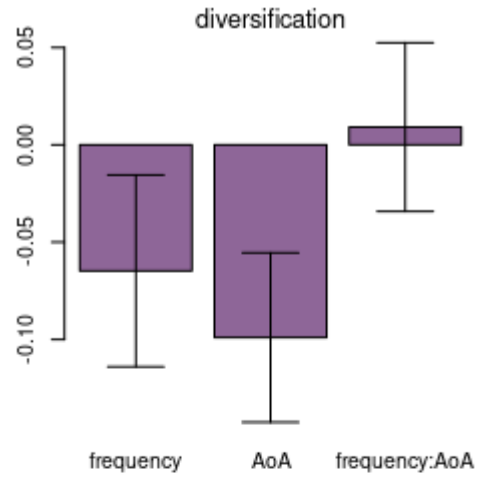
1. **Diversification**
   - Do words become more polysemous? (cf. Study 2, Baumann et al. 2023, *EMNLP*)
2. **Displacement**
   - Do words shift far in the semantic space? (cf. Baumann et al. 2023, *Cognitive Linguistics*)
3. **Variability**
   - Do word meanings fluctuate from period to period? (cf. Hamilton et al. 2016, *ACL*; Cassani et al. 2021, *Cognitive Science*)

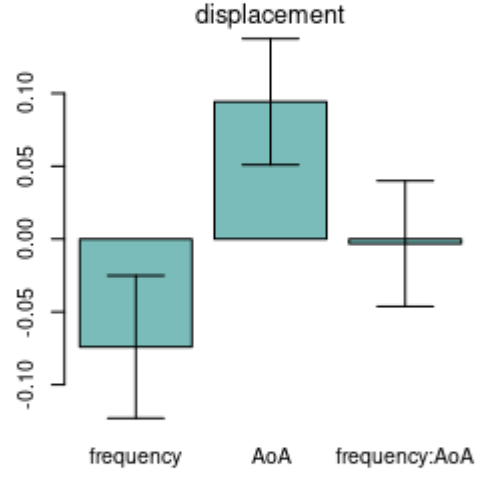


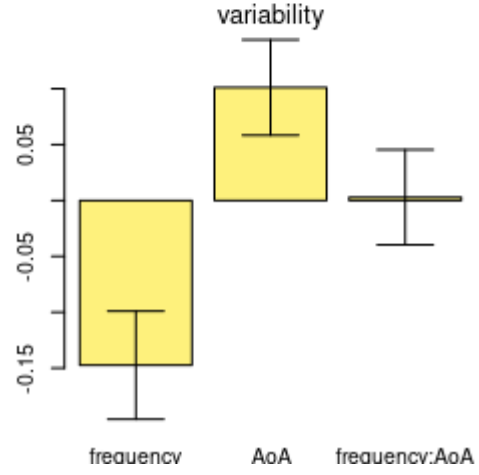
Do trajectories approach the center, are they long, and are they wiggly?

Study 3: frequency and acquisition

- Preliminary findings (cf. Baumann et al. 2024, *CogSci*)
- **Frequency** uniformly **demotes/stabilizes** semantic dynamics, independent from acquisition
- Age of acquisition is more complex:
  - **Early acquired** words show **less variability/fluctuation** and **less displacement**
  - But: **early acquired** words are more likely to **diversify** than late acquired words
  - NB: effect remains even if controlling for concreteness



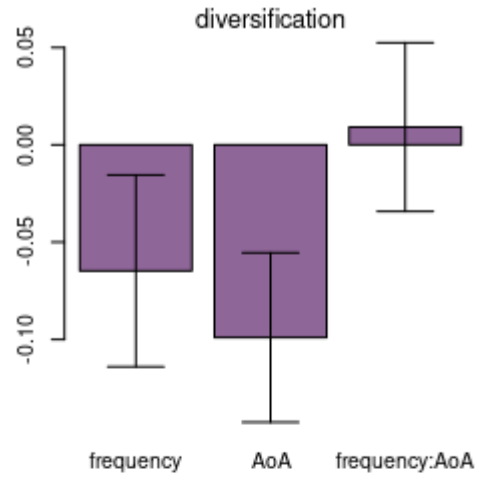


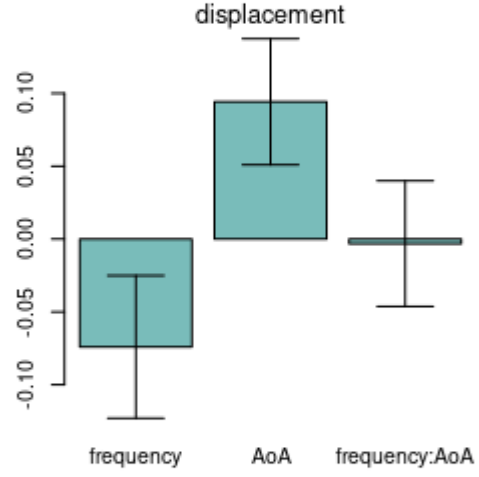


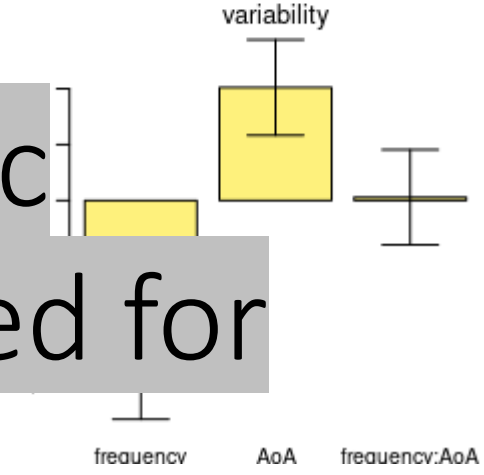
universität wien

Study 3: frequency and acquisition

- Preliminary findings (cf. Baumann et al. 2024, *CogSci*)
- **Frequency** uniformly **demotes/stabilizes** semantic dynamics, independent from acquisition
- Age of acquisition is more complex:
  - **Early acquired** words show **less variability/fluctuation** and **less displacement**
  - But: **early acquired** words are more likely to **diversify** than late







Résumé 3: frequency demotes semantic dynamics even if acquisition is controlled for

Résumé

- **Frequency inhibits** semantic **variability, displacement**, and **diversification**
- Still somewhat puzzling:
  - synchronically, frequent words have more senses (cf. Zipf), but diachronically they *don't* tend to become more polysemous
  - theoretical model predicts that this is because frequent words are optimized for easy transmission, but early acquired words are *more* likely get more senses
- **Digitized data & computational methods** for quantifying meaning great for testing such hypotheses theoretically & empirically based on many words

universität wien

Thank you!