Medienlinguistische Methodik
Einführung

Arne Rubehn
Lehrstuhl für Multilinguale Computerlinguistik
Universität Passau

14.10.2025

Wer bin ich?

Arne Rubehn

Wer bin ich?

~~Dr.~~ Arne Rubehn

Wer bin ich?

~~Prof. Dr.~~ Arne Rubehn

Wer bin ich?

~~Prof. Dr.~~ Arne Rubehn

gerne per du

Wer bin ich?

~~Prof. Dr.~~ **Arne Rubehn**

gerne per du

seit 2023: wiss. Mitarbeiter & Doktorand, Lehrstuhl für Multilinguale Computerlinguistik, Universität Passau

2019-2023: M.A. Computerlinguistik, Universität Tübingen

2015-2019: B.A. Allgemeine Sprachwissenschaft & Latein, Universität Tübingen

Wer seid ihr?

https://partici.fi/92489810

Semesterausblick

14.14.	Einführung
21.10.	Modellierung
28.10.	Datenerhebung: Beobachtungen & Befragung
04.11.	Datenerhebung: Experimente
11.11.	Korpuslinguistik
18.11.	Annotationen
25.11.	Textanalyse
02.12.	Diskursanalyse
09.12.	Daten aus Social Media
16.12.	Sprachmodelle
13.01.	Qualitative Inhaltsanalyse
20.01.	Offene Forschung
27.01.	Zusammenfassung & Klausurvorbereitung
03.02.	**Klausur (120 min. – MGP mit SE Werbesprache)**

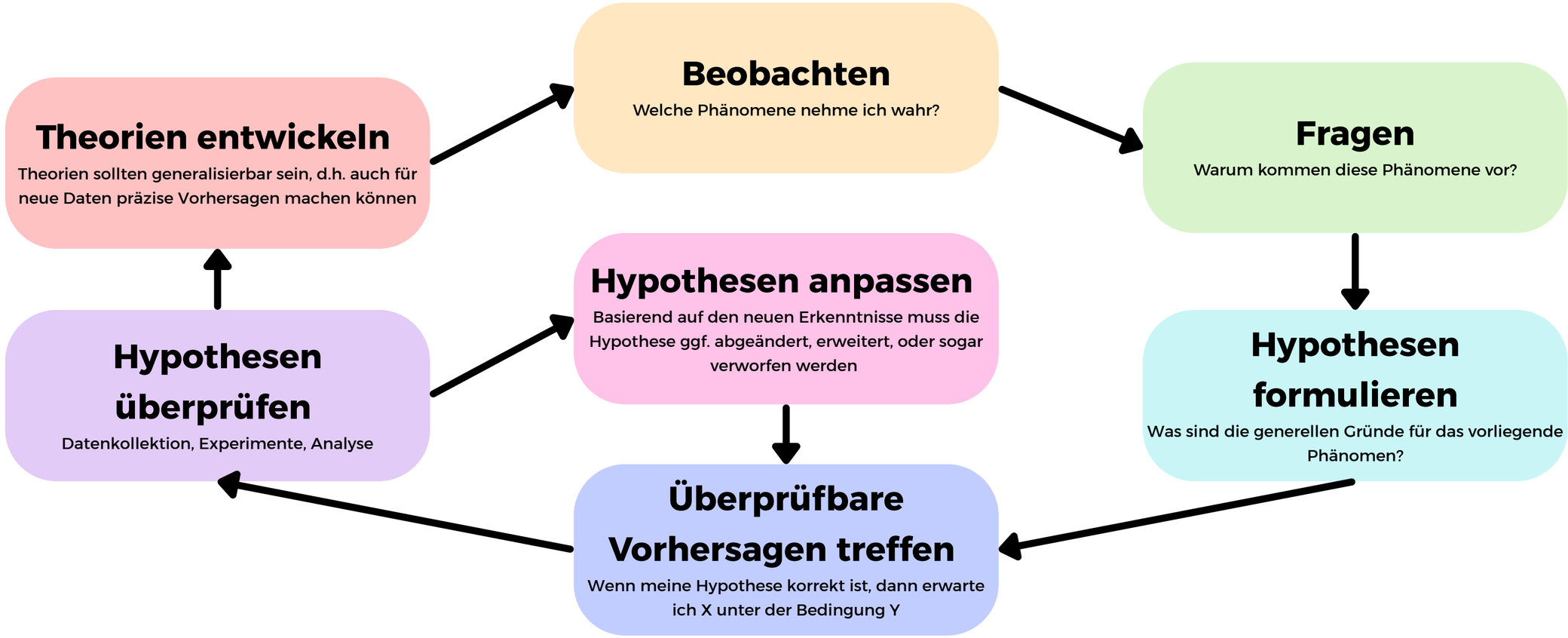
Raum: HK 28, SR 002
Uhrzeit: 14:00-16:00 (c.t.)
Leistungsnachweis:
Klausur (Modul)
3 Studienleistungen (unbenotet)
Credits: 10 CP (Modul)

Es besteht **keine Anwesenheitspflicht**.

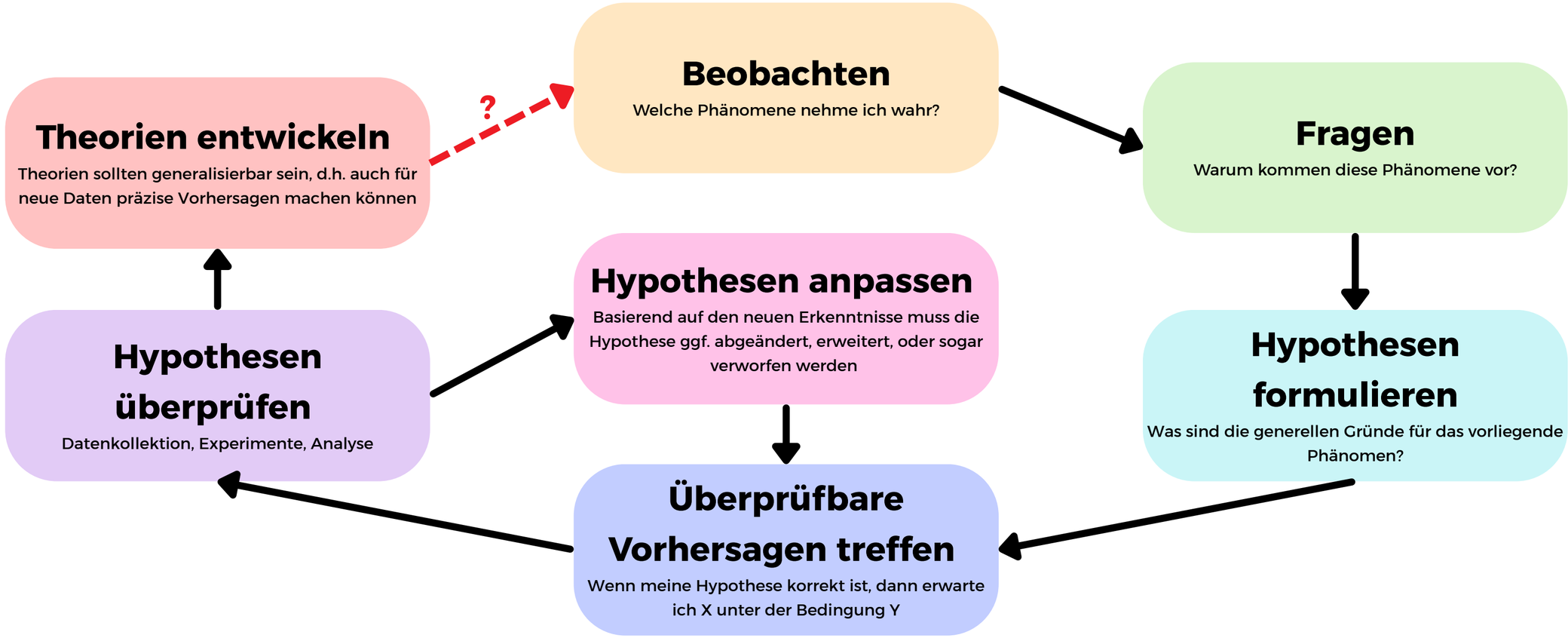
Es wird **keine zusätzliche Lektüre** erwartet.

mental health matters.

Die wissenschaftliche Methode

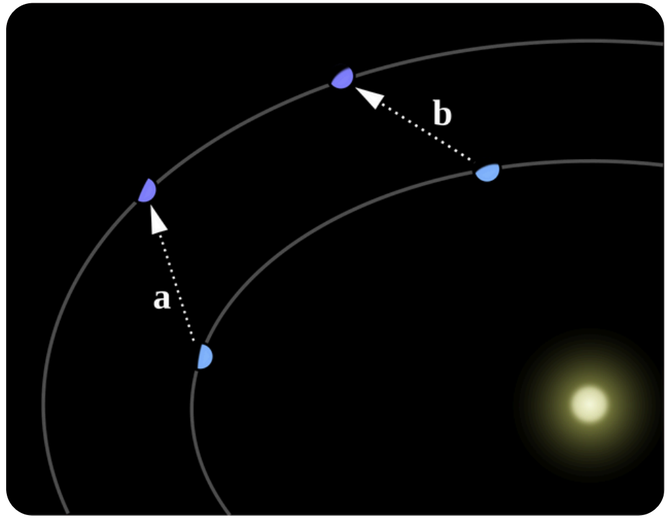


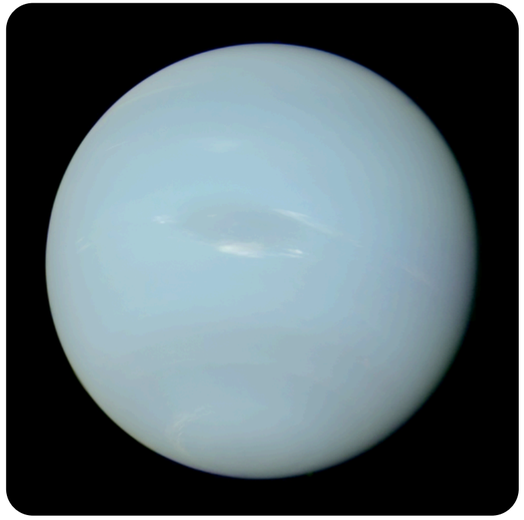
Die wissenschaftliche Methode

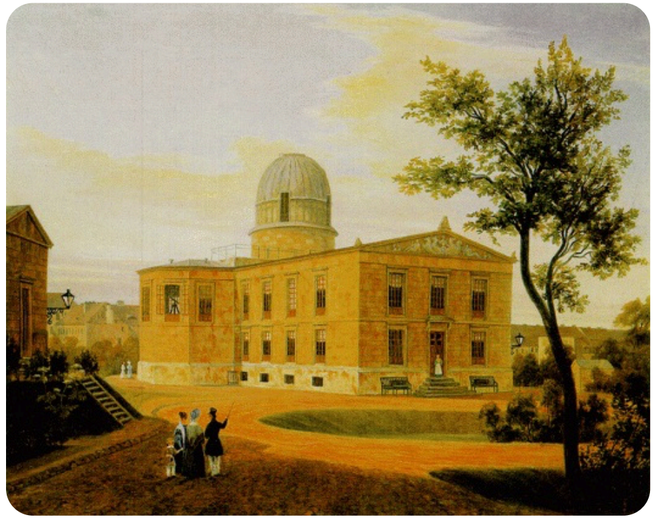


Die wissenschaftliche Methode

Die Existenz und Position des Planets Neptun konnte durch Newtons Gravitationstheorie **vorausgesagt** werden!







Die wissenschaftliche Methode

In der wissenschaftlichen Methode kann man vier grundlegende Elemente unterscheiden:

Charakterisierungen. Beobachtungen, Definitionen, Messungen.

Hypothesen. Theoretische Erklärungen für die beobachteten Phänomene.

Vorhersagen. Testbare Prozesse, die sich aus der Hypothese ableiten.

Experimente. Test der o.g. Elemente.

Die wissenschaftliche Methode

Die wissenschaftliche Methode stellt keine genaue Anleitung dar, sondern beschreibt eher die generellen Abläufe von guter wissenschaftlicher Arbeit.

Der eben vorgestellte Workflow beschreibt naturwissenschaftliche, bzw. **experimentelle Forschung** wohl besser als sozial- und geisteswissenschaftliche.

Sie fußt auf den drei wichtigen Grundsätzen der **Ehrlichkeit, Offenheit** und **Falsifizierbarkeit.**

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Published research findings are sometimes refuted by subsequent evidence, with ensuing confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims [6–8]. However, this should not be surprising. It can be proven that most claimed research findings are false. Here I will examine the key factors that influence this problem and some corollaries thereof.

The Essay section contains opinion pieces on topics of broad interest to a general medical audience.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful. "Negative" is actually a misnomer, and the misinterpretation is widespread. However, here we will target relationships that investigators claim exist, rather than null findings.

As has been shown previously, the probability that a research finding is indeed true depends on the prior probability of it being true (before doing the study), the statistical power of the study, and the level of statistical significance [10,11]. Consider a 2 × 2 table in which research findings are compared against the gold standard of true relationships in a scientific field. In a research field both true and false hypotheses can be made about the presence of relationships. Let R be the ratio of the number of "true relationships" to "no relationships" among those tested in the field. R is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R + 1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2 × 2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2 × 2 table, one gets $PPV = (1 - \beta)R/(R - \beta R + \alpha)$. A research finding is thus

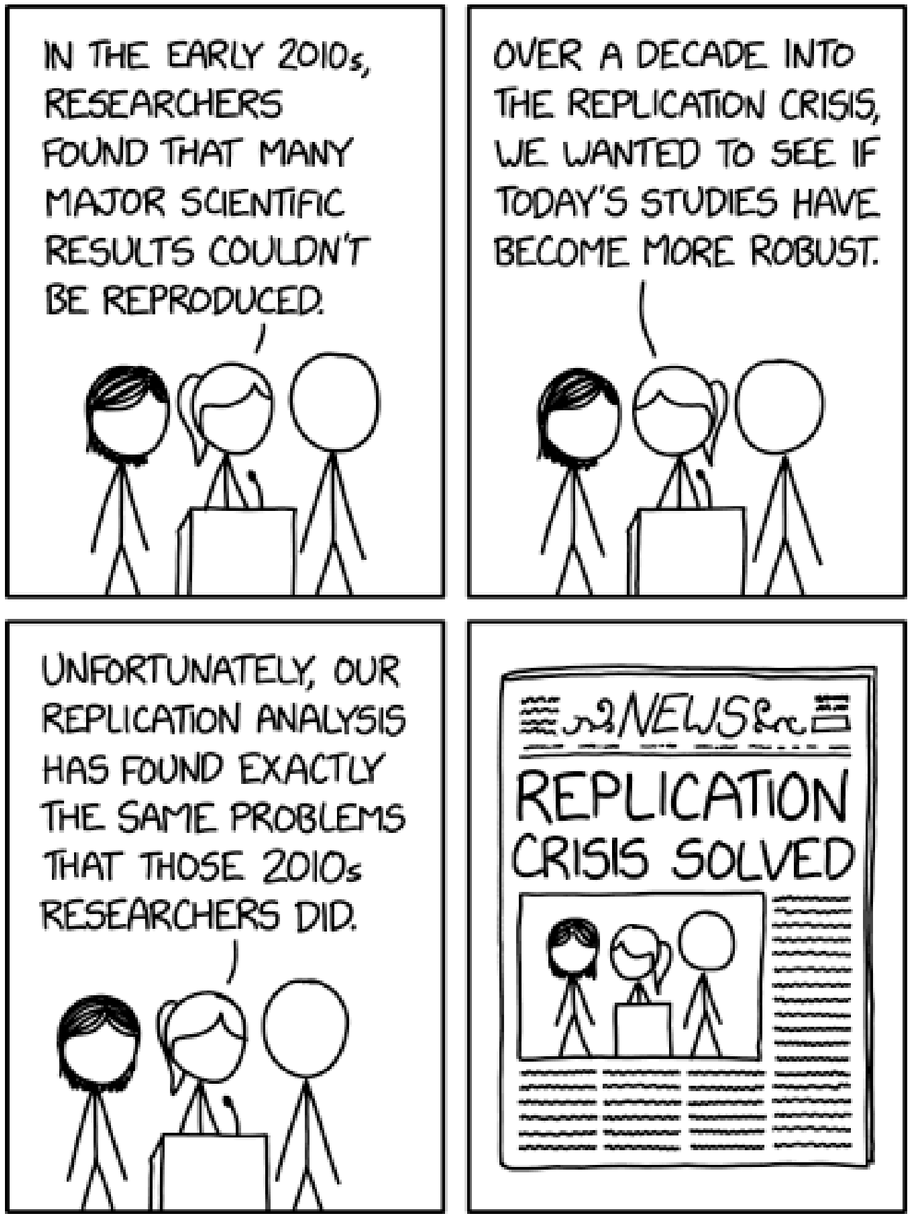
Citation: Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8): e124.

Abbreviation: PPV, positive predictive value

John P. A. Ioannidis is in the Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, and Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States of America. E-mail: jioannid@cc.uoi.gr

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020124



Semesterausblick

Das Seminar wird unter dem Hauptthema **Korpuslinguistik** stehen.

Es wird also darum gehen, wie wir den **Sprachgebrauch** mithilfe von großen Textsammlungen untersuchen können.

Des Weiteren sollen **wissenschaftliche Prozesse** im Allgemeinen diskutiert werden.

Welche Gedanken habt ihr zu diesen Themen?

Semesterausblick

Modellierung. Was können wir uns unter Daten und Modellen in der Wissenschaft vorstellen?

Datenerhebung. Wie können wir Forschungsdaten durch Beobachtungen, Befragungen oder Experimente gewinnen?

Korpuslinguistik. Was sind Korpora und wie können wir mit ihnen arbeiten?

Annotationen. Wie und zu welchem Zweck können Daten annotiert werden?

Semesterausblick

Textanalyse. Welche Techniken gibt es zur Untersuchung von Textdaten?

Diskursanalyse. Wie können Diskurse wissenschaftlich untersucht werden?

Daten aus Social Media. Wie kann ich Beiträge aus Social Media als Forschungsdaten nutzen? Was ist dabei zu beachten?

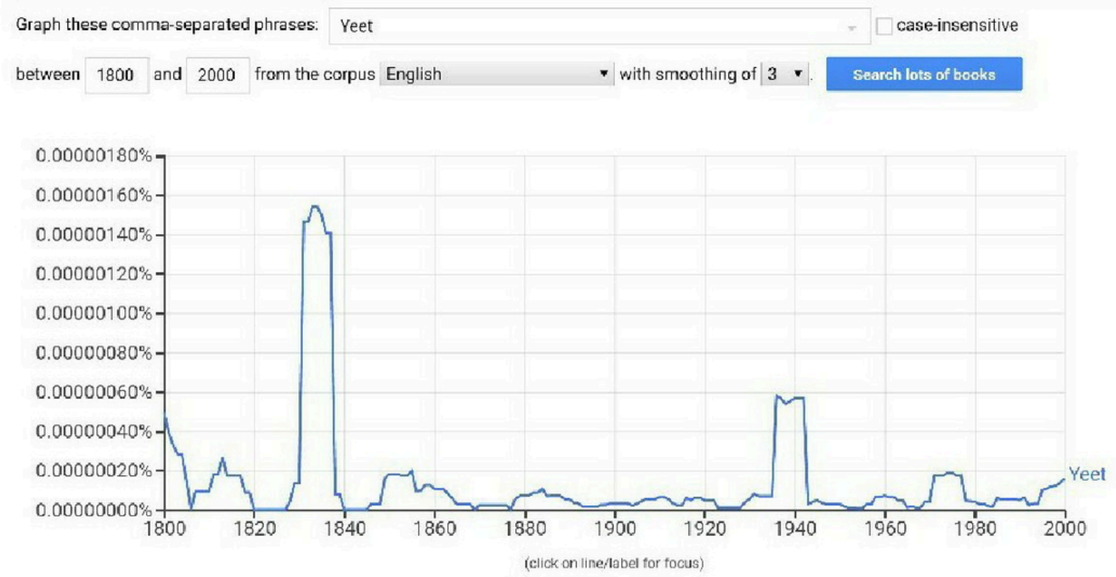
Sprachmodelle. Wie funktionieren Modelle wie (Chat-)GPT? Welche Chancen und Risiken gibt es?

Semesterausblick

Qualitative Inhaltsanalyse. Wie funktioniert sie und welche Fragen kann sie beantworten?

Offene Forschung. Wie stellen wir sicher, dass Analysen reproduzierbar und Daten wiederverwendbar sind?

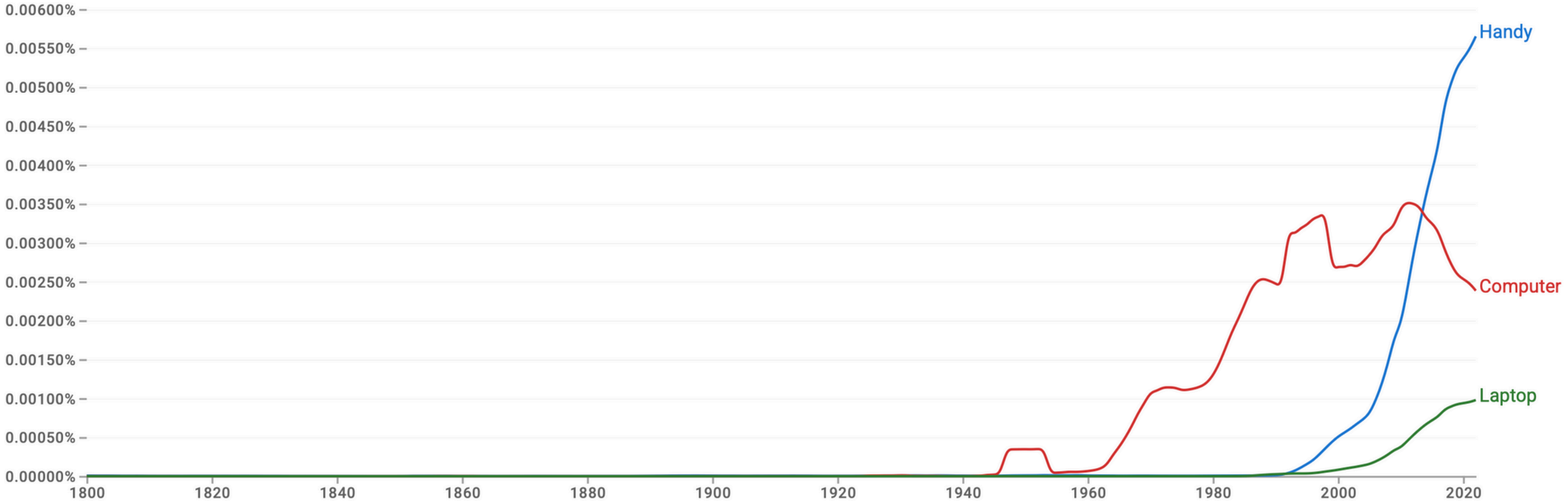
Gibt es Themen oder Aspekte, die euch besonders interessieren?



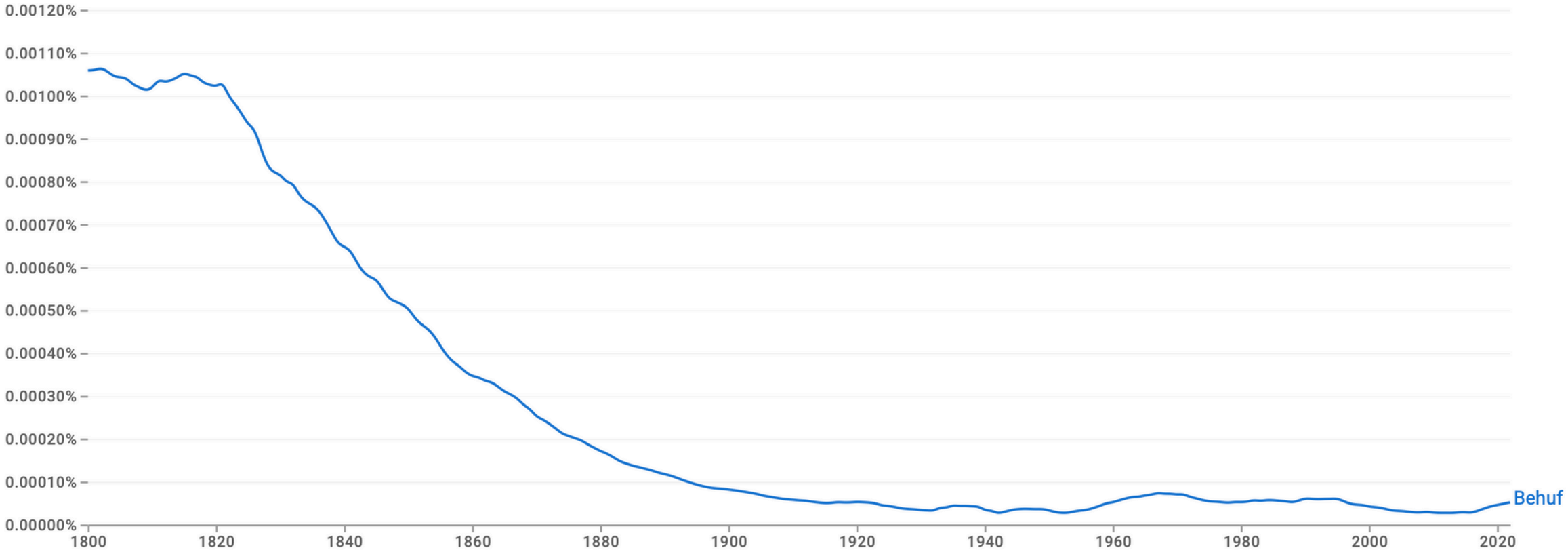
1836:

https://www.reddit.com/r/meme/comments/d462cf/summary_of_the_1836_yeet/

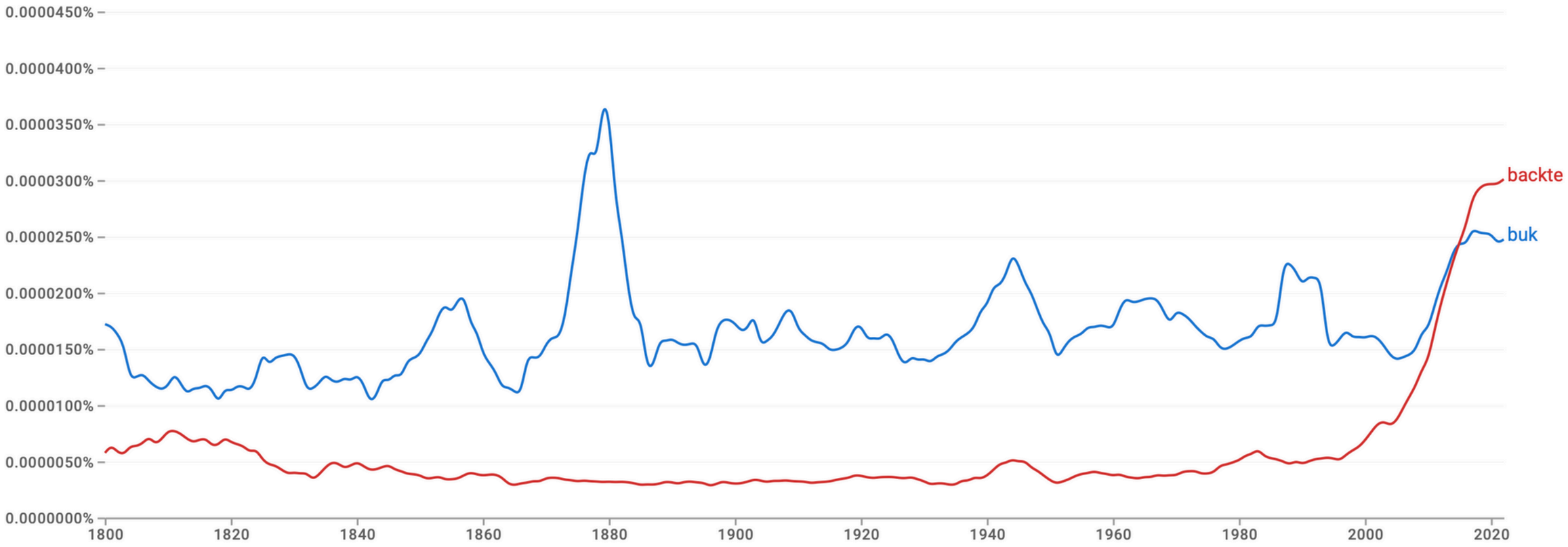
Google Books Ngram Viewer



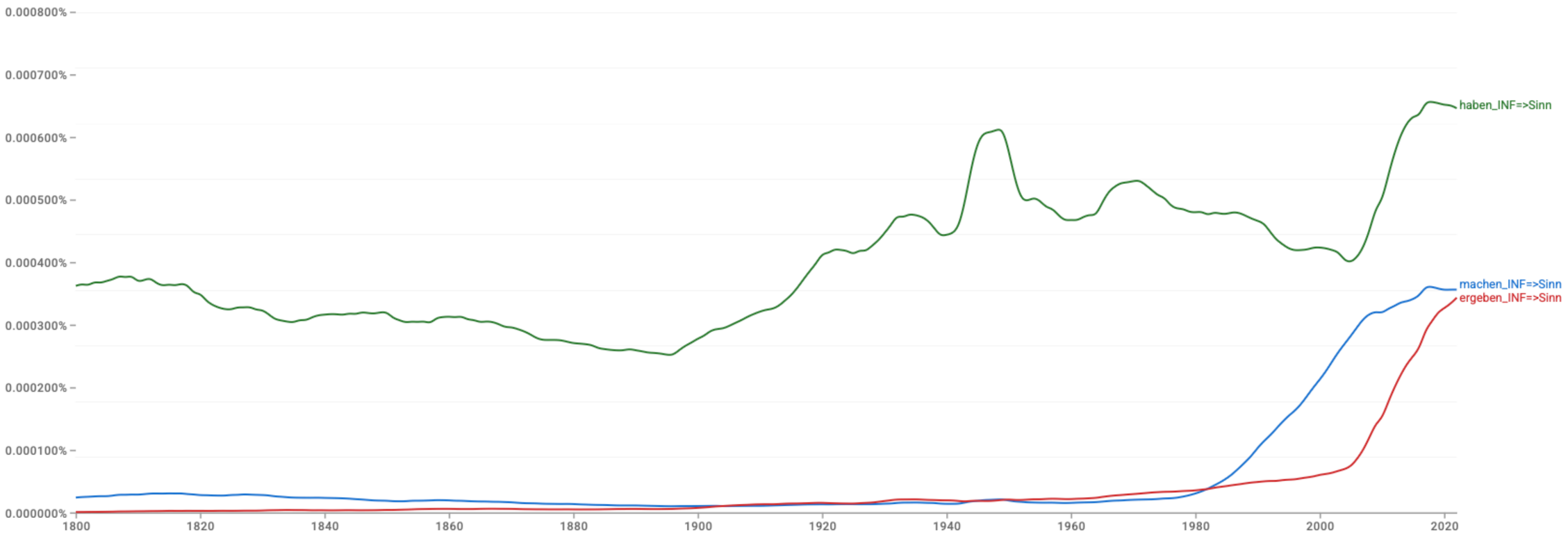
Google Books Ngram Viewer



Google Books Ngram Viewer



Google Books Ngram Viewer



Google Books Ngram Viewer

Probiert es selbst aus!

https://books.google.com/ngrams/

Google Books Ngram Viewer

Probiert es selbst aus!

https://books.google.com/ngrams/

Welche Fragen könnte man mithilfe eines solchen Korpus beantworten?

Welche Informationen werden benötigt, um einen solchen Korpus zusammenzustellen?