

# Digitalgestützte Textanalyse in Forschung und Lehre

(Digitale) Chancen für den Lateinunterricht

Freiburg, 28.9. – 29.9.2023

#Chancen verknüpfen: Digital Humanities und Lateinunterricht

**Dr. Andrea Beyer,  
Humboldt-Universität zu Berlin**



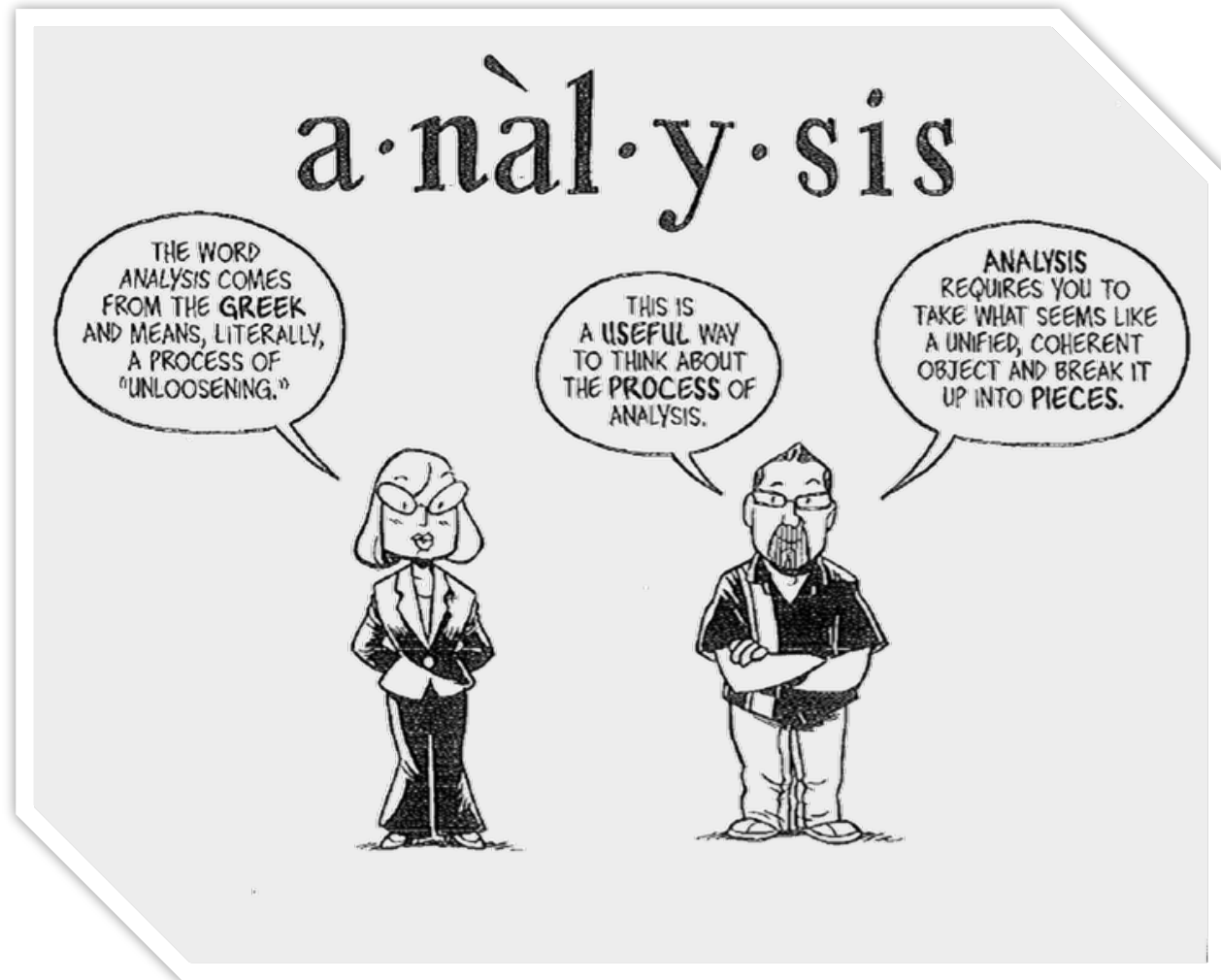
— Digital Research for All —  
**DAIDALOS**

Gefördert durch  
**DFG**

Deutsche  
Forschungsgemeinschaft

# VIER Schwerpunkte und EIN Spezialfall

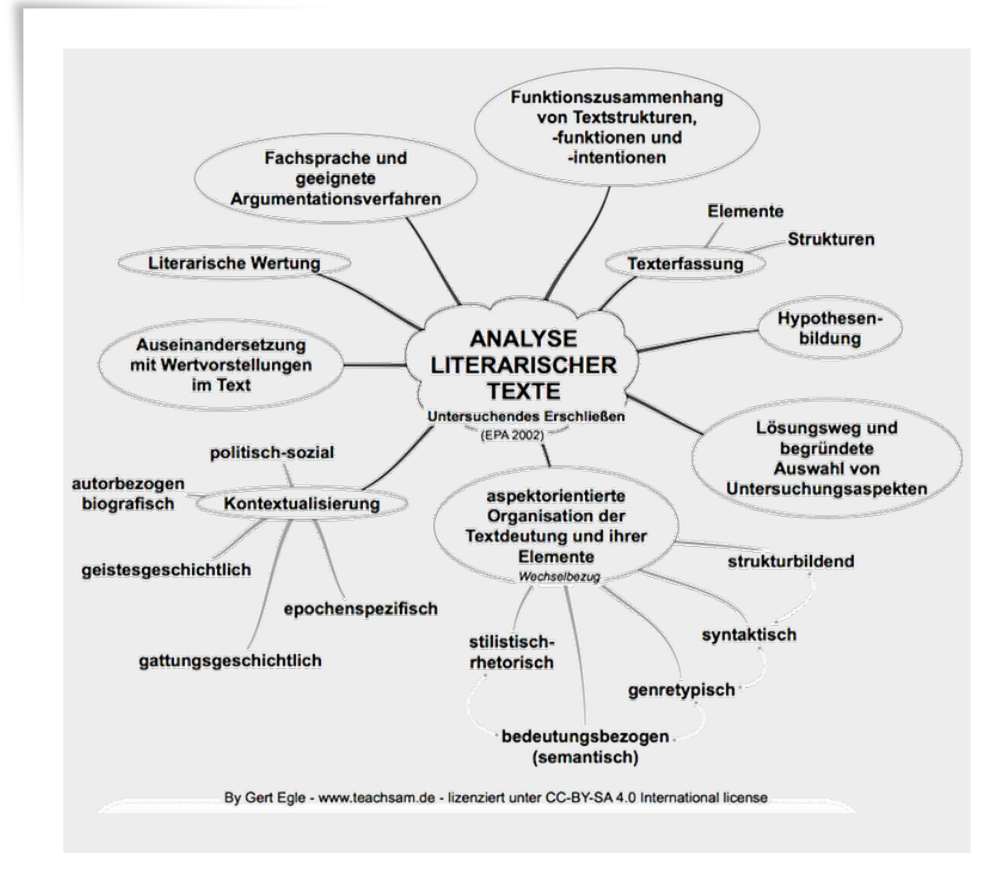
- # Textanalyse
- # Digitalgestützt
- # Forschung
- # Lehre
- # Lateinunterricht



[https://bullygoodwriting.files.wordpress.com/2013/10/analysis\\_comic.png?w=960](https://bullygoodwriting.files.wordpress.com/2013/10/analysis_comic.png?w=960)

# Textanalyse

Definition, Formen, Ziele



[https://teachsam.de/deutsch/d\\_schreibf/schr\\_schule/txtinterpr/mmf/images/analyse%20literarischer%20texte%20untersuchendes%20erschliessen%20760px.png](https://teachsam.de/deutsch/d_schreibf/schr_schule/txtinterpr/mmf/images/analyse%20literarischer%20texte%20untersuchendes%20erschliessen%20760px.png)

# Textanalyse ...

---

- ... ist
  - ein Prozess, um sich ein Textverständnis zu erarbeiten
  - ein Textgenre zur Dokumentation von Textverständnis
- ... erfordert
  - deklaratives, prozedurales und exekutives Wissen und
  - ein sinnvolles **Wechselspiel** aus **Makro- und Mikroanalysen** (distant & close reading)
- ... zielt
  - auf (**hermeneutisch** gewonnenes) **Textverständnis** und
  - auf (weiterführende) **Interpretation** des Verstandenen

# Formen der Textanalyse

- manuell, semi-automatisch (mixed) oder automatisch
- Fokus:
  - **linguistisch:** Phonologie, Morphologie, Syntax, Semantik, Lexik
  - **literaturwissenschaftlich:** intra- und intertextuelle Interpretation, intra- und interkulturelle Annäherungen, Rezeption & Transformation
  - **editionsphilologisch:** Kollation, kritischer Apparat
  - **translatologisch:** verschiedene Paradigmata
  - **Kommentar / Metadaten:** Personen, Orte, Zeit, Werk, Testimonia etc.
- Level der Analyse ist je nach Können und Arbeitsschwerpunkt unterschiedlich

# Digitalgestützt

Digitalisate, Hilfsmittel und Methoden



Agroecius, De orthographia (ca. 450): Wortfrequenzen in Form eines Anführungszeichens visualisiert

# Digitalgestützt, d. h. ...



... Einsatz von fachunabhängigen digitalen **Hilfsmitteln**

- Zur **Digitalisierung** eines Textes, z. B. Scan-Software
- Zur Beschaffung eines Textes, z. B. Suchmaschinen, Foren
- Zur Darstellung eines Textes, z. B. Office-Programme, Lernplattform
- Zur **Informationsgewinnung** zum Text, z. B. Chatbots, Wikipedia
- Zur Analyse eines Textes, z. B. Lern-Apps, Lernplattform, Tabellenkalkulation
- Zur Übersetzung eines Textes, z. B. Online-Übersetzungen
- Zur Visualisierung von Interpretationsergebnissen, z. B. Mind-Map-Tools
- Zur **Reflexion** über einen Text, z. B. Abstimmungstools, Tools für kollaboratives Schreiben



# Digitalgestützt, d. h. ...



... Einsatz von **fachspezifischen** digitalen **Hilfsmitteln**

- Zur **Digitalisierung** eines Textes, z. B. [Transkribus](#)
- Zur Beschaffung eines Textes, z. B. [Perseus](#), [Scaife](#), [DigiLibLT](#)
- Zur Darstellung und Nutzung eines Textes, z. B. digitale Editionen wie in der [Digital Latin Library](#), TEI-Editoren
- Zur **Informationsgewinnung** zum Text, z. B. [ToposText](#)
- Zur **Analyse** eines Textes, z. B. Annotationstools wie [Inception](#) und [Recogito](#), Analysetools wie [Alpheios](#) und [A Latin Macronizer](#)
- Zur Übersetzung eines Textes, z. B. [Ugarit](#) als Übersetzungsumgebung





# Digitalgestützt, d. h. ...



... Einsatz von fachspezifischen digitalen **Methoden ohne Code**

## Data Literacy

- Zum **Lesen** des Textes, z. B. [Scaife](#) (inkl. Analysetools)
- Zur Einschätzung der Textkomplexität, z. B. [Machina Callida](#)
- Zur **morphologischen Analyse** einzelner Formen, z. B. [Scaife](#), [Alpheios](#)
- Zur syntaktischen Analyse, z. B. [The Ancient Greek and Latin Dependency Treebank](#) ([Perseids](#))
- Zur **Lexik**, z. B. [Logeion](#), [Navigium](#), [LiLa](#)
- Zur Suche paralleler Textstellen, z. B. [Tesserae](#)
- Zur Paraphrasensuche, z. B. [Paraphrasis](#)
- Zur Konkordanz-Ansicht, z. B. [Machina Callida](#) (Übung „Beispielkontext“)
- Zur Visualisierung von Kookkurrenzen, z. B. [eAqua](#)



— Digital Research for All —



Gefördert durch  
DFG  
Deutsche  
Forschungsgemeinschaft

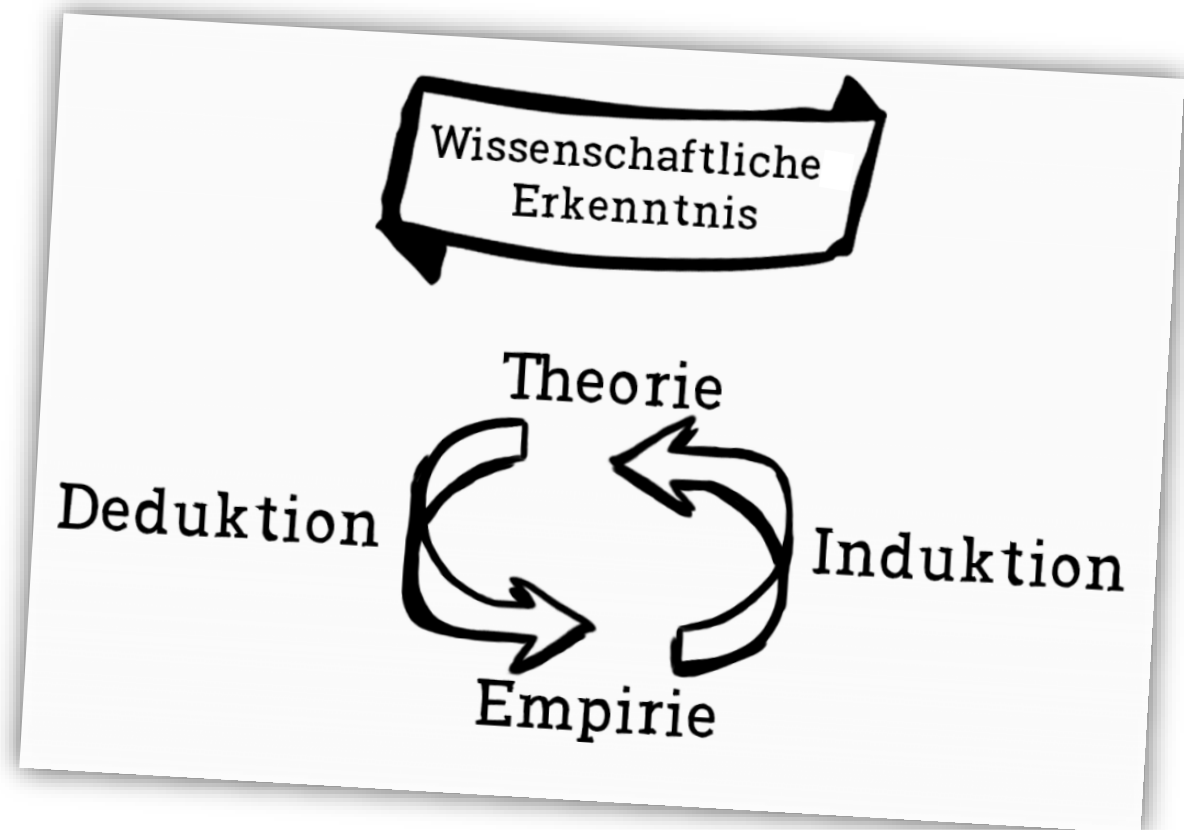
# Digitalgestützt, d. h. ...

... Einsatz von (generischen) digitalen **Methoden mit Code**

- Zur **Digitalisierung** von Texten, z. B. Optical Character Recognition (OCR)
- Zur Zerlegung von Texten bis auf Wort- und Zeichenebene (Tokenisation)
- Zur Zerlegung von Wortformen bis zum Wortstamm (Stemming)
- Zur **Zuweisung einer Wortform zu einem Lemma** (Lemmatisation)
- Zur Bestimmung der Wortart (POS-Tagging)
- Zur Analyse der Syntax (Dependency Parsing)
- Zur Informationsentnahme, z. B. Named Entity Recognition
- Zur Bestimmung von Themen, z. B. Topic Modelling
- Zur Analyse von wertenden Informationen, z. B. Sentiment Analysis
- Zur **Generierung von Text(antworten)**, z. B. Transformer-Modelle

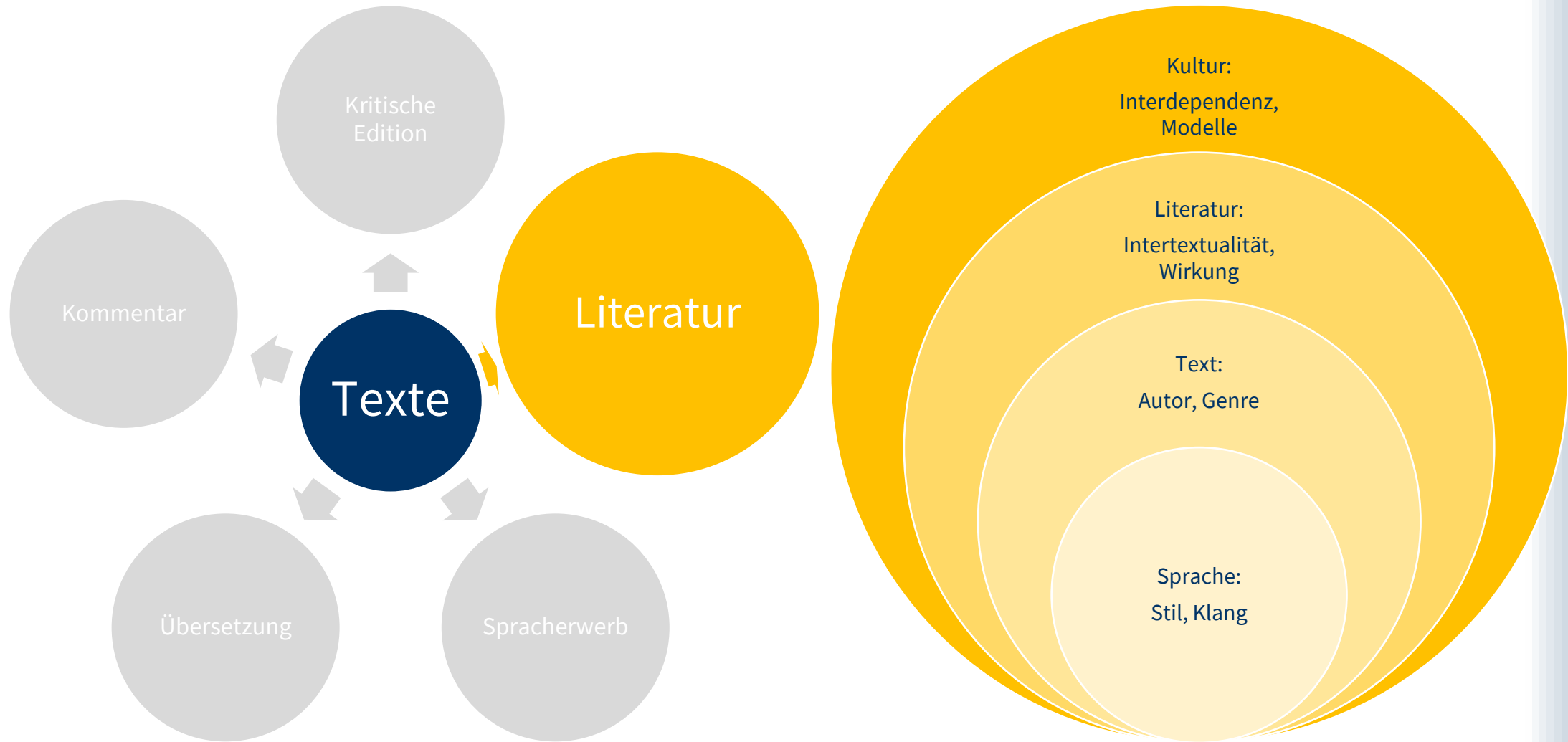
# Forschung

Klassische Philologie als  
textbasierte Wissenschaft



<https://blogs.houu.de/wissarbeiten/viii-forschung/01-forschungsmethoden/>

# Forschungsfelder



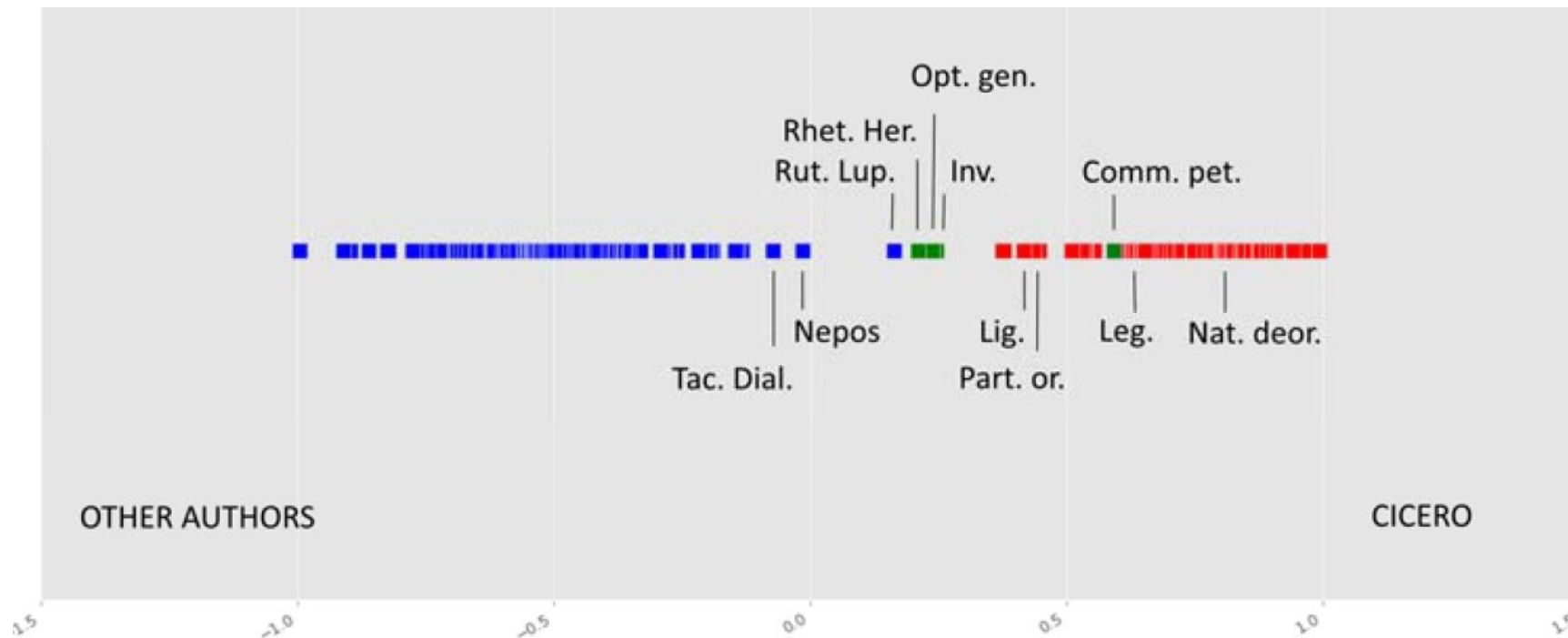
— Digital Research for All —



Gefördert durch  
Deutsche  
Forschungsgemeinschaft

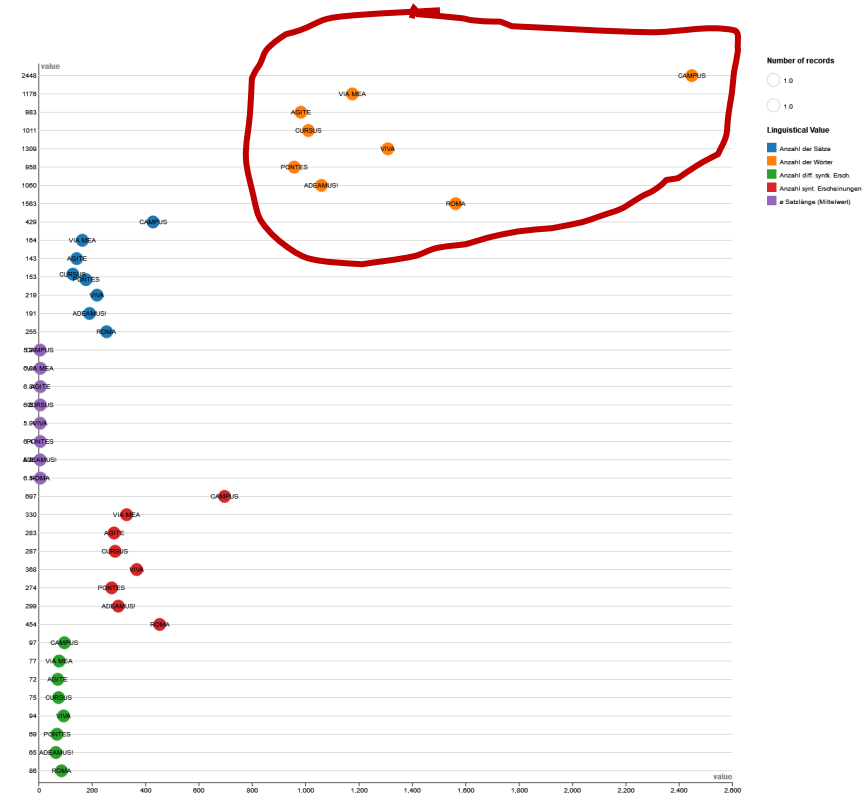
# Autorschaft des *Commentariolum*

„*Commentariolum petitionis* was written, at least largely, by Marcus Cicero. It is possible that his brother Quintus participated in writing it, but we do not consider any major input by him probable, since it would be unlikely that the brothers had such a uniform style as that in the text, and since our analyses very strongly indicate a Ciceronian authorship.” (Vainio et al. 2019, 37)



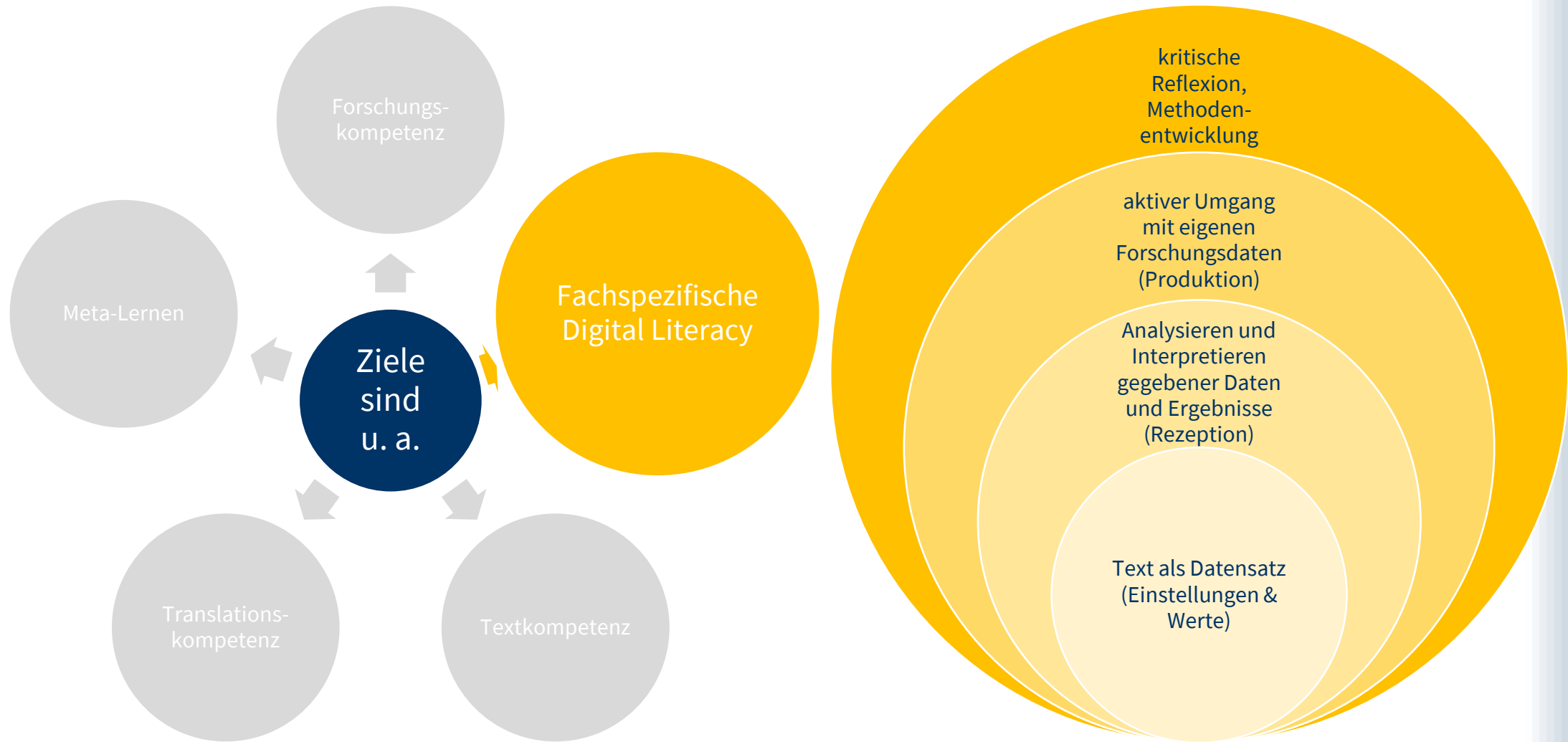
# Lehre

Qualifizierung für Forschung und Beruf



8 Lehrbücher im Vergleich: linguistische Werte, z. B. Anzahl der Wörter in 6 Lektionen / Lehrbuch (umkreist)

# Lehrziele



— Digital Research for All —

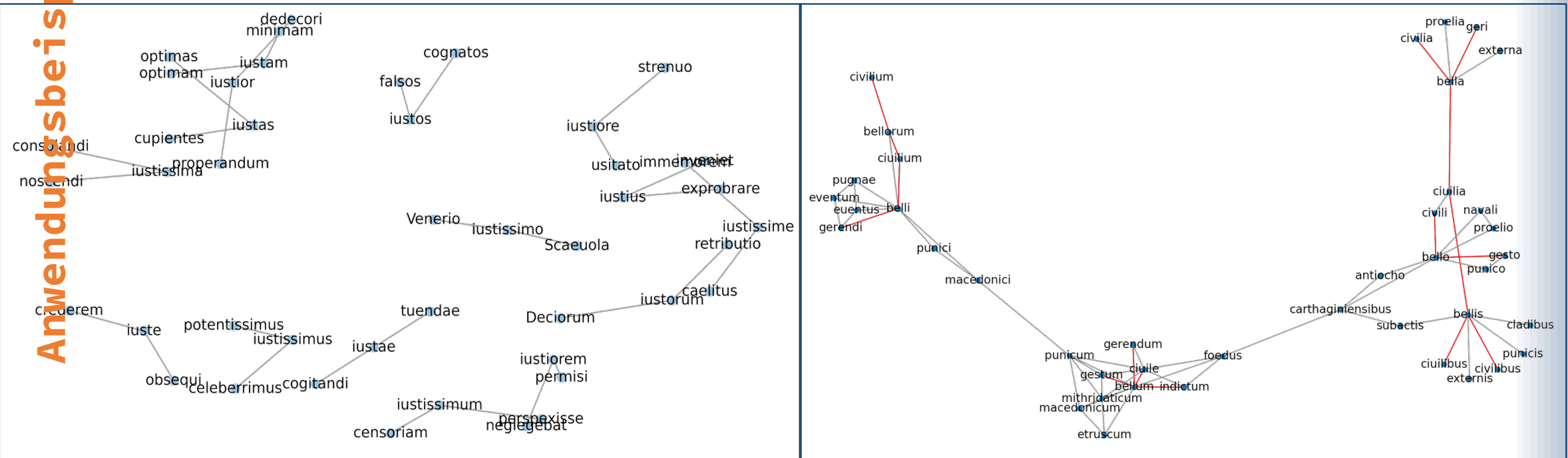


Gefördert durch  
Deutsche Forschungsgemeinschaft

# Kollokation *bellum iustum*

„Interpretieren Sie die beiden Graphen hinsichtlich der Frage, ob *bellum iustum* ein antikes Konzept ist.“

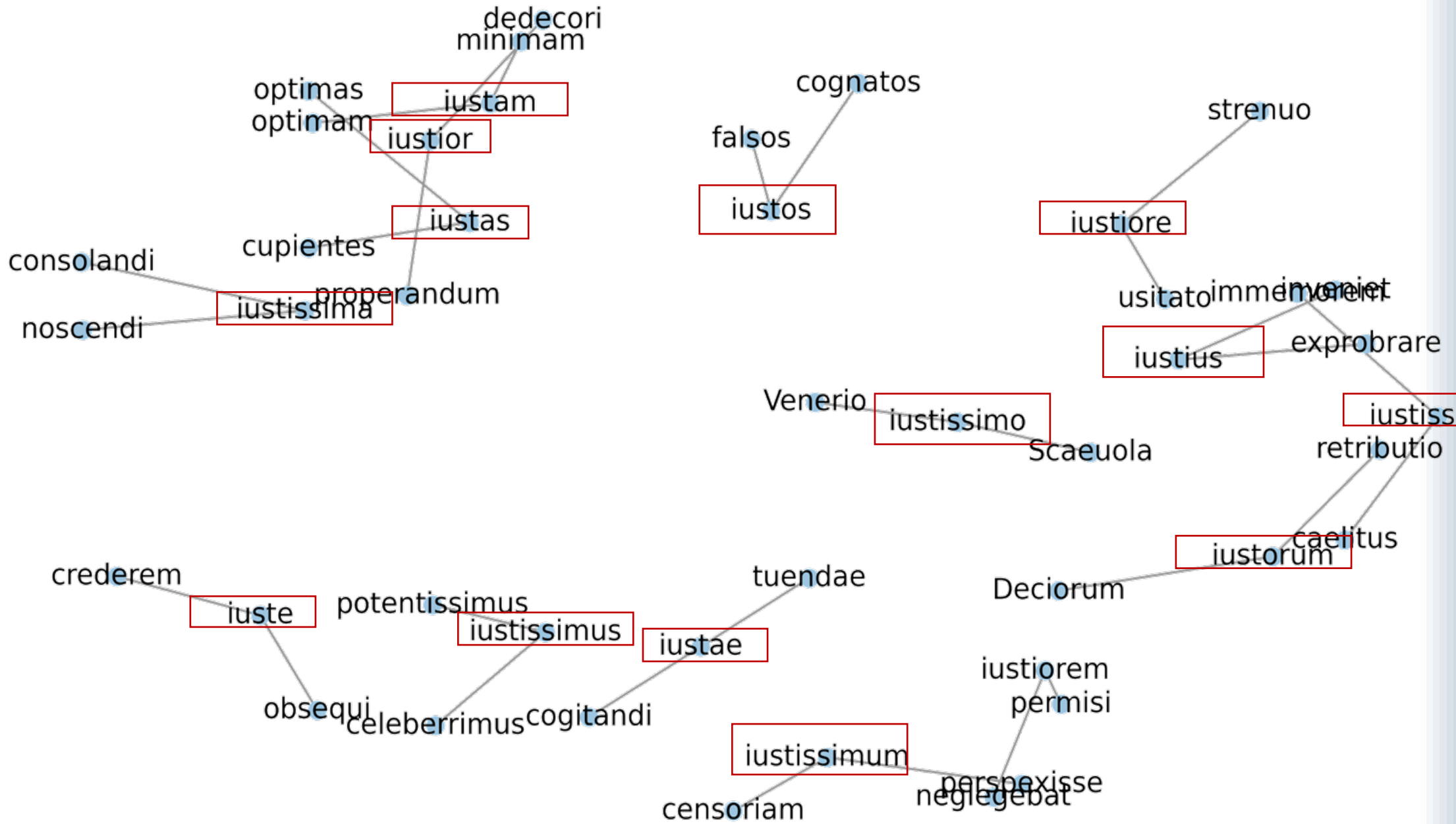
Anwendungsbeispiel

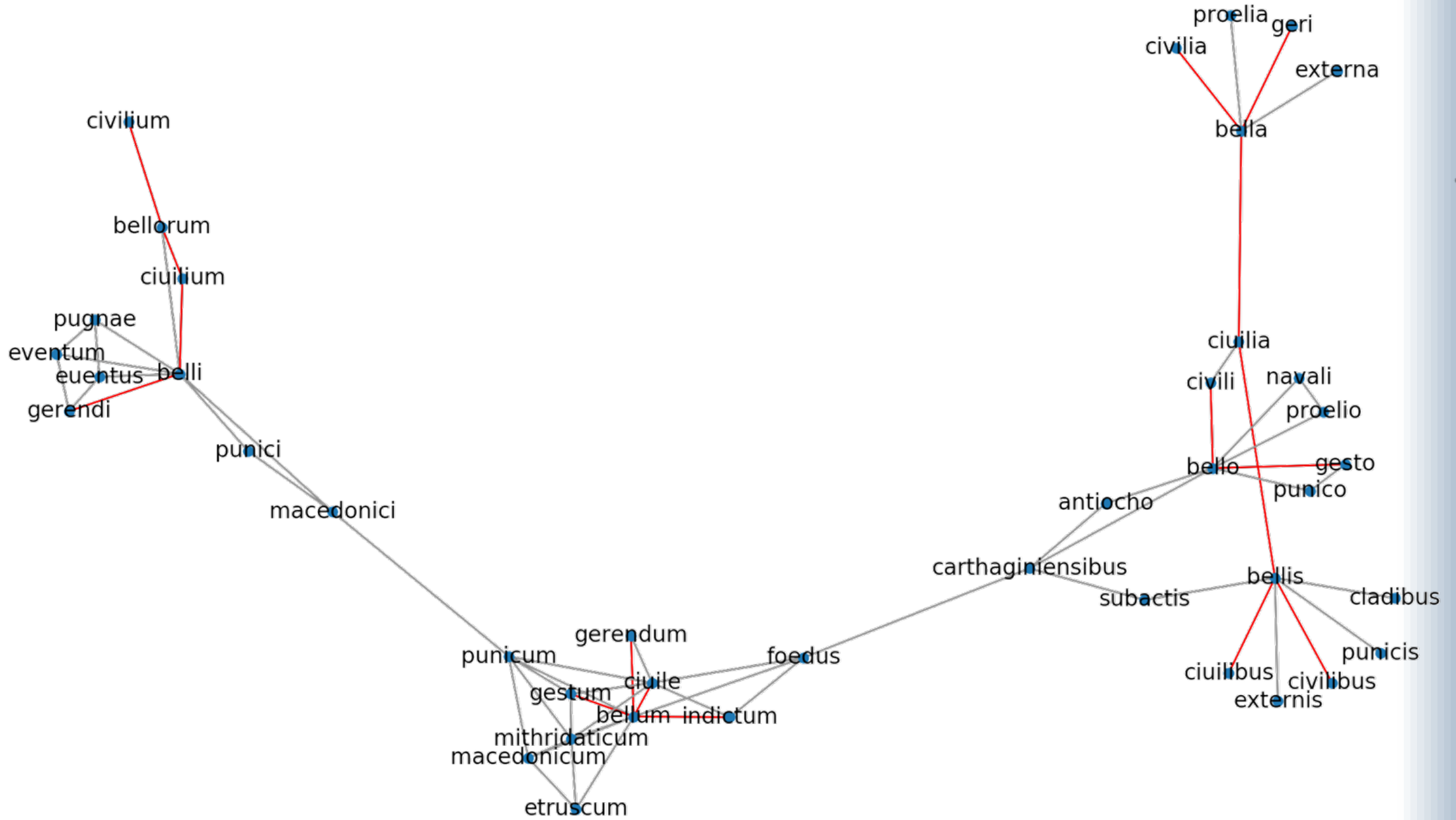


DRaIdalos  
Digital Research for All

DFG  
Deutsche Forschungsgemeinschaft  
Gefördert durch







DAIIDALOS  
— Digital Research for All —

Gefördert durch  
**DFG** Deutsche  
Forschungsgemeinschaft

# Lateinunterricht

Digitalgestützte Textanalyse  
im Literaturunterricht



Wie unterscheiden sich die Reden Caesars und Catos in Sallusts Monographie Coniuratio Catilinae?



In Sallusts Monographie [...]

1. Caesars Rede:

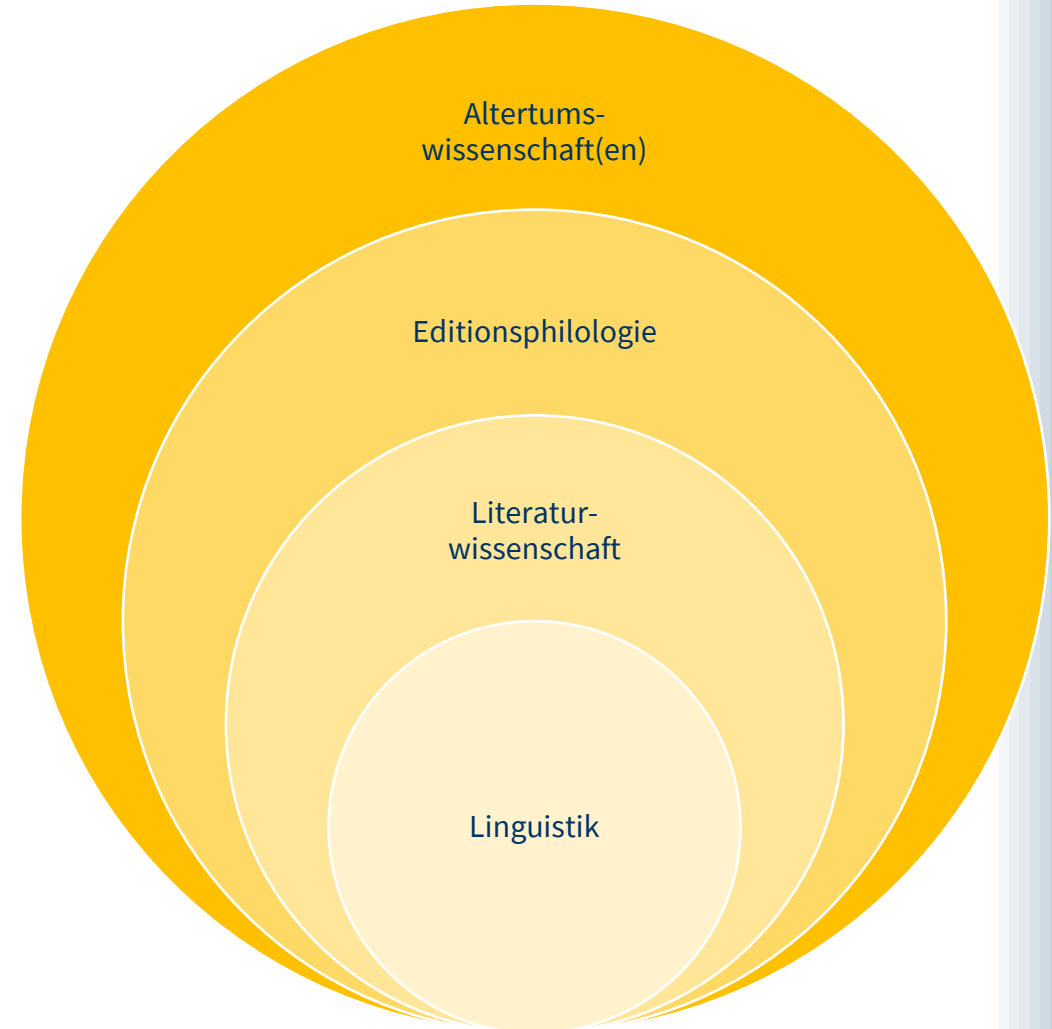
- Caesar versucht, die Anklagen gegen Catilina zu mildern und ihn vor der Verurteilung zu schützen.
- Er argumentiert für Mäßigung und Gnade und appelliert an das Mitgefühl des Senats.
- [...]

2. Catos Rede:

- Cato verurteilt Catilina scharf und fordert energische Maßnahmen gegen die Verschwörung.
- [...]

Diese Reden sind ein Schlüsselteil von Sallusts Werk und [...]

# Ziele des Literaturunterrichts Latein



— Digital Research for All —

Gefördert durch



Deutsche  
Forschungsgemeinschaft

# Persönlicher Stil oder bewusster Kontrast?

## (Sallusts Reden, Cat. 51-52)

### Reden in der Geschichtsschreibung: Funktion und Stil

„Sallust ordnet in seiner Monographie *De coniuratione Catilinae* die Reden der beiden – keineswegs gleichrangigen – Politiker Caesar und Cato direkt hintereinander an. Ein Vergleich liegt nahe...

1. Arbeitet mit Textbelegen die Argumentation der Redner heraus und visualisiert sie, z. B. tabellarisch.
2. Untersucht den Stil der Redner, indem ihr
  - a. die Ergebnisse der Machina Callida euch gegenseitig erklärt,
  - b. die Antworten von ChatGPT gemeinsam reflektiert und auswertet,
  - c. auf dieser Basis Stilmerkmale festlegt und die beiden Reden selbst untersucht.
3. Entwerft eine Argumentation, bei der ihr entweder zeigt, dass
  - a. Sallust versucht hat, den persönlichen Stil eines Redners nachzuahmen, oder
  - b. dass er versucht hat, die Reden stilistisch stark voneinander abzugrenzen, um den Kontrast und damit den erzählerischen Erfolg („auf des Messers Schneide“) zu erhöhen.

# Ergebnisse der fachspezifischen NLP-Software



Linguistik & Literaturwissenschaft

## Rede Caesars, Cat. 51

Textkomplexität

- Anzahl der Wörter: **1035**
- Wörter pro Satz ( $\emptyset$ ): **32.34**
- Lexikalische Dichte: **0.57**
- Anzahl der Nebensätze: **53**
- Anzahl der Infinitivkonstruktionen: **10**
- Anzahl der Partizipien: **27**

## Rede Catos, Cat. 52

Textkomplexität

- Anzahl der Wörter: 881
- Wörter pro Satz ( $\emptyset$ ): 20.49
- Lexikalische Dichte: 0.6
- Anzahl der Nebensätze: 34
- Anzahl der Infinitivkonstruktionen: 2
- Anzahl der Partizipien: 16

Sprachliche Unterschiede sind erkennbar. Sallust lässt Caesar länger sprechen, weist ihm durchschnittlich längere Sätze mit mehr Nebensätzen zu und legt Caesar auch in Relation zur Gesamtextlänge mehr Partizipien in den Mund (2,61 % zu 1,82 % der Wörter sind Partizipien).



# Welche sprachlichen Unterschiede gibt es[...]?



Die sprachlichen Unterschiede zwischen den Reden von Caesar und Cato in Sallusts „Coniuratio Catilinae“ sind in der antiken lateinischen Rhetorik und im Schreibstil deutlich erkennbar.

Caesars Rede:

- Caesars Schreibstil ist bekannt für seine Eloquenz und Geschmeidigkeit. **allgemein**
- Er verwendet oft lange und kunstvolle Sätze sowie rhetorische Figuren wie Parallelismus und Anaphora ([...]). **allgemein**
- Caesar neigt dazu, subtile Argumente und Nuancen in seinen Sätzen einzubauen. **allgemein**

Catos Rede:

- Catos Schreibstil ist hingegen direkter und schroffer. **allgemein**
- Er verwendet kurze, prägnante Sätze, die seine Entschlossenheit und Strenge unterstreichen. **allgemein**
- Cato verzichtet oft auf die verzierte Rhetorik, die Caesar verwendet, und geht stattdessen direkt zur Sache. **allgemein**

[...]

# Woran machst du diese Unterschiede fest?



ChatGPT

[...]. Hier sind einige Möglichkeiten, wie diese Unterschiede erkannt werden können:

1. **Satzlänge und -struktur:** Durch das Identifizieren von kurzen, prägnanten Sätzen in Catos Rede im Gegensatz zu den komplexeren und kunstvolleren Sätzen in Caesars Rede kann man den Unterschied im Schreibstil feststellen.
2. **Verwendung rhetorischer Figuren:** Caesar verwendet oft rhetorische Figuren wie Parallelismus, Anaphora und Chiasmus, um seine Argumente zu verstärken. Cato hingegen neigt dazu, diese weniger häufig oder überhaupt nicht zu verwenden.
3. **Ton und Ausdruck:** Der Ton, den die beiden Redner in ihren Reden anschlagen, ist unterschiedlich. Caesars Ton ist oft milder und einfühlsamer, während Cato einen strengeren und energischeren Ton anschlägt.
4. **Verwendung von Adjektiven und Adjektivphrasen:** Die Art und Weise, wie Adjektive und Adjektivphrasen in den Reden verwendet werden, kann ebenfalls auf die Unterschiede im Schreibstil hinweisen. Caesar verwendet möglicherweise mehr beschreibende und nuancierte Adjektive, während Cato sich auf direktere Adjektive beschränkt.

Diese Unterschiede können [...]. Die antike Rhetorik und die Literaturkritik sind hilfreiche Werkzeuge, um diese Unterschiede zu identifizieren und zu interpretieren.



— Digital Research for All —  
**daidalos**



Gefördert durch  
Deutsche  
Forschungsgemeinschaft



# Woran machst du diese Unterschiede fest?



[...]. Hier sind einige Möglichkeiten, wie diese Unterschiede erkannt werden können:

1. Satzlänge und –struktur: Durch das Identifizieren von **kurzen, prägnanten Sätzen** in Catos Rede im Gegensatz zu den **komplexeren und kunstvolleren Sätzen** in Caesars Rede kann man den Unterschied im Schreibstil feststellen.
2. Verwendung rhetorischer Figuren: Caesar **verwendet oft rhetorische Figuren** wie Parallelismus, Anaphora und Chiasmus, um seine Argumente zu verstärken. Cato hingegen neigt dazu, diese **weniger häufig oder überhaupt nicht** zu verwenden.
3. Ton und Ausdruck: Der Ton, den die beiden Redner in ihren Reden anschlagen, ist unterschiedlich. Caesars **Ton ist oft milder und einfühlsamer**, während Cato einen **strengeren und energischeren Ton** anschlägt.
4. Verwendung von Adjektiven und Adjektivphrasen: Die Art und Weise, wie Adjektive und Adjektivphrasen in den Reden verwendet werden, kann ebenfalls auf die Unterschiede im Schreibstil hinweisen. Caesar verwendet **möglicherweise mehr beschreibende und nuancierte Adjektive**, während Cato sich auf **direktere Adjektive** beschränkt.

Diese Unterschiede können [...]. Die antike Rhetorik und die Literaturkritik sind hilfreiche Werkzeuge, um diese Unterschiede zu identifizieren und zu interpretieren.

**Rot = allgemeine, unspezifische Aussagen**

# Stilmerkmale: Satzlänge und -struktur



## Machina Callida

- Caesar (1035 W.):
  - Wörter pro Satz (Ø): 32.34
  - Anzahl der Nebensätze: 53
  - Anzahl der Infinitivkonstruktionen: 10
  - Anzahl der Partizipien: 27
- Cato (881 W.)
  - Wörter pro Satz (Ø): 20.49
  - Anzahl der Nebensätze: 34
  - Anzahl der Infinitivkonstruktionen: 2
  - Anzahl der Partizipien: 16

## ChatGPT

- Caesar:
  - „komplexeren und kunstvolleren Sätzen in Caesars Rede“
- Cato:
  - „kurzen, prägnanten Sätzen in Catos Rede“



# Stilmerkmale: Satzlänge und -struktur



## Machina Callida

- Caesar (1035 W.):
  - Wörter pro Satz (Ø): 32.34
  - Anzahl der Nebensätze: 53
  - Anzahl der Infinitivkonstruktionen: 10
  - Anzahl der Partizipien: 27
- Cato (881 W.)
  - Wörter pro Satz (Ø): 20.49
  - Anzahl der Nebensätze: 34
  - Anzahl der Infinitivkonstruktionen: 2
  - Anzahl der Partizipien: 16

## ChatGPT

- Caesar:
  - „komplexeren und kunstvolleren Sätzen in Caesars Rede“
- Cato:
  - „kurzen, prägnanten Sätzen in Catos Rede“

Stilmerkmale gemeinsam konkretisieren



Daidalos  
Digital Research for All



Gefördert durch  
Deutsche  
Forschungsgemeinschaft

# Fragen ???

## Sehr gerne, falls noch Zeit ist ...

Interesse an KI und wie ein LLM funktioniert?

@chatbot: warum kannst du latein et quo vadis?

11.10.23 um 18.30 Uhr online (<https://eveeno.com/123921049>)

Infos zu KI und KI-Bildung: <https://zenodo.org/record/8381513>

# Literatur

- Beyer, Andrea, u. a. Basiswissen Lehrerbildung: Latein unterrichten. Herausgegeben von Stefan Kipf u. a., Klett/Kallmeyer, 2021.
- Baker, Pam. ChatGPT für Dummies. John Wiley & Sons, 2023.
- Egger, Roman, und Enes Gokce. „Natural Language Processing (NLP): An Introduction“. Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications, herausgegeben von Roman Egger, Springer, 2022, S. 307–34. Springer Link, [https://doi.org/10.1007/978-3-030-88389-8\\_15](https://doi.org/10.1007/978-3-030-88389-8_15).
- Girardet, Klaus M. „Gerechter Krieg“. Von Ciceros Konzept des bellum iustum bis zur UNO- Charta“. Gymnasium, Nr. 114, 2007, S. 1–36.
- Jänicke, S., u. a. On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges. The Eurographics Association, 2015. <http://dx.doi.org/10.2312/eurovisstar.20151113>.
- Liebsch, Ann-Catherine. „Cogitamus ergo sumus: Kognitiv aktivierende Aufgaben im Lateinunterricht“. Pegasus-Onlinezeitschrift, 2020, S. 1–33. <https://doi.org/10.11588/pegas.2020.0.78238>.
- Moretti, Franco. Distant Reading. Verso Books, 2013.
- Nguyen, Lan Anh Thuy, und Anita Habók. „Tools for Assessing Teacher Digital Literacy: A Review“. Journal of Computers in Education, Januar 2023. Springer Link, <https://doi.org/10.1007/s40692-022-00257-5>.
- Nünning, Vera, und Ansgar Nünning. Methoden der literatur- und kulturwissenschaftlichen Textanalyse: Ansätze – Grundlagen – Modellanalysen. Springer-Verlag, 2016.
- Ralle, B. Lernaufgaben entwickeln, bearbeiten und überprüfen. Waxmann, 2014.
- Schüller, Katharina, u. a. „Future Skills: Ein Framework für Data Literacy“. AStA Wirtschafts- und Sozialstatistisches Archiv, Bd. 13, Nr. 3–4, 2019, S. 297–317. <https://doi.org/10.1007/s11943-019-00261-9>.
- Vainio, Raija, u. a. „Reconsidering Authorship in the Ciceronian Corpus through Computational Authorship Attribution“. Ciceroniana online, Bd. 3, Nr. 1, 2019, S. 15–48.
- Zhao, Wayne Xin, u. a. A Survey of Large Language Models. arXiv:2303.18223, arXiv, 11. September 2023. arXiv.org, <http://arxiv.org/abs/2303.18223>.