

Berlin Mathematics Research Center

MATH+



berlin-brandenburgische
AKADEMIE DER WISSENSCHAFTEN

EF5-4: The Evolution of Ancient Egyptian

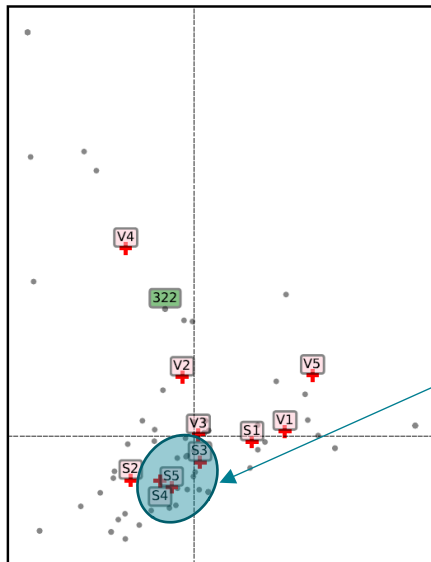
T. S. Richter, R. Birk, M. Weber*, K. Fackeldey



more general

Aims

aims of comparative text analysis



Historical changes: What differentiates the use of language before some event compared to the language after this event?

Plagiarism: What differentiates the first part of the text from the second part?

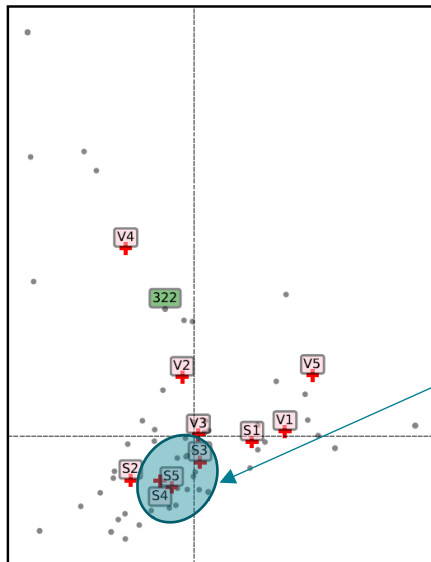
Intellectual Content, Sequencing, Authorship, Genre, Linguistic Diachrony...

Clusteranalysis



Aims

aims of comparative text analysis

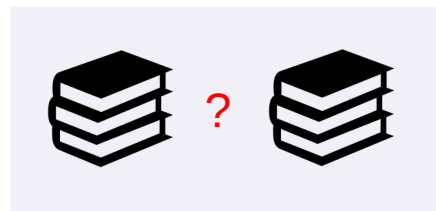


Historical changes: What differentiates the use of language before some event compared to the language after this event?

Plagiarism: What differentiates the first part of the text from the second part?

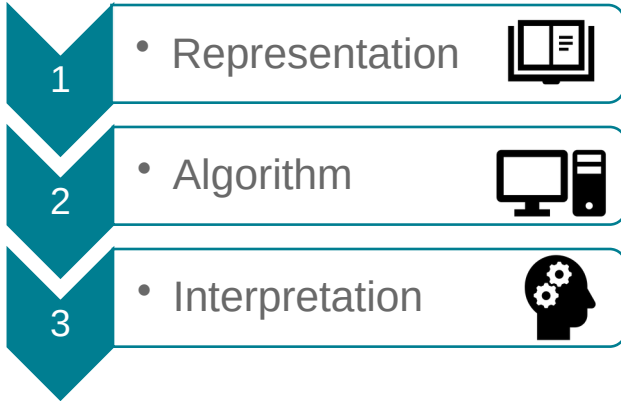
Intellectual Content, Sequencing, Authorship, Genre, Linguistic Diachrony...

Clusteranalysis



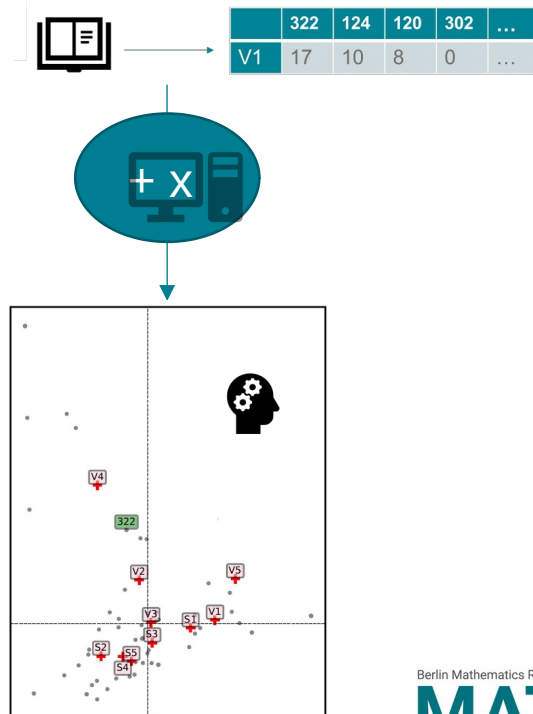
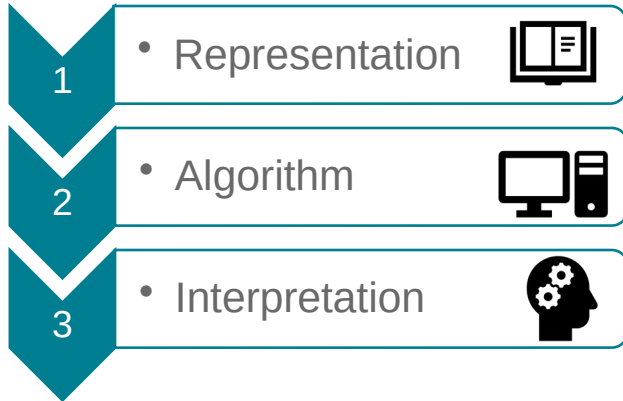
General Approach of Quantitative Studies

operationalization

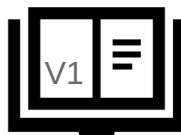


General Approach of Quantitative Studies

operationalization



1. Representation

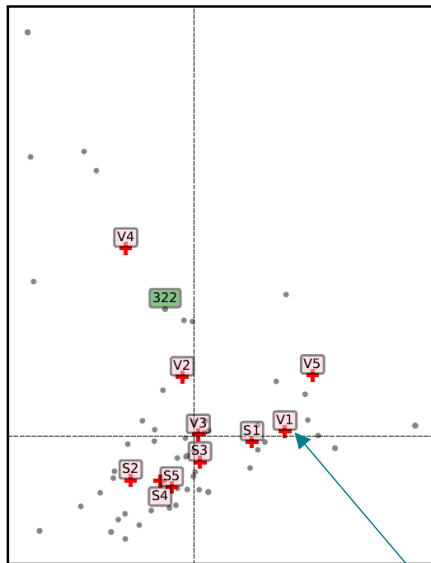


| | | | | | |
|------|-----|-----|-----|-----|-----|
| | 322 | 124 | 120 | 302 | ... |
| text | 17 | 10 | 8 | 0 | ... |

„everything that can be said about the text“

2. Algorithm

task: „represent texts AND operations on texts in machine“



| | | | | | |
|------|-----|-----|-----|-----|-----|
| | 322 | 124 | 120 | 302 | ... |
| text | 17 | 10 | 8 | 0 | ... |



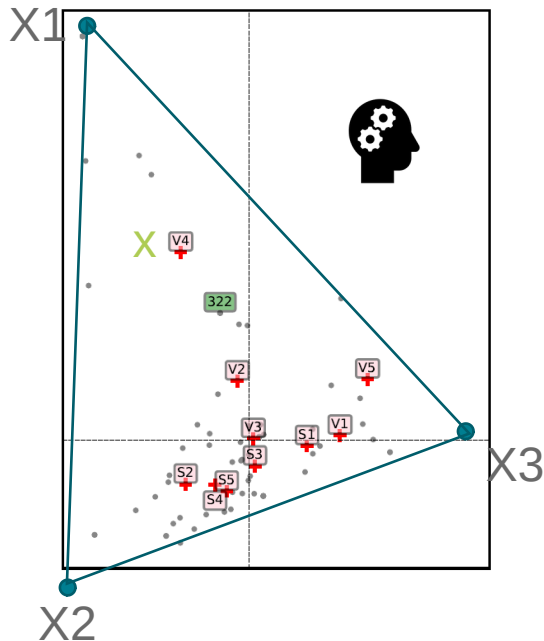
FIELD RULES

$a(b+c) = ab + bc$
 $a + b = b + a$
 $a + (-a) = 0$
 $a + (b+c) = (a+b) + c$
 $(ab)c = a(bc)$
 $ab = ba \dots$

| | | |
|------|-----|-----|
| | x | y |
| text | 0.9 | 0.1 |

Interpretation

coordinates have a meaning in terms of clustering



$$x = c_1 X_1 + c_2 X_2 + c_3 X_3$$

S. Röblitz, M. Weber (2013): "Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification"

Evelyn Gius (I-V) / Jaques Hadamard (I-III)

- I) Existence of solution
- II) Uniqueness of solution
- III) Continuity of solution
- IV) „Too complex“ – too few texts
- V) Gain in knowledge

E. Gius (2019): „Computationelle Textanalysen als fünfdimensionales Problem: Ein Modell zur Beschreibung von Komplexität“

J. Hadamard (1902): «Sur les problèmes aux dérivées partielles et leur signification physique.»

What's wrong... with the representation?



| | | | | | |
|------|-----|-----|-----|-----|-----|
| | 322 | 124 | 120 | 302 | ... |
| text | 17 | 10 | 8 | 0 | ... |

„Everything that can be said about the text“

What's wrong... with the representation?



| | | | | | |
|------|-----|-----|-----|-----|-----|
| | 322 | 124 | 120 | 302 | ... |
| text | 17 | 10 | 8 | 0 | ... |

„Everything that can be said about the text“



What's wrong... with the representation?



→

| | | | | | |
|------|-----|-----|-----|-----|-----|
| | 322 | 124 | 120 | 302 | ... |
| text | 17 | 10 | 8 | 0 | ... |

„Everything that can be said about the text“



What's wrong... with the representation?



| | 322 | 124 | 120 | 302 | ... |
|------|-----|-----|-----|-----|-----|
| text | 17 | 10 | 8 | 0 | ... |

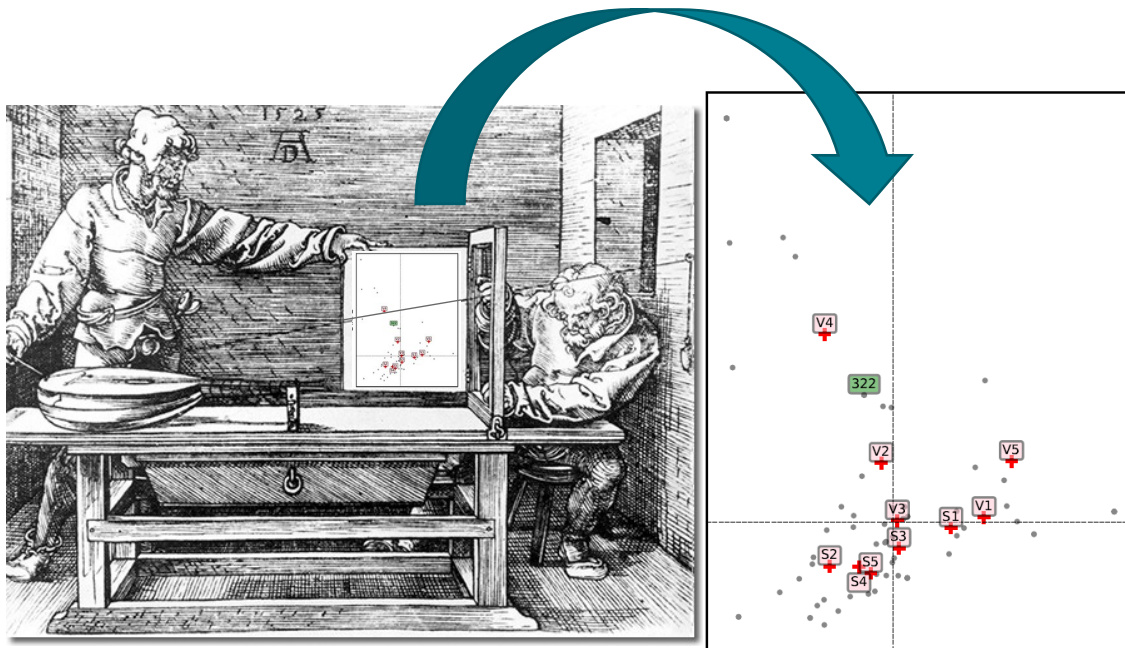
„Everything that can be said about the text“



| | T1 | T2 | T3 | T4 | ... |
|------|----|----|----|----|-----|
| text | 0 | 1 | 0 | 1 | ... |

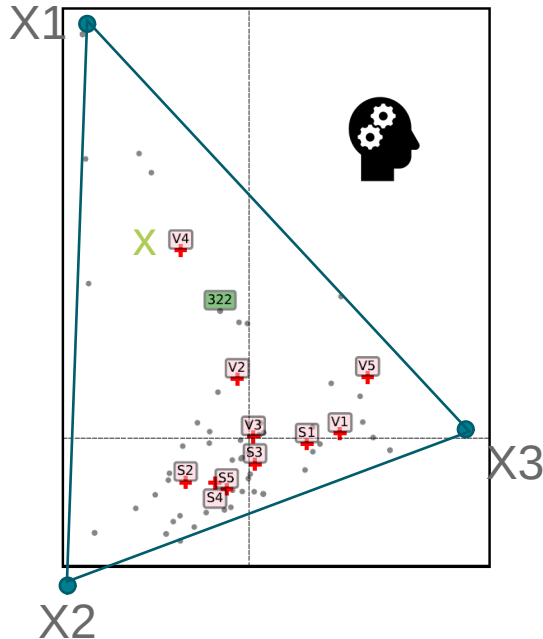
Complexity Reduction

Task: “represent texts AND operations on texts in machine”



What' s wrong ... with the interpretation?

coordinates have a meaning in terms of clustering



$$x = c_1 X_1 + c_2 X_2 + c_3 X_3$$

$$x = V_1 * V_2 + S_1 * S_2$$

What are the differences between the commonalities of V_1 and V_2 compared to the commonalities of S_1 and S_2 ?



Abu Dscha'far Muhammad ibn
Musa Chwārizmī (**Algorismi**)

al-Kitab al-muhtasar fi hisab

al-gabr wa-l-muqabala (824)

M. Weber, K. Fackeldey: The Complexity of Comparative Text Analysis - "The Gardener is always the Murderer" says the Fourth Machine, arXiv:2012.07637 [cs.CL], 2020