New Methods in the Humanities

TextMining in ancient Greek Corpora

Dr. Stylianos Chronopoulos/ Dr. Michaela Rücker
Overview

1\textsuperscript{st} short Introduction to the project eAQUA
2\textsuperscript{nd} One tool: the search mask for exploratory data analysis
3\textsuperscript{rd} Presentation of a UseCase for an analysis of a concept, the example is the concept of „rod” in Antiquity
4\textsuperscript{th} summary of the entire research
eAQUA: Extraction of structured Knowledge from Ancient Sources

- was funded by the BMBF program „Wechselwirkungen zwischen Natur- und Geisteswissenschaften“ (2008-2011; http://www.eaqua.net/index.php)
- the project was a cooperation between the Department of Ancient History Leipzig, the Department of Natural Language Processing Leipzig and the Department of Classical Philology Leipzig
- Aim: to adapt available text mining technologies for an automatic extraction of structured knowledge from ancient sources
eAQUA developed new tools and methods for processing large digitized corpora - in particular for researching ancient knowledge transfer:

- Search function: new search strategies through semantic relationships
- Text completion: automatic suggestions for missing letters or words in fragmentary texts (e.g. papyri)
- Citation graph: search for possible quotations in ancient literature
- Mental Maps: visually supported analysis of transfer and transformation of concepts (concept change) across space and time
The Data

Greek literature 8th century BC - 1500 AD: Thesaurus Linguae Graecae (TLG); Perseus Digital Library

Latin literature 200 BC - 1300 AD: The Packard Humanities Institute (PHI 5); Patrologia Latina (PL); Perseus Digital Library; Bibliotheca Teubneriana Latina (BTL)

Documentary texts of the ancient world (papyri, inscriptions): Packard Humanities Institute (PHI 7); Duke Papyrus Archive; Codex Sinaiticus; Epiduke

Non-classical Texts: Perseus Digital Library
The eAQUA Word-Search for exploratory data analysis
Visualization reduction in complexity: speed and efficiency
The results of the explorative search

I) statistical information concerning the search term: occurrence, class of frequency, words with same normalised form and words with same base form

II) paradigmatic context of the search term: words with similar context

III) syntagmatic context of the search term: the graph and the lists of co-occurrences and neighbors

IV) quotations
significant co-occurrences

A selection of quotations
Significant co-occurrences

• frequent occurrence of two terms from a text corpus alongside each other in a certain order

• it assumes interdependency of the two terms

• in TextMining we differentiate between co-occurrence (semantic units together in one sentence) and neighborhood (semantic units are located directly next to each other)

Significance measures

• in most instances the calculation is based on the Log-likelihood measure

• it is based on the aggregate frequency of words in the selected corpus
The Visualization – word net

- the graph is a visualization of the syntagmatic relations
- the co-occurrences are displayed as nodes being connected through edges representing the significance
- that formed a netlike structure, the co-occurrences are connected with each other depending on their occurring
- The result is a word net where non-significant nodes are as far away as possible from each other and at the same time as close as possible to their significant co-occurrences
Creation and Analysis of Concept and Concept Change

The Concept: Rods
Most Commonly Used Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>baculum</td>
<td>βάκτρον/báktron</td>
</tr>
<tr>
<td>bathron</td>
<td>ὅξαβδος/rhábdos</td>
</tr>
<tr>
<td>caduceus</td>
<td>κηρύκειον/kērykeion</td>
</tr>
<tr>
<td>rudis</td>
<td>σχῆπτρον/sképtron</td>
</tr>
<tr>
<td>lituus</td>
<td></td>
</tr>
<tr>
<td>stimulus</td>
<td></td>
</tr>
</tbody>
</table>

Many terms = many concepts?
Most Commonly Used Terms

• ῥάβδος, ἡ: rod, wand, etc.; lighter than the βακτηρία or walking-stick

Special uses:

❖ magic wand, as that of Circe; fishing-rod; limed twig, for catching small birds; shaft of a hunting-spear; staff of office, like the earlier σκῆ-τρον; wand borne by the ῥαψῳδός; rod for chastisement; shepherd's staff or crook; measuring-rod; stitch

• βάκτρον, τό (baculum): stick, cudgel

• σκῆ-τρον, τό: staff or stick, used by the lame or aged; staff or baton, esp. as the badge of command, sceptre; as a symbol of royalty, kingly power

• κηρύκειον, τό: caduceus

• καμπύλη, ἡ: crook, lituus; auch Shepherd's crook

aus: LSJ
Visualization

Bsp.: tag clouds
Tag-Clouds/Word-Clouds

= a visual representation for text data, typically used to depict keyword metadata (tags) on websites, or to visualize free form text

• tags are usually single words, and the importance of each tag is shown with font size or color

• it is useful for quickly perceiving the most prominent terms and for locating a term alphabetically to determine its relative prominence
The concept „rod“ in eAQUA
entering of the search word in to the search field

The word and its co-occurrences

1. η βακτηρία
• eAQUA-search field = semantic search

• Task: finding new correlations by a direct connection to the ancient texts compared to the preselection of texts in dictionaries and their restricted contexts

• entering of the search word in to the search field: βακτηρία

• Selection of the corpus: TLG

statistical information:

• occurrence in the selected corpus: 209

• Class of frequency: 14

• Words with same normalised form: βακτηρία (158); Βακτηρία (24); ΒΑΚΤΗΡΙΑ (1); βακτήρια (1);

• Words with same base form: βακτηρίαν; βακτηρία; βακτηρίας; βακτηρίαι; βακτηρίας; Βακτηρία; Βακτηρίαν; βακτηρίη; Βακτηρίας; βακτηρίαιν; Βακτηρίαι
• The paradigmatic context, represented by “words with similar context”: χλανίς; πιλίδιον; φαιός; χιτωνίσκος; ἀπήσθιον; βαία; στιβὰς; εὐφυθμος

• a lemmatized keyword-search is not possible
• this picture shows the graph of the feminine search term βακτηρία, ἦ, rod, stick

• it is the visualization of the search

• the initial view of the graph clearly displays visible connections between words - to enable a quick view the co-occurrences in this first sight are limited

• mark of two word nets – two different contexts?

• the visualization also displays the other forms of the search term and other terms for „rod“ - Kampyle and Rhabdos
• Below the graph there is a list of the calculated co-occurrences, the visualization is based on this computation

• in the list I selected as one example the words of one context that you can also see in the graph: χιτωνίσκος; χλανίς; πιλίδιον; φαιδς; εὐρυθμος; ἀπαλόν

• behind the words you see in parentheses the number of quotation
  • that means for instance a combination of the search term βακτηρία and the word χιτωνίσκος
  • both words occur together 4 times
The quotations: βακτηρία und χιτωνίσκος

Belegstellen für βακτηρία und χιτωνίσκος

Anzahl 4 (βακτηρία : 209 , χιτωνίσκος : 27)

Gefundene Belegstellen

1. λευκή χλανίς, φαιός χιτωνίσκος καλός, πυλίδιον ἀπαλόν, εὐρύθυμος βακτηρία, βαία τε πέξα:
   Source: ANTIIPHANES Comic. (0410) (NULL - 4 B.C.): Fragmenta
   Vide: NULL
   Document citation: //33tit33/3 to //33tit33/5 (Schema: Fragment/line)

2. λευκή χλανίς, φαιός χιτωνίσκος καλός, πυλίδιον ἀπαλόν, εὐρύθυμος βακτηρία, βαία τράπεζα:
   Source: ANTIIPHANES Comic. (0410) (NULL - 4 B.C.): Fragmenta
   Vide: NULL
   Document citation: //Antal/1tit/3 to //Antal/1tit/5 (Schema: Play/fragment/line)

3. λευκή χλανίς, φαιός χιτωνίσκος καλός, πυλίδιον ἀπαλόν, εὐρύθυμος βακτηρία, βεβαία τράπεζα, τί μακρά δεῖ λέγειν;
   Vide: Cf. et ATHENAEUS Epigr. (0141)
   Document citation: //12tit63/55 to //12tit63/57 (Schema: Book/Käibel paragraph/line)

4. λευκή χλανίς, φαιός χιτωνίσκος καλός, πυλίδιον ἀπαλόν, εὐρύθυμος βακτηρία, βαία τράπεζα, τί μακρά δεῖ λέγειν;
   Vide: Cf. et ATHENAEUS Epigr. (0141)
   Document citation: //2,2/94/22 to //2,2/94/23 (Schema: Volume/page/line)
Ἀντιφάνης δ’ ἐν Ἀνταίῳ περὶ τῆς τῶν φιλοσόφων τρυφερότητος
dιαλεγόµενος φησιν· (Α.) ὦ ’τάν, κατανοεῖς τίς ποτ’ ἐστὶν οὗτοσὶ ὁ
gέρων; {Β.} ἀπὸ τῆς µὲν ὀψεως Ἑλληνικὸς· λευκὴ χλανίς, φαιὸς
χιτωνίσκος καλὸς, πιλίδιον ἀπαλόν, εὔρυθµος βακτηρία, βεβαία
τράπεζα—τί µακρὰ δεῖ λέγειν; ὅλως αὐτὴν ὁρᾷ γὰρ τὴν Ἀκαδηµίαν
dοκῶ.

And Antiphanes, in his Antaeus, speaking of the
luxurious habits of the philosophers, says – My friend,
now do you know who this old man is called? By his
look he seems to be a Greek. His cloak is white, his
tunic fawn-coloured, his hat is soft, his stick of
moderate size, his table scanty. Why need I say more, I
seem to see the genuine Academy.

(translated by C.D.Yonge)
The concepts in the visualization – the interpretation of the graph

Βακτηρία as accessory of clothing – attribute for:

- Greek
- sophisticated / elegant
- old man
- philosopher & member of the academy
- = identification with the academy
the complete graph for the search term *bakteria* also displays word nets clearly recognizable

that could be references to particular contexts AND

I also find interesting co-occurrences for further research for instance

1) at the first glance the mention of Poseidon and Charon seems to be very interesting, their connection with the term „rod“ is associated with special contexts

the name Charon could refer to the ferryman of greek mythology: as πορθμεύς (*porthmeús*; Minyas EpGF fr. 1) or *portitor* (‘customs agent’, Verg. Aen. 6,298 [4. 221]), transports the dead, brought to him by Hermes, across a reed-lined underworld lake for burial, in a boat with oars and pole
But: the view in to the ancient sources shows another context

- the scholion to Aristophanes „Wealth“ (Plutos) 278 illustrates the athenian practice of the appointment of judges to a man named Charon: ὁ δὲ Χάρων τὸ ξύμβολον: περὶ τοῦ παραδιδομένου τοῖς εἰσιοῦσιν εἰς τὸ δικαστήριον συμβόλου Αριστοτέλης ἐν τῇ Ἀθηναίων πολιτείᾳ σύντω γράφει „τοῖς γὰρ δικαστηρίοις χρῶμα ἐπιγεγραπται ἐφ’ ἐκάστῳ ἐπὶ τῶ σφηκίσκω τῆς εἰσόδου. ὁ δὲ λαβὼν τὴν βακτηρίαν βαδίζει εἰς δικαστήριον τὸ ὁμόχρουν μὲν τῇ βακτηρίᾳ, ἔχον δὲ τὸ αὐτὸ γράμμα ὅπερ ἐν τῇ βαλάνῳ. ἐπειδὰν δὲ εἰσέλθῃ, παραλαμβάνει σύμβολον δημοσία παρὰ τοῦ εἰληχότος ταύτην τὴν ἀρχήν“.

- the scholion refers on the one hand to Aristoph. Plut. 278: ἐν τῇ σορῷ νυνὶ λαχὼν τὸ γράμμα σου δικάζει, σὺ δ’ σὺ βαδίζεις, ὁ δὲ Χάρων τὸ ξύμβολον δίδωσιν.

„Cario: It's in the tomb that it's your lot to judge. Why don't you go there? Charon has given you your ticket [Symbolon – gr. ξύμβολον; Anmerkung d. A.]. ( E. O’Neill)
• and on the other hand to Aristotle, *Athenaion Politeia* 63-69, who described the whole process of the appointment of judges

Athen. Pol. 65: ... τοῖς γὰρ δικαστηρίωις χρωματα[τα]· ἐπιγέγραπται ἐκαστῷ ἐπὶ τῷ σφηκίσκῳ τῆς εἰσόδου· ὁ δὲ λαβὼν τὴν βακτηρίαν βαδίζει εἰς τὸ δικαστήριον τὸ ὀμόχρων μὲν τῇ βακτηρίᾳ, ἔχ[ον δὲ τὸ αὐτὸ γράμμα ὡς ἐν τῇ βαλάνῳ]. ἐπειδὰν δὲ εἰς ἑλθὼ, παραλαμβάνει σύμβολον δημοσίᾳ παρὰ τοῦ εἰληχότος ταύτῃ νῦν τὴν ἀρχήν.

„And the man himself having again shown it to the attendant then goes inside the barrier, and the attendant gives him a staff of the same color as the court bearing the same letter as the one on the acorn, in order that it may be necessary for him to go into the court to which he has been assigned by lot [Symbolon – gr. σύμβολον; note of the author]; for if he goes into another, he is detected by the color of his staff.“ (H. Rackham)

• in a different way from expecting the result of the research is a mix of an institutionell with a mythological context

• the combination of close (text/source) and distant (visualization) reading
second example: initial view of the graph for the search term ῥάβδος (occurrence: 883; ῥάβδον: 914; ῥάβδῳ: 587)

- the graph displays a structure using word nets

U.a. New Testament, Epistle to the Hebrews 9,4: „It contained the golden altar of incense and the ark of the covenant covered entirely with gold. In this ark were the golden urn containing the manna, Aaron’s rod that budded, and the stone tablets of the covenant.“
The concepts for ῥάβδος

- Magic (the goddess Circe) - ῥάβδῳ
- Feast (bacchantic)
- shepherd
- education
- monarch
Summary
• at this point it was not possible to show you the complete research of all terms
• so you have to believe me when i tell you some results:
  • mixing of the contexts by joint appearances of the different terms: among others 40 quotations for ῥάβδος and βακτηρία
  • In recognition of the importance of the definition as walking stick it seems reasonable to assume that the origin of the term is due to the mobile way of life and was transferred to other professionals
  • wandering physicians, priests, philosophers, beggars…
  • symbolize hiking activities
Inhaltliche Kontexte

Greise/ Pädagogen
als Stütze für
Kranke

Spielzeug für Kinder

Strafe/ Tötung

Instrument

Attribut für

Priester
Blinde
Könige
Bettler
Götter
Wanderer
Philosophen
Hirten
Ärzte
Magier
Richter
Symbolische Dimension

Weisheit

Magie

Medizin

Macht/Herrschaft
THANK YOU FOR YOUR ATTENTION