Corpus Studies in Ancient Greek: Exploiting "Thesaurus Linguae Graecae" and other corpora.

Dariya Rafiyenko / Ilja Serzants (Leipzig University) 5.4.2018, Riga
Corpora and Tools Online

- Thesaurus Linguae Graecae (www.stephanus.tlg.uci.edu)
- Perseus Digital Library (www.perseus.tufts.edu; review: https://ride.i-d-e.de/issues/issue-8/perseus/)
- LOFTS „The Leipzig Open Fragmentary Texts Series“ (www.dh.uni-leipzig.de/wo/lofts)
- PROIEL „Pragmatic Resources in Old Indo-European Languages“ (~hf.uio.no > Ifikk > English > Research > Projects > Proiel)
- eAQUA (www.eaqua.net)
Thesaurus Linguae Graecae (TLG)
Quick Links

- The TLG Team
- History of the TLG
- Online LSJ
- Culliffee’s Homeric Lexicon
- Powell’s Lexicon of Herodotus
- The TLG Canon
- The Abridged TLG
- Encoding
- Beta to Unicode Guide
- Unicode Proposals
- Online LBG

Our mission and our projects

News and Updates

March 29, 2018: Seventy three (73) new texts added. Lemma recognition is now up to 98.3%.
TLG: Main Information

a Special Research Program at the University of California, Irvine

a large digital corpus of literary texts

a comprehensive digital library of Greek literature from antiquity to the present era

founded in 1972
TLG texts became available to the scholarly community first on magnetic tapes (in the mid 1970s)
later in CD ROM format
four editions: 1985 (A), 1988 (C), 1992 (D) and 2000 (E)
from 2004 is no longer licensing the corpus in CD ROM format
from 2001 available online and focuses its resources on web dissemination

Available per subscription
A subcorpus together with the extensive bibliographical database (Abridged TLG, ca. 1,000 works from 70 authors) is open to public
TLG: Text collection

texts written in Greek

from Homer (8 c. B.C.) to the fall of Byzantium in AD 1453, or actually up to the 20th cent. AD

Today the Online TLG contains more than 110 million words from over 10,000 works associated with 4,000 authors

is constantly updated and improved with new features and texts

!! one must bear in mind that only 22 mln. words up to the 4AD => more than 80% is Byzantine
TLG: Advantages and Limitations

+ very accurate (there are almost no errors)
the biggest corpus as of today (ca. 110 mln words)
the best search options available
a lot of metadata to the texts

– very proprietary (user can be blocked due to "extensive" usage)
not clear in word count
made by philologists, not by corpus linguists
the metadata to the texts are not consistent (esp. genres)
no apparatus criticus (even though TLG uses editions as basis)
TLG: Structure

**TEXT DATA** (linguistic texts)

vs.

**METADATA** (CANON, information about these texts)

**Metadata:**
- For authors
- For works
- For publications
TLG Canon (metadata)
Metadata for authors

obligatory
  - canon number
  - name
  - date

optional
  - further names
  - generic epithet
  - Geographic epithet
  - non-standard generic epithet(s)
  - links to other authors or works ("cf.")
  - links to Scholia, Vita(e), Lexicogr. in other authors or works

inherited from work(s)
  - word count
Metadata for works and publications

obligatory
  canon number
  title
  publication (editor, publication title, series, year of publication)
  word count (!! has been changed in 2017)

optional
  generic epithet
  links ("cf.") to other authors or works
  edition type (NULL vs. dipl.)
  language (NULL vs. vern.)

inherited from author
  date
  geographic epithet

- 0007.067  De liberis educandis [Sp.] (1a-14c)
- 0007.068  Quomodo adolescens poetas audire debeat (14d-37b)
- 0007.079  Septem sapientium convivium (146b-164d)
- 0007.080  De superstitione (164e-171f)
Metadata: Date

with certainty

  e.g.
  1-2 AD
  1-2? AD
  1/2 AD
  1/2? AD

uncertain

  e.g.
  varia
  incerta
  ante-
  post-
Optional Metadata

Further names (author): STESICHORUS (0292) "vel Teisias"

Generic epithet (author):
   Homer (0012) Epic
   Solon (0263) Nomograph, Poeta
   History: Hist., Biogr., ScrEccl.
   Cf. stephanus.tlg.uci.edu > Helppdf > Canon.tags ;
   stephanus.tlg.uci.edu > Helppdf > TLG Classifications B.and.S

Non-standard generic epithet (author):
   DIONYSIUS II (0247) Tyrannus Siceliae

Generic epithet (work):
   Euripides Encom, Lyr : Epinicum in Alcibiadem (fragmenta)
   Trag: Alcestis, Medea, Heraclidae
Canon Search: Home page
Canon search: „Plato“
Canon: Information on author

SEARCH THE TLG CANON

ANATOLIUS in Author

Date
Generic Epithet
Geographic Epithet

Results per page: 50

CATEGORY
Filter by
Author
Geographic Epithet
Generic Epithet
Date

RESULTS
View As: List | Map | Timeline
Select to search
Select all
1. ANATOLIUS Math., Phil. {2577}
Anatolius Laodicensis
A.D. 3
Math., Phil.
Alexandrius, Caesariensis
Episcopus Laodiceae

QUICK LINKS

CROSS REFERENCE (0)

MY SEARCH SELECTIONS (1)

Go to Text Search | Clear all

PLATO Phil. {0059} ×
-Euthyphro (001) ×
-Apologeta Socrates (002) ×
-Crito (003) ×
-Phaedo (004) ×
-Cratius (005) ×
-Thearetus (006) ×
-Sophista (007) ×
-Politicus (008) ×
-Parmenides (009) ×
-Euthydemistes (007)
Canon: Information on work(s) of an author
Canon search: Information on work(s) of an author
TLG: Text data
Text is a sequence of characters (letters, digits, spaces etc.), a STRING characters form words; words form phrases.

Words can be LEMMATIZED, which means that all inflected forms of a word are grouped together so that they are analysed as a single item, identified by the word's lemma, or dictionary form.

!! Homonymic forms:
e.g. ποδαλγῶν:
- masc./fem./neut. gen. pl. of the adjective ποδαλγός / ποδαλγής
- part. praes. act. masc. nom./voc. Sg. from verb ποδαλγέω
- gen. pl. of the noun ποδάλγη
Beta Code and Beta Escapes

special code; basically, a Latin transliteration; widely used before the arrival of the Unicode

Ἀ)ΝΔΡΟ/Ϲ = ἀνδρός
*ω*ς = ΩΣ

Beta escapes: different textual and metatextual information (tlg.uci.edu > Help > Help1), as, for example:
- milestones indicating the current citation in the text
- codes indicating formatting
- sigla in text
- the semantics of the text

@1 Page End
@6 blank line
[2αβ]2 editorial conjecture
#13 asteriskos
PLATO, Crito. [0039.003]

~x43"a" @("S*W") *TI/ THNIK.A DE A)F(I=CAI, W)= *KRI/TWN; H) OU) PR
@("K*R") *PA/NU ME N OU)=N.
@("S*W") *PHN.IKA MA/LISTA;
@("K*R") *ORQROS BAQU/S.
@("S*W") *QAUMA/ZW O(PWS H)QE/LHSE/ SOI O( TOU)= DESMWTHRIOU
FULAC U(PAKOU=SAI.
@("K*R") *SUNH/QHS H)/DH MOI/ E)STIN, W)= *SW/KRATES, DIA TO PO
LA/KIS DEU=RO FOITA=N, KAI/ TI KAI EU)ERGE/THTAU U(P E)MOU=.
@("S*W") *YARTI DE H/(KEIS H) PA/LAI,
@("K*R") *EPIEIKKW=S PA/LAI.

~v

@("K*R") *EID=TA PW=S OU)K EU)QU S E)PH/GEIRA/S ME, A)LLA SLICH=
KA/QHSAI:
@("K*R") *OU) MAI/ TON *DIA/ W)= *SW/KRATES, OUD A)N AU) TO S H
EN TOSAU/TH TE AGRPNI/A KAI/ LU/PH ED=NAI, A)LLA KAI/ SOU=)
QAUMA/ZW AUSQANO/MENOS WS H/DE/WS KAQUE/DEIS; KAI/ E)PI/TH
OUK H) GEIRON I)(NA W(S H)DISTA DIA GHS. KAI/ POLLA/KIS ME N DI-
KAI PRO/TERON E)N PANTI TW= BI/W H)DAIMO/NISA TOU= TRO/POI
DE MA/LISTA E)N TH= NU=N PARESTW/SH/ SUMFORA=, W(S R(A/DI/WS
KAI PRA/WS FE/REIS. 2
@("S*W") *KAI CAI R.A)N, W)= *KRI/TWN, PLAHI/MELI S E)I/H A)GANAK
THLIKOU=TON O)NTA EI) DEI= H) DH TEL/EUTA=N.
PLATO, Crito. (0059.003)

43.

(a) @ *S*JF. *TI/ THTNIKA/DE A)F=CAI, W) = *KRI/TWN; H) OU) PRW\ E)TI
    @ *K*.R. *PA/NUMEN OU)=N
    @ *S*JF. *PHNIKA MALISTA;
    @ *K*.R. *ORQROS BAQUS.
    @ *S*JF. *QAUAM/ZW O/PWS H)QE/LHSE/ SOI O (TOU= DESMWTHRI/OU
    FU/LAC U (PAKOU=SAI.
    @ *K*.R. *SUNH/QHS H)/DH MOI/ E)STIN, W)= *SW/KRATES, DIA/ TO POL-
    LAKIS DEU=R O FOITA=N, KAI/ Ti KAI/ EU)ERGE/HTAI U(P E)MOU=.
    @ *S*JF. *ARTI DE/ H(KEIS H) PA/LAI;
    @ *K*.R. *JPEIEIKW=S PA/LAI. (10)
(b) @ *S*JF. *EI)=TA PW=S OU)K EU)QU S E)PH/GEIRA/S ME, A)JLLA/ SIGH-
    KA/QHSAI;
    @ *K*.R. *OU/MA/ TO/N *DIA, W)= *SW/KRATES, OU)D A)N AU)/TO S H)
    E)N TOSAU/TH TE A)GRUPNI/A KAI/ LU/PH/ EI)=NAIL, A)JLLA/ KAI/ SOU=1
    QAUMA/ZW AISQANO/MENOS W(S H(DE/WS KAOEU/DEIS KAI/ E)PI/TH
    OU)K) H)/GEIRON I/NA W(S H/DISTA DIA/LHSE/ KAI/ POLLAKIS M E N DE
    KAI/ PRO/TERON E)N PANTI/ TW= BI W, HU)/DAIMO/NISA TOU= TRO/POI
    KAI/ PRA/WS FE/REIS. @1
    @ *S*JF. *KAI/ GA[R A)N, W)= *KRI/TWN, PLHMMELE/S EI)/H A)GANAKT.
    THLIKOU=TON O)N TA) E)I) DEI=H) DHI TELEUTA=N.
43. (a) ΣΩ. Τι τηνακάδε ἀφίξη, ὢ Κρίτων; ἢ οὐ πρὸ ἐπὶ ἄτιν; ΚΡ. Πάνω μὲν οὖν; ΣΩ. Περικά μαλάτα; ΚΡ. Ὀρθὴς σχεσίς; ΣΩ. Θεαμάζομαι ὧς ἠθέλησεν οἱ ἀδελφοί σου ἐπιστρέψασθαι. ΚΡ. Συνήθης ἦδη μια ἐστών ὡς τόθεν ἄντον; διὰ μᾶλλον ψευδώνοις ἀφιέρωσαν, καὶ τι καὶ ἐγώ. ΣΩ. Αἱ σπάνιοι τῆς ἡμέρας; ΚΡ. Επειδή τοῦτο πάλαι. (10) (b) ΣΩ. Εἴτε πῶς σὰς εὐθὺς ἐπηγείρισες, οὐκ ἔγνως κάθησιν; ΚΡ. Οἱ μὲ τὸν Δία, ὡς Ἑκάστρες, χαίρετε ἐν τοιαύτῃ τῇ ἄγραφῳ καὶ λύπῃ εἰς τὸν θαμμένον αἰσθάνομεν ὡς ἤρέσσατο καὶ οὐκ ξέρετε ὧν ἐν ᾧ ἤλειθεν ἄγαθον; καὶ πρὸς ἐν αὐτῷ τὸ βίον κρατοῦσαν, τὰ δὲ μάλλον ἐν τῇ νυν παραπάσχει σαρκικὸς καὶ καταφέρεις. ΣΩ. Καὶ γὰρ ἂν ὃς Κρίτων, πλημμέλως σφυκτόν ἄκοιμητος, ὅτι πέφυκεν ὁ ἐν αὐτῷ ἡ ἀνθρώπην ἀγαθόν εἰ ἐν ἀρνητικὸν τὴν περιπάτησιν.
Possibilities of the TLG
TLG: Functions

1. Canon search

2. Text browsing; comparison of two texts

3. Text search

4. Linking to external resources: lexica, online databases, online translations, morphological annotation etc.

5. Visualization (of location, of statistical information from the corpus)

6. Learning tools
TLG: Visualization

- There are 4,270 words in this work corresponding to 1,467 unique forms and 804 lemmata.
- The most frequent wordforms are: ο, σωκράτες, κρίτων (more).
- The most frequent lemmata are: λέγω, Σώκας, -σος, ο, πολύς, πολλά, πολύ (more).
- The most over-represented lemmata in this work are: Σώκας, -σος, ο, σωκράτης, -ίως, ὁ, δοκεῖ (more).
- The most under-represented lemmata in this work are: θεός, -ος, ὦ, γίγνομαι, εαυτοῦ, -ης, -ο (more).
- There are 0 unique occurrences in this work (more).
TLG: Parallel Browsing

**PARALLEL BROWSING**

PLATO (0059)

Phaedo (004)

**NOVUM TESTAMENTUM (0031)**

Evangelium secundum Marcum (002)

---

**Text Structure**

PLATO, Phaedo. (0059,004)

57.

(a) **ΕΥ.** Αὐτὸς ὁ Φαῖδος, παρεγένον Σκίττον ἐκείνη τῇ ἡμέρᾳ ἢ τὸ φαινόμενον ἔπειν ἐν τῷ δεσμητηρίῳ, ἢ ἄλλον τοῦ ἱκουόνος.

ΦΑΙΔ. Αὐτὸς ὁ Σκίττον ὁ Χάνταμος.

ΕΥ. Τί ὁ σύ ἀν δὲ οὕτω ἂν εἴην ὁ ἀνήρ πρὸ τοῦ θανάτου; καὶ πῶς, εἰπεῖτε; ἤδεις γὰρ ἐν ἐγώ ἀκούσαιμι. καὶ τὰς ὁπίσθιας παῦει, τὰσ ἐπιγραφεῖς τὰ νὰν Ἴθινα, ὅτα τὰς ἔχεις ἀρίστας χρόνου συνογώ.

(b) ἔκειτο δύο ἡμῖν ὁ μήπω ἐργάζει, οὗτος τὰ περὶ τούτων, καὶ ἤδεις ἀν δὲ ὁ πολλοί πῶς ἀπαθnutrition": 

58.

(a) **ΦΑΙΔ.** Οὐδὲ τὰ περὶ τῆς δίκης ἄρα ἔπευξεν ὅτι τὸν ποιεῖν.
1. access to and search in the bibliographical information
2. compilation of a list of authors / works where to search
# TLG: Metadata Structure

<table>
<thead>
<tr>
<th>author work</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>canon number</td>
<td>obl.</td>
</tr>
<tr>
<td>**name</td>
<td>title**</td>
</tr>
<tr>
<td>further names</td>
<td>opt.</td>
</tr>
<tr>
<td>date</td>
<td>obl.</td>
</tr>
<tr>
<td>geographic epithet</td>
<td>opt.</td>
</tr>
<tr>
<td>generic epithet(s)</td>
<td>opt.</td>
</tr>
<tr>
<td>non-standard generic epithet(s)</td>
<td>opt.</td>
</tr>
<tr>
<td>links to other authors or works (&quot;cf.&quot;)</td>
<td>opt.</td>
</tr>
<tr>
<td>links to Scholia, Vita(e), Lexicogr. in other authors or works</td>
<td>opt.</td>
</tr>
<tr>
<td>publication</td>
<td>NA</td>
</tr>
<tr>
<td>edition type</td>
<td>NA</td>
</tr>
<tr>
<td>language</td>
<td>NA</td>
</tr>
<tr>
<td>word count</td>
<td>(=works)</td>
</tr>
</tbody>
</table>
TLG: Text Search
TLG: Text Search Possibilities

1. Textual search
go through the actual text files, character by character, to find an instance of the string!! Very slow

2. Word index search
uses a precompiled index of all full and partial Greek words in the corpus

3. Lemmatized word index search
uses a precompiled and lemmatized word index of all Greek words in the corpus
1. Subcorpus to search
   - whole corpus
   - a list of authors
   - a list of works

2. Wildcard search (or regular expressions, http://stephanus.tlg.uci.edu/helppdf/wildcard.pdf)
   - alternatives: ANADU|ANEDU searches for ANADU or ANEDU
   - character classes: AN[AE]DU searches for ANADU or ANEDU
   - word boundaries: ^ANA searches for ANA as a prefix (word index)

3. Boolean / proximity search:
   Word A and word B (after 5 lines) except word C
   = A and B (within five lines after A) and C (within five lines either before or after B)
CASE STUDY № 1 :
THE CONCEPT OF GOUT
Preliminaries

  – whether there existed a dedicated language of medicine in Greek, and if so,
  – how was it designed.

• The second goal is to analyze the concept of gout along the following aspects:
  – different lexemes, parts of speech, text genres, authors, historical periods, pragmatic factors, etc.
Preliminaries

• The concept of *gout* is chosen purposly because its lexemes do not exhibit any polysemy.

• Probably, the most frequently cited chronicle disease of the Classical Antiquity (Stamatu 2005: 356).
Preliminaries

• The number of occurrences in the corpus (appr. 1500) is statistically significant; though still all occurrences can be treated individually.
• The lexemes are attested from 5 BC until 16 AD. A diachronic analysis of the concept is possible.

• Nearly 70% of the occurrences are attested in texts which can be characterized as medical (medical treatises, collections of medical prescriptions, etc.)
• The other 30% occur in clearly non-medical texts.
• Therefore the differences between the medical and non-medical usage of the concept can be determined.
Gout: lexemes

- 21 lexemes denoting the concept of gout:
  - 7 nouns: ἀρθρῖτις, ποδάγρα, ποδαγρία, ποδάλγη, ποδαλγία, χειράγρα and χειραλγία;
  - 9 adjectives: ἀρθριτικός, ποδαγρικός, ποδαγρός, ποδαλγικός, ποδαλγός, ποδαλγής, χειραγρικός, χειραγρός and χειραλγός;
  - 5 verbs: ποδαγράω, ποδαγριάω, ποδαλγέω, ποδαλγιάω and χειραγρέω

- What are the differences between them?
TLG search

• ποδα[λγ][γρ] | αρθρ[ιη]τ | χειρα[λγ][γρ]
  (textual search, 1571 matches)

• ποδα [λγ] [γρ] | αρθρ [ιη] τ | χειρα [λγ] [γρ]
  (word index, 1576 matches)
# Gout: data base

<table>
<thead>
<tr>
<th>token</th>
<th>ἀρθρίτιδα</th>
<th>ποδαγρικαῖς</th>
</tr>
</thead>
<tbody>
<tr>
<td>stem</td>
<td>ARTHRIT</td>
<td>PODAGR</td>
</tr>
<tr>
<td>POS</td>
<td>substantive</td>
<td>adjective</td>
</tr>
<tr>
<td>lexeme</td>
<td>ἀρθρῖτις</td>
<td>ποδαγρικός</td>
</tr>
<tr>
<td>time</td>
<td>Middle Byz.</td>
<td>Rom.</td>
</tr>
<tr>
<td>type of context</td>
<td>medical texts</td>
<td>medical texts</td>
</tr>
<tr>
<td>author</td>
<td>Hippiatrica</td>
<td>Galenus</td>
</tr>
<tr>
<td>work</td>
<td><em>Hippiatrica Cantabrigiensia</em></td>
<td><em>In Hippocratis aphorismos commentarii vii</em></td>
</tr>
<tr>
<td>TLG-Number</td>
<td>tlg:0738.006</td>
<td>tlg:0057.092</td>
</tr>
<tr>
<td>citation</td>
<td>2,15,1</td>
<td>18a,84,3</td>
</tr>
<tr>
<td>text</td>
<td>Διάρρινον πρὸς μάλιν ἀρθρίτιδα Ἰπποκράτους.</td>
<td>τῆς ἔξωθεν ὑγρότητος, ὅντα πυκνά καὶ σκληρὰ τὰ τοιαύτα μόρια, τῷ τοιούτῳ λόγῳ καὶ μεθίησι μόγις. διὰ τοῦτο οὖν καὶ ταῖς ποδαγρικαῖς φλεγμοναῖς ὁ Ἰπποκράτης ὁρὸν ἔθετο λύσεως παντελούς οὐ τὴν τεσσαρεσκαιδεκάτην ἡμέραν, ἀλλὰ τὴν τεσσαρακοστήν.</td>
</tr>
</tbody>
</table>
Gout: data base

<table>
<thead>
<tr>
<th>(a) word-level annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>parts of speech</td>
</tr>
<tr>
<td>noun</td>
</tr>
<tr>
<td>adjective</td>
</tr>
<tr>
<td>verb</td>
</tr>
<tr>
<td>lexemes</td>
</tr>
<tr>
<td>(21)</td>
</tr>
<tr>
<td>stems</td>
</tr>
<tr>
<td>ARTHRIT</td>
</tr>
<tr>
<td>PODAGR</td>
</tr>
<tr>
<td>PODALG</td>
</tr>
<tr>
<td>CHEIRAGR</td>
</tr>
<tr>
<td>CHEIRALG</td>
</tr>
</tbody>
</table>
# Gout: results for medical texts

<table>
<thead>
<tr>
<th></th>
<th>adjective</th>
<th>substantive</th>
<th>verb</th>
<th>PODAGR</th>
<th>ARTHRIT</th>
<th>PODALG</th>
<th>CHEIRAGR</th>
<th>CHEIRALG</th>
</tr>
</thead>
<tbody>
<tr>
<td>medical texts</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>historiographical texts</td>
<td>-</td>
<td>+</td>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>theological context</td>
<td>-</td>
<td>+</td>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Historia Naturalia</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>language descriptions</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comica</td>
<td>-</td>
<td></td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>philosophy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>astrology</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>biographies</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Cluster „Medical Texts“**

**Lemma:**
- + favored (standardised residual > 2)
- - dispreferred (standardised residual < 2)
### Gout: results for non-medical texts

<table>
<thead>
<tr>
<th></th>
<th>adjective</th>
<th>substantive</th>
<th>verb</th>
<th>PODAGR</th>
<th>ARTHRIT</th>
<th>PODALG</th>
<th>CHEIRAGR</th>
<th>CHEIRALG</th>
</tr>
</thead>
<tbody>
<tr>
<td>medical texts</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>historiographical texts</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>theological context</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><em>Historia Naturalia</em></td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>language descriptions</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td><em>Comica</em></td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>philosophy</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>astrology</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>biographies</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Cluster „Vitae“**

- + favored (standardised residual > 2)
- - dispreferred (standardised residual < 2)

**Lemma:**

- + favored (standardised residual > 2)
- - dispreferred (standardised residual < 2)
„Vitae“: PODALG as the disease of the emperors

Results: overview of the concept

PODAGR
- attested throughout the corpus
- 55% of attestations (866)
- mainly ποδάγρα and ποδαγρικός
- semantically neutral designation of the disease

ARTHRT
- attested throughout the corpus
- 31% of attestations (492)
- mainly ἄρθριτικός and ἄρθρητις, no verb
- mainly in medical texts

PODALG
- attested from the Hellenistic period
- esp. in historiographical and theological types of context
- 11% of attestations (170)
- i.e., as a disease of the emperors

CHEIRAGR
- attested from the Roman period
- esp. in astrology
- 2% of attestations (34)

CHEIRALG
- attested from the Early Byzantine period
- used as counterpart to PODALG
- 0.5% of attestations (8)
CASE STUDY № 2 : PREPOSITIONS
Prepositions by periods
Why corpus/usage-based method?

- The transition from the classical language to early Koiné and then further developments into NT Koiné are gradual;
- the language of the early Hellenistic period is still very similar to the Classical language because:
  (a) the semantics and formal properties of grammatical categories do not change abruptly;
  (b) and speakers often have the choice to employ a more conservative meaning/construction given the literary tradition (e.g. in order to yield specific pragmatic effects).
<table>
<thead>
<tr>
<th>Rank</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Homerus, Hesiodus, Aeschylus</td>
</tr>
<tr>
<td>2</td>
<td>Thucydides, Lysias, Aeschines, Isocrates, Demosthenes, Isaeus, Xenophon, Gorgias, Plato, Aristophanes</td>
</tr>
<tr>
<td>3</td>
<td>Polybius, Archimedes, Diodorus, Nicolaus Damascenus, Chrysippus, Menander</td>
</tr>
<tr>
<td>4</td>
<td>Longus, Appian, Arrianus, Josephus Flavius, Plutarchus, Heliodorus, Lucianus, Flavius Philostratus, Dio Chrysostomus, Chariton</td>
</tr>
<tr>
<td>5</td>
<td>Novum Testamentum</td>
</tr>
<tr>
<td>6</td>
<td>Ioannes Malalas, Ioannes Antiochenus</td>
</tr>
</tbody>
</table>

**Diagram:**

- Title: Prepositions by authors
- Y-axis: Authors
- X-axis: Frequency per 1,000 words
- Points plotted on the graph represent the frequency of prepositions used by each author.
THANK YOU!!