



# Papyrological texts and linguistic research

## SunoikisisDC Digital Classics: Session 9

*Marja Vierros*

&

*Polina Yordanova-Nanev*



UNIVERSITY OF HELSINKI  
FACULTY OF ARTS

March 20, 2025 16:00-17:30 GMT

# Outline

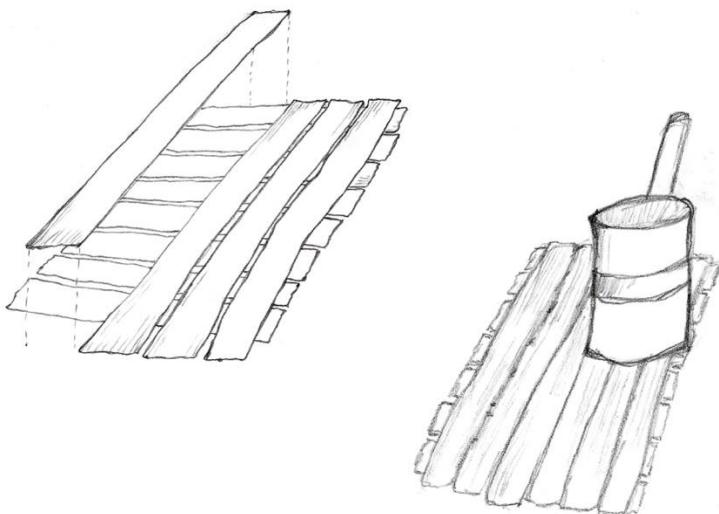
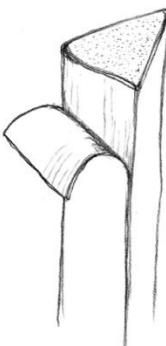
---

- Introduction
  - What are papyri, what is papyrology?
  - Why study Greek language through papyrological sources?
  - Case study 1 with PapyGreek Search on orthography > phonology
- Treebanking (i.e. linguistic annotation)
  - what is treebanking?
  - case study 2 on morphosyntax with PapyGreek Search
  - case study 3 on word order with KTB
- Other projects and tools
- Exercise



# Papyri?

- *Cyperus papyrus* (papyrus sedge)
- was manufactured into a lightweight writing material





Complaint about an attack in a bath house

P.Hels. I 2 (195–192 BCE)

<http://papyri.info/ddbdp/p.hels;1;2>

## Papyri – important primary written source

- widespread writing material in antiquity
- survived mostly from Egypt
  - Greek texts from about c. 300BCE–600CE
- documentary and literary texts
- documentary texts include, e.g.
  - letters
  - administrative texts (contracts, receipts and accounts, etc.)
- great material for studying the **use and development of the Greek language**



# Papyri as data



documentary papyrus texts have been edited and published since 19th century



these texts were made electronically available in late 20th century (CD-ROM)



The Duke Databank of Documentary Papyri (DDbDP) is now available and queryable in **papyri.info** aka *The Papyrological Navigator*



ca. 50 000 documents (5M tokens) with digital transcription +editorial markup in TEI EpiDoc XML

# what can the *papyri.info* query interface do?

Papyri.info

string searches (strings of characters), i.e.

- parts of words
- specific forms of words or phrases
- proximity of several different strings

lemmatised searches

- all declined forms of a certain word

Boolean operators and regular expressions give a boost

filtering by date, place of origin, language, etc.

Browse:

DDbDP HGV APIS DCLP

Authors TM Number

or Search:

Data Bibliography

## Refine Search

[New Search](#)[Search](#)

within  chars

and or not then near lex  -

- Convert from betacode as you type  
 ignore capitalization  
 ignore diacritics/accents

Text  Metadata  Translations

Series  Vol.   
or

Collection  ID #

Author

Work

Substring: οικονομ

[X](#)

Target: text

No Caps: On

No Marks: On

1168 hits.

Identifier	Title	Locat
<a href="#">aegyptus 96</a> 10	Frammento ...	Fundort: Theadelphia Ägypten); Se Krokodilopolis (Ars
	...περί τε οἰκ[ο]νομιῶ[v]	
<a href="#">analpap 21 22</a> 79	Inizio ...	unbekannte
<a href="#">analpap 26</a> 377_2	Ordine ...	Antinoopolis (E Ägypt
	...οἴκο(νόμος)-	
<a href="#">apf 53 168 2</a>	P.Ross. ...	Arsinoiton Poli Ägypten) Arsi

Browse: DDbDP HGV APIS DCLP Authors TM Number

or Search: Data Bibliography

## Refine Search

New Search

Search

within  
chars

and or not then near lex  
regex abbr start-not end-not

clear -

- Convert from betacode as you type
- ignore capitalization
- ignore diacritics/accents

Text  Metadata  Translations

Series --- All values ---

Vol.

Collection --- All values ---

ID #

Author

--- All values ---

Work

--- All values ---

Substring: οικ  
Target: text  
No Caps: On  
No Marks: On

1168 hits.

### Identifier

aegyptus 96

10

...περί τε οἰκ

analpap 21 22

79

analpap 26

377.2

...οἰκο(νόμος)

apf 53 168 2

Lemma: οἰκονόμος

X

Target: text

No Caps: On

No Marks: On

314 hits.

Identifier	Title	Location	Date
<a href="#">apf 53 168 2</a>	P.Ross. ...	Arsinoiton Polis (Arsinoites, Ägypten) Arsinoiton Polis	630 CE
	...οἰκονόμω τῆς		
<a href="#">basp 48 103</a>	Receipt ...	Oxyrhynchos	587 CE - 588 CE
	...Φεὶβ οἰκονόμος		
	...Φὶβ οἰκονόμου		
<a href="#">basp 54 48</a>	New ...	Oxyrhynchos ? (Oxyrhynchites, Ägypten)	186 BCE
	...τῶν οἰκονόμων μήτε		
<a href="#">basp 58 157</a>	A ...	Oxyrhynchos (Oxyrhynchites)	527 CE
	5. [καὶ οἰκονόμου]		
	...οἰκονόμου		
<a href="#">bgu 1 102</a>	keiner	Soknopaiu Nesos (Arsinoites)	161 CE
	...Καίσαρος ο[ἰ]κονόμου(*)		
<a href="#">bgu 1 311</a>	keiner	Theognis (Arsinoites)	601 CE -

# what can the *papyri.info* query interface NOT do?

1

find grammatical  
forms

2

find syntactic  
relations

3

target to variation  
between original  
spelling and  
regularised spelling



<https://papygreek.com/>

A platform for the linguistic study of Greek papyri,  
including a **grammar**, **annotated texts**, and a **search tool**.

PapyGreek Treebanks Texts People Grammar Search ADMIN

Note: this is a pre-release version of the grammar. [Sign out](#) to see the current release.

### A Grammar of Documentary Greek Papyri

Parent chapter: - Number: 0

[SAVE](#) [DELETE](#) [NEW CHAPTER](#)

#### Welcome!

Welcome to the Digital Grammar of Greek Documentary Papyri. This is a work produced by the team of the [PapyGreek project](#) at the University of Helsinki. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 758481).

About the Grammar and the corpus

The manual treebank annotation in PapyGreek corpus mostly follows Celano (2014c) for the morphological and syntactic layer (the advanced syntactic/semantic layer is not present). Some special practices, differences and deviations can be found in [this document](#). The most notable differences are:

- numerals are all marked as numeral (instead of adjectives)
- pronouns are all marked as pronouns (instead of adjectives)

About the queries and trees

- layer differences, orig/reg
- saved searches
- Embedded trees

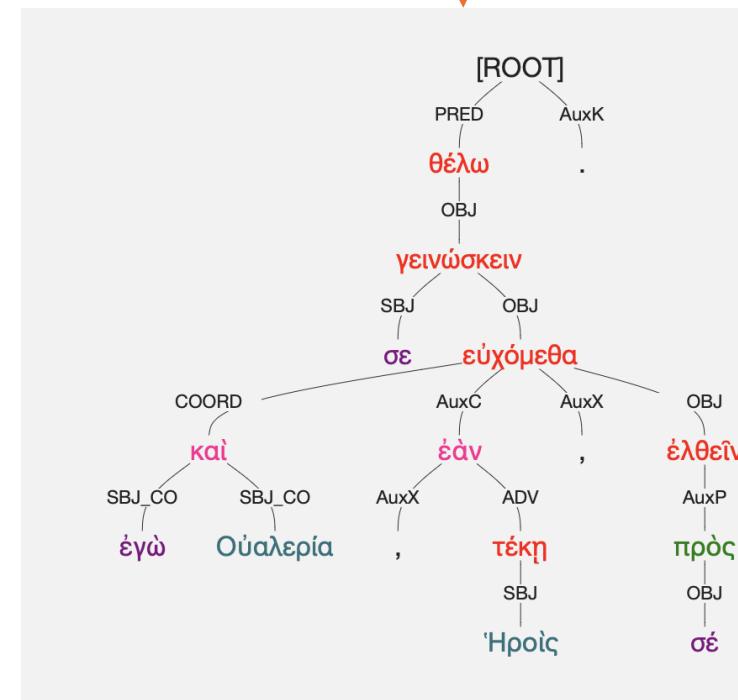
How to cite

Vierros, Marja, Sonja Dahlgren, Erik Henriksson, Polina Yordanova, and Sari Kock, eds. 2023. Digital Grammar of Greek Documentary Papyri. Helsinki: University of Helsinki. Zenodo.

<https://doi.org/10.5281/zenodo.XXXX>  
(Available online at <https://papygreek.com/grammar>, Accessed on 2023-26-01).

(changes possible, so access date important; stable versions?)

Vierros et al. (2023)



Search ?

SAVED SEARCHES

Original Regularized

relation=COORD,lemma\_plain=δε

relation=PRED%

relation=Aux%

\*

PapyGreek Treebanks 3.0:  
<https://zenodo.org/records/8428823>



European Research Council  
Established by the European Commission

# PapyGreek Search

- combines several (linguistic) levels with different size of data:
  - editorial interventions (based on the EpiDoc XML)
    - → orthography → phonological research
  - morphology
    - manually annotated + automatically analysed (machine learning)
  - syntax
    - manually annotated treebanks
    - dependency structures and word order, too
  - contextual information
    - time, place (HGV data; [whole papyri.info](#))
    - writer, author, addressee, scribal official ([treebanked documents](#))



# Orthographic level

- EpiDoc <reg> and <orig> within <choice> results in three types of corrections (all from O.Claud.1.1):
  - ὅστρακον → ὅστρακον (osstrakon > ostrakon) **deletion**
  - δώσι → δώσει (dōsi > dōsei) **addition**
  - τοῦτω → τοῦτο (toutō > touto) **replacement**
- Multiple edits common within a word (e.g. λαμπάνιν → λαμβάνειν )
- Edit sequences (linked to word-id) stored in MySQL tables
  - can be queried also by context (surrounding letters, word-initial, word-final)

II spc Mons Claudianus

ῶς(\*) ἐν τις εὕρη  
τουτω(\*) τὸ ὄστρα-  
κον(\*), δώσι(\*) στρατη-  
ρα(\*).  
5 ἔρωσο(\*) .

## Apparatus

- Λ 1. l. ὕς
- Λ 2. l. τοῦτο
- Λ 2-3. l. ὄστρα |κον
- Λ 3. l. δώσει
- Λ 3-4. l. στρατῆ |ρα
- Λ 5. l. ἔρωσο

<b>id</b>	<b>word_id</b>	<b>operation</b>	<b>before</b>	<b>original</b>	<b>regularized</b>	<b>after</b>
1	1	replace		ῷ	ῷ	ῷ
2	5	replace	τοῦτ	ω	ο	
3	7	delete	ὅ	σ		στρακον
4	9	add	δώσ		ε	ῃ
5	10	delete	στ	ρ		ατῆρα
6	12	add	ἔρ		ρ	ωσο

# Case study 1

Orthographic queries:  
/o, u/ allophonic  
variation

Erik Henriksson, Sonja Dahlgren, Marja Vierros. 2025.  
*Phonological Variation in Greek Papyri: Two Case Studies Using PapyGreek Search.* In: Reggiani, N. (ed.) *Digital Papyrology*, vol. 3. De Gruyter.

# /o, u/ allophonic variation

## simplified background information

- Greek papyri from Egypt have a lot of variation between graphemes **ο**, **ω**, **ου** representing sounds /o, u/
- in Egyptian-Coptic /o, u/ had allophonic status (stressed o, unstressed u)
  - → in Greek texts the unstressed **ο**, **ω** /o/ was often replaced by **ου** /u/, and the stressed **ου** by **ο**, **ω**
  - Greek and Egyptian stress systems were different, and the Greek one was changing
- in Greek, case merger happening between the genitive and the dative cases
  - **-ω [o]** = dative singular, **-ου [u]** = genitive singular (in the 2nd declension group)

# /o, u/ allophonic variation

## PapyGreek Search

- orig=οὐ, reg= ω:
  - **1266** hits
- *Regular expressions* give more nuance: singular dative was sometimes written –not pronounced– with an *iota adscript*: ωι (replace 1266 with **1616**)
- orig=ω / ωι, reg οὐ:
  - **1335** hits
- in word final position:
  - **841** hits (63%)
  - <https://papygreek.com/search/167>
- the other direction: 70 % of instances were word final

form=-ου+ω

regex:form=-(ω|ωι)+ου

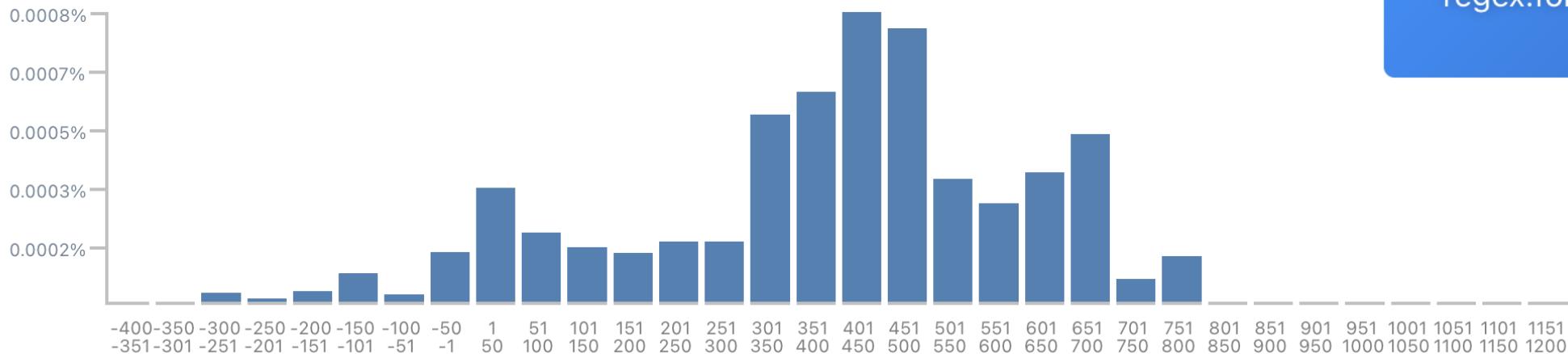
regex:form=-(ω|ωι)\$+ου\$<^\$

RELATIVE FREQUENCIES

ABSOLUTE FREQUENCIES



regex:form=-ou\$+(\$|wl)\$<^\$

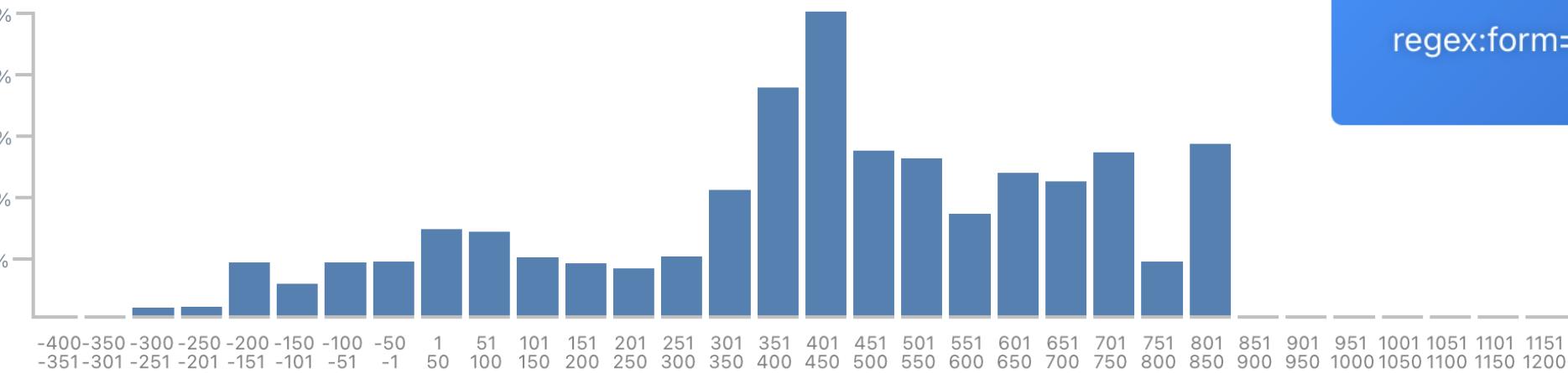


RELATIVE FREQUENCIES

ABSOLUTE FREQUENCIES



regex:form=-((\$|wl)\$)+ou\$<^\$



# Results?

- After several queries on different grapheme variations of the /o/ and /u/ sounds, the Coptic stress pattern can be seen behind many, if not most, of the instances found
  - happens in the case endings perhaps due to the word-final position being often unstressed (esp. in disyllabic words)
  - coarticulation (certain consonants affecting the vowel quality) is also one possible explanatory factor
- variation between /o/ and /u/ was relatively rare before the Roman period, when Coptic began to appear
- Egyptian phonological influence and Coptic orthographic practices potentially behind the variation in bilingual settings
- Results are still inconclusive, more statistical analyses needed

---

Sunoikisis Digital Classics - Spring 2025  
Session 9: Linguistics

# Papyrological texts and linguistic research

## Treebanking



Digital Grammar of  
Greek Documentary Papyri (PapyGreek)  
ERC Starting Grant, No. 758481



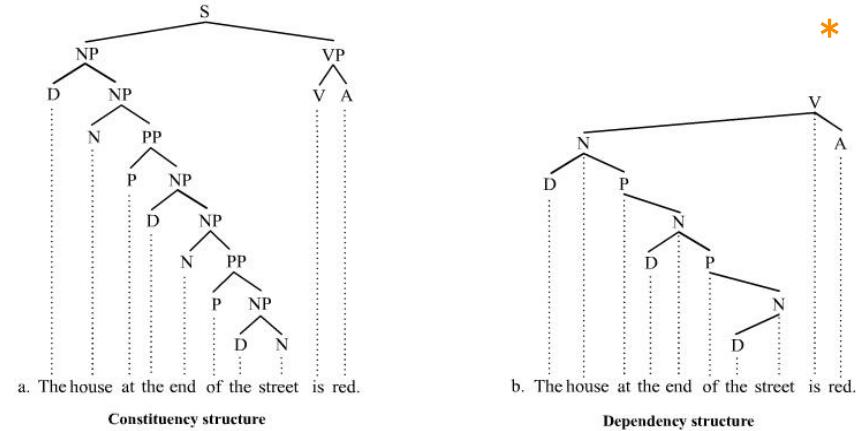
UNIVERSITY OF HELSINKI  
FACULTY OF ARTS



Polina Yordanova-Nanев  
[polina.yordanova@helsinki.fi](mailto:polina.yordanova@helsinki.fi)

# Introduction to Treebanking

- Linguistic analysis of a sentence in the form of hierarchical parse (syntax) tree;
- Constituency grammar:
  - focus is phrase structure
- Dependency grammar:
  - focus is on the individual word - better for languages with variable WO, easier to use in cross-linguistic comparisons, more efficient for dependency parsing tasks;
- Most Ancient Greek and Latin TB projects are dependency-based



\*Img source: Tjo3ya ([https://commons.wikimedia.org/wiki/File:The\\_house\\_at\\_the\\_end\\_of\\_the\\_street.jpg](https://commons.wikimedia.org/wiki/File:The_house_at_the_end_of_the_street.jpg))

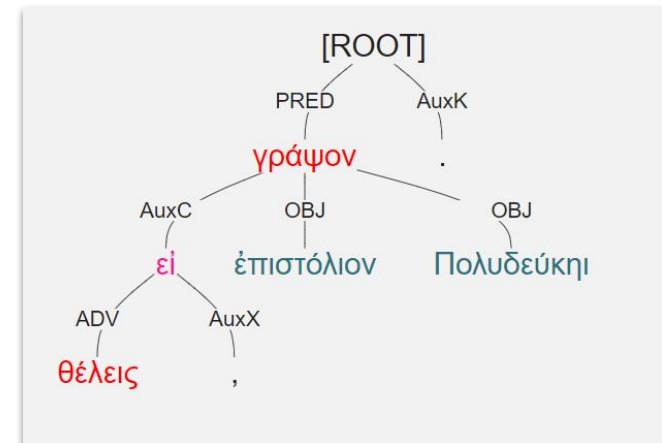


# Introduction to Treebanking

- “Treebank” - database of syntactic trees annotated according to an established formalism
  - Annotation can be done manually, automatically through parsing algorithms, or semi-automatically with human verification
  - Often in XML, but relational databases are also gaining popularity
  - Applications in computational linguistics, NLP, machine translation models etc.
-

# Introduction to Treebanking - Dependency TB

- Main finite verb (or a coordinator) is the head of the tree (ROOT);
- Head <-> dependent relationships between all other words;
  - “Head” = “parent” = “governor”;
  - “Dependent” = “child” = “descendent”;
  - “Siblings” - same parent node.
- Each word can have **only one head**;
- Words can have one, multiple or zero dependents;
- Relation labels indicate the syntactic function.





# Annotation

- Manual annotation is usually done through a user interface, e.g. Arethusa
- Information recorded as XML:

**ARETHUSA**

The screenshot shows a dependency tree for the sentence: Ισιδώρα Ἀσκληπιάδῃ τῷι ἀδελφῷι χαίρειν καὶ ύγιαίνειν διὰ παντός. The root node [ROOT] branches into PRED and AuxK. PRED further branches into SBJ (Ισιδώρα) and OBJ (Ἀσκληπιάδῃ). The OBJ node has two children: ATR (ἀδελφῷι) and COORD. The COORD node has two children: καὶ and OBJ\_CO (χαίρειν). The OBJ\_CO node has two children: ATR (χαίρειν) and OBJ\_CO (ύγιαίνειν). The OBJ\_CO node for ύγιαίνειν has three children: AuxP (διὰ), ADV (παντός), and the word παντός itself.

Ισιδώρα Ἀσκληπιάδῃ τῷι ἀδελφῷι χαίρειν καὶ ύγιαίνειν διὰ παντός.

```
<sentence id="1">
<word id="1" form="Ισιδώρα" lemma="Ισιδώρα" postag="n-s---fn-" relation="SBJ" wordid="242186" head="11"/>
<word id="2" form="Ἀσκληπιάδῃ" lemma="Ἀσκληπιάδης" postag="n-s---md-" relation="OBJ" wordid="242187" head="11"/>
<word id="3" form="τῷι" lemma="ό" postag="l-s---md-" relation="ATR" wordid="242188" head="4"/>
<word id="4" form="ἀδελφῷι" lemma="ἀδελφός" postag="n-s---md-" relation="ATR" wordid="242189" head="2"/>
<word id="5" form="χαίρειν" lemma="χαίρω" postag="v--pna---" relation="OBJ_CO" wordid="242190" head="6"/>
<word id="6" form="καὶ" lemma="καὶ" postag="c-----" relation="COORD" wordid="242191" head="11"/>
<word id="7" form="ύγιαίνειν" lemma="ύγιαίνω" postag="v--pna---" relation="OBJ_CO" wordid="242192" head="6"/>
<word id="8" form="διὰ" lemma="διά" postag="r-----" relation="AuxP" wordid="242193" head="7"/>
<word id="9" form="παντός" lemma="πᾶς" postag="a-s---mg-" relation="ADV" wordid="242194" head="8"/>
<word id="10" form="." lemma="." postag="u-----" relation="AuxK" wordid="242195" head="0"/>
<word id="11" insertion_id="0009e" artificial="elliptic" relation="PRED" form="[0]" head="0"/>
</sentence>
```



## relation: syntactic function

## postag: morphology

## head: immediate governor in tree hierarchy

```
<sentence id="1">
  <word id="1" form="Ισιδώρα" lemma="Ισιδώρα" postag="n-s---fn-" relation="SBJ" head="11"/>
  <word id="2" form="Ασκληπιάδη" lemma="Ασκληπιάδης" postag="n-s---md-" relation="OBJ" head="11"/>
  <word id="3" form="τῶι" lemma="ό" postag="l-s---md-" relation="ATR" head="4"/>
  <word id="4" form="ἀδελφῶι" lemma="ἀδελφός" postag="n-s---md-" relation="ATR" head="2"/>
  <word id="5" form="χαίρειν" lemma="χαίρω" postag="v--pna---" relation="OBJ_CO" head="6"/>
  <word id="6" form="καὶ" lemma="καί" postag="c-----" relation="COORD" head="11"/>
  <word id="7" form="ύγιαίνειν" lemma="ύγιαίνω" postag="v--pna---" relation="OBJ_CO" head="6"/>
  <word id="8" form="διὰ" lemma="διά" postag="r-----" relation="AuxP" head="7"/>
  <word id="9" form="παντός" lemma="πᾶς" postag="a-s---mg-" relation="ADV" head="8"/>
  <word id="10" form"." lemma"." postag="u-----" relation="AuxK" head="0"/>
  <word id="11" insertion_id="0009e" artificial="elliptic" relation="PRED" form="[0]" head="0"/>
</sentence>
```

# Case study 2

Henriksson, E. and Vierros, M. 2025. "PapyGreek Search: Exploring the Language of Greek Papyri."

In: Reggiani, N. ed. *Digital Papyrology III: The Digital Critical Edition of Greek Papyri: Issues, Projects, and Perspectives*. Berlin, Boston: De Gruyter, pp. 163-184.

Morphosyntactic queries:  
grammatical case discrepancy between the article and the noun?

# grammatical case discrepancy between the article and the noun?

- Continuing from case study 1: looking more closely to the genitive / dative case variation:
- the definite article (in masc+neut) followed the 2nd declension:  
τοῦ /tu/ = genitive; τῷ, τῷι /to/ = dative
- if the definite article and its head noun were both written in the “wrong case”, it can strengthen the argument of actual case merger, and if written differently, the phonological explanations
  - NB. many caveats here, but we will set them aside for now

# Queries

---

- e.g. to find cases where a **head is in singular genitive** and it **has an article with orthographic variation ω > ou** (the original writer wrote a dative ending but an editor regularised it into genitive because the article and the noun should agree in their case)
  - the singular definite article: “pos=l\_s%”
  - combine with spelling variation  
“-(ω|ωι)+ou”
  - (add restriction of word final position)

```
postag=__s____g%
```

```
postag=l_s%,regex:form=-(ω|ωι)+ou
```

Papyrus: article in the dative

ἐν τῷ οἴκῳ τῷ Ἀμμωνοῖ

the article is in the dative case

- due to analogy from the previous words?
- due to not knowing that a genitive was needed?
- due to similar pronunciation of the dative and genitive endings?

Regularised text: article in the genitive

ἐν τῷ οἴκῳ τοῦ Ἀμμωνοῖ

In this phrase, the genitive head declines according to the 3rd declension, so the genitive ending is -οῖ

# To make the study complete...

- one would need several queries for different scenarios
  - gen/dat vs. dat/gen
  - the head (also or only) has the orthographic problem
  - how do feminine and plural articles and nouns behave
  - include other words of the noun phrases (e.g. intervening adjectives)
- PapyGreek Search is capable of handling the queries
  - problem is that the treebanked data is small
    - the queries including head-dependent –relations only target the PapyGreek treebanked corpus
    - morphological and orthographic queries target the whole DDbDP

---

# Case study 3



# My thesis project

- The Greek of the papyri is different from literary Greek, but how so?  
Word order in the NP as a (relatively) narrow focal point;
- 5 corpora: 2 literary (prose and verse), 2 of documentary papyri (manually and automatically annotated one), and NT Greek (Koine);
- Beyond word order: challenges in using TBs for research and potential responsible approaches;
- Side quest - a querying tool / data management platform for treebanks



Developed with Jamie Norrish

Built on the basis of Kiln  
(which you have met as EFES)

Data management system

Additional multi-step annotation

Test Suite

Search Engine

Home WORD SEARCH SENTENCE SEARCH DOCUMENT SEARCH



## Search

### Word

Search word

Search word forms

### Current filters

Corpus: papygreek (x)

Column

### Facets

#### Metadata

+

#### Morphology

+

#### Syntax

+

#### Semantics

+

#### First Governor

+

#### First

+

#### Dependent

+

#### Other

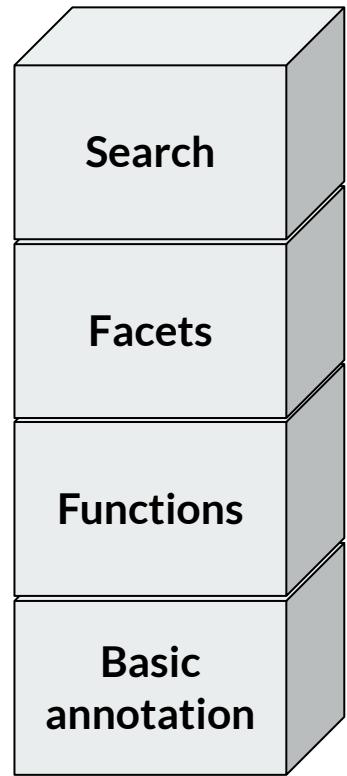
+

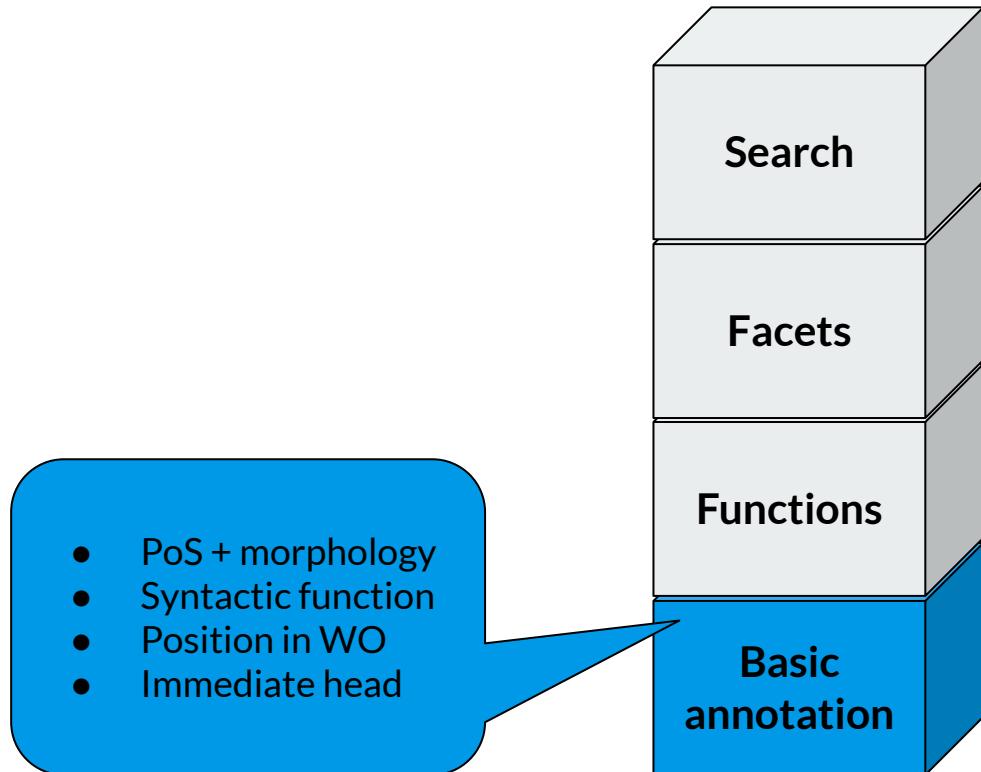
301 – 350 / 4485 of 4485 total rows

	Match	Sentence	Text	Genre	Century	Corpus
ἡσυχάζω	περὶ μὲν τούτων ἡσυχάζω ·	bgu.16.2625	Letter	1bc	papygreek	
Ἡρᾶτι	σύ αδες σύν τῇ ἀδελφῇ σοι ως ἔγραψες Ηρᾶτι τουνειν πιοσήσεν ήμῶν , ίνα διακονήσει ήμῶν .	bgu.1.261	Letter	2ad	papygreek	
Ἡρᾶς	Ἡρᾶς ὁ ἐπιστάτης και Ἀρτεμᾶς ὁ γραμματεὺς και Ἡρακλῆς τοῦ Ἐβρίου ἐντεύχεσαι ως ἐνυγκλουμένοι ὑπὸ τοῦ παρά σου πρός ἀριθμητικόν · [0]	bgu.16.2613	Letter	1bc	papygreek	
Ἡρᾶς	Ἡρᾶς δὲ παρῆσται τὰ ((τάλαντα)) ι δὰ τὸ Τυρ->άνιον μικώς μετριάζειν ίνα λέμψων παρ' αὐτού σχράματα ·	bgu.16.2646	Letter	1bc	papygreek	
Ἡροῖς	γιγνώσκειν σε θέλω ἐγώ και Ούαλερια έλαν Ηροίς τέρπειν ειναδιεθει έλασιν	bgu.1.261	Letter	2ad	papygreek	

# Demo of KTB



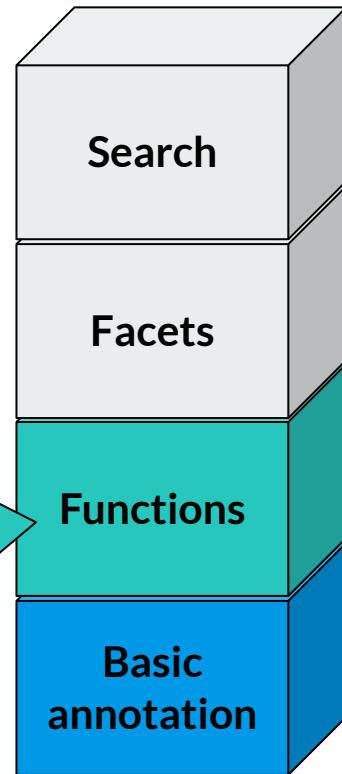




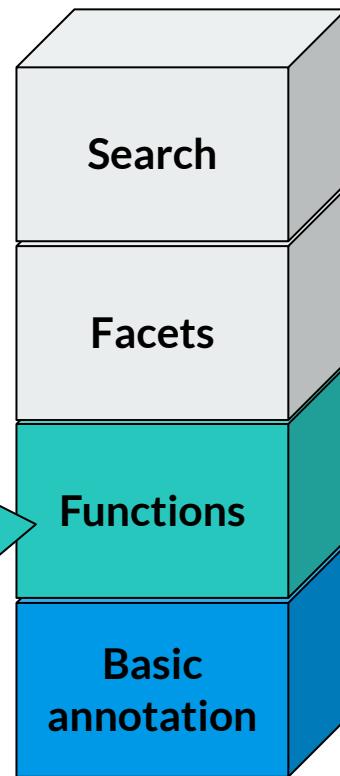
**“Shortcuts”:**

- “item is...”
  - Vowel / diphthong
  - Proper name
  - Patronym
  - ...
- Has article...
- Agrees with governor
- Dependency directly

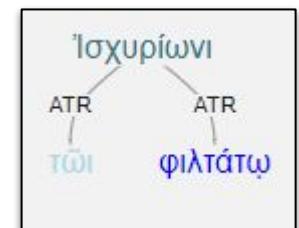
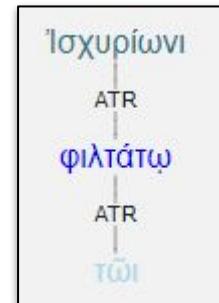
or  
solely through  
coordinators



- “Shortcuts”:**
- “item is...”
    - Vowel / diphthong
    - Proper name
    - Patronym
    - ...
  - Has article...
  - Agrees with governor
  - Dependency directly  
or  
solely through  
coordinators



**When does a word have an article?**





## What is a governor?

### “Shortcuts”:

- “item is...”
  - Vowel / diphthong
  - Proper name
  - Patronym
  - ...
- Has article...
- Agrees with governor
- Dependency directly

or

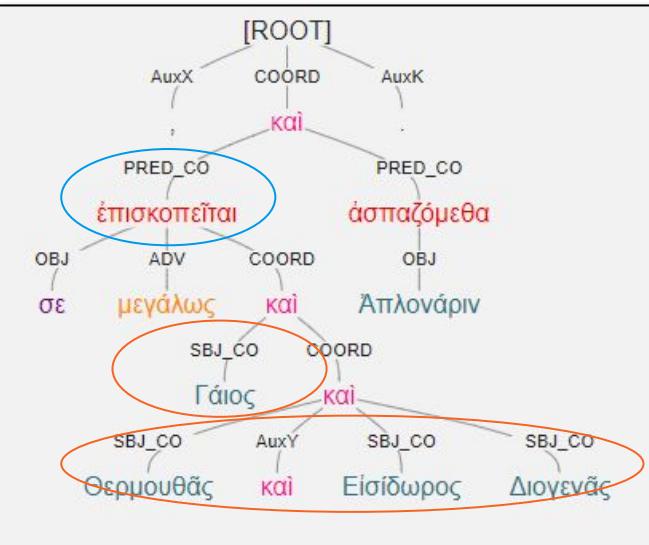
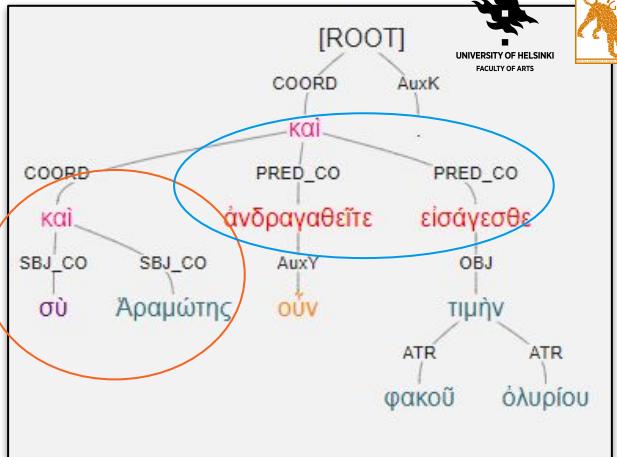
solely through  
coordinators

Search

Facets

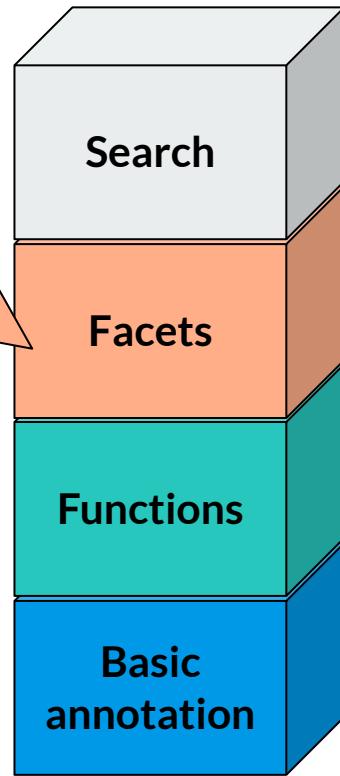
Functions

Basic  
annotation



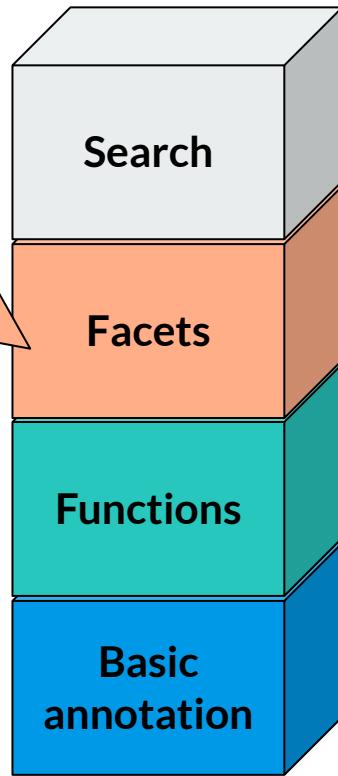
≈ 200 indexed fields

- Metadata:
  - Century, genre, author...
  - Language proficiency
- Sentence:
  - Basic word order
  - Subordinate clauses count/depth...
- Word:
  - Semantic type (PROIEL, AI-assisted analysis)
  - Is in greeting/dating formula
  - Governor pos/relation
  - ...



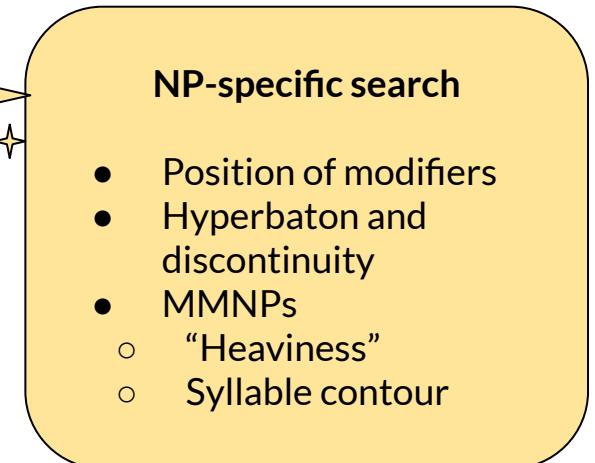
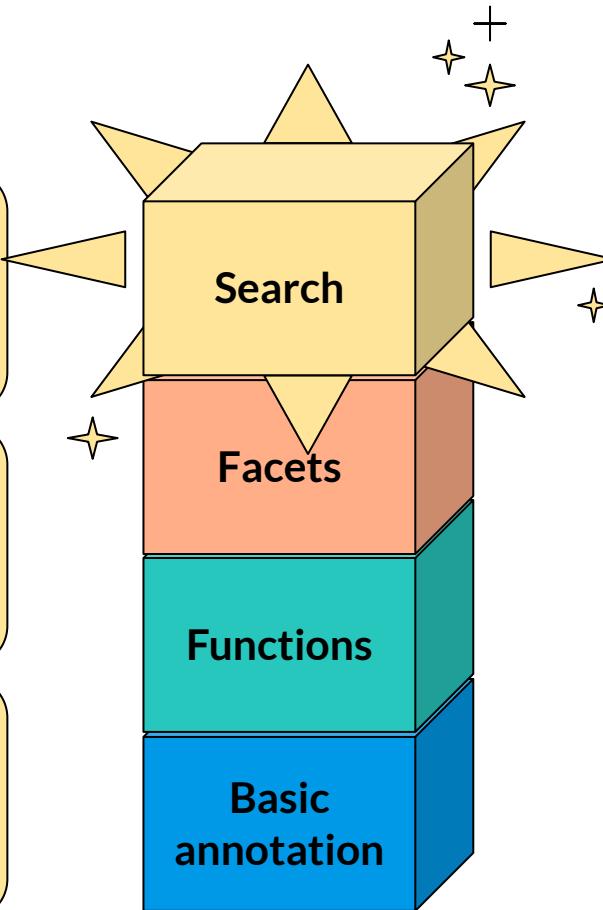
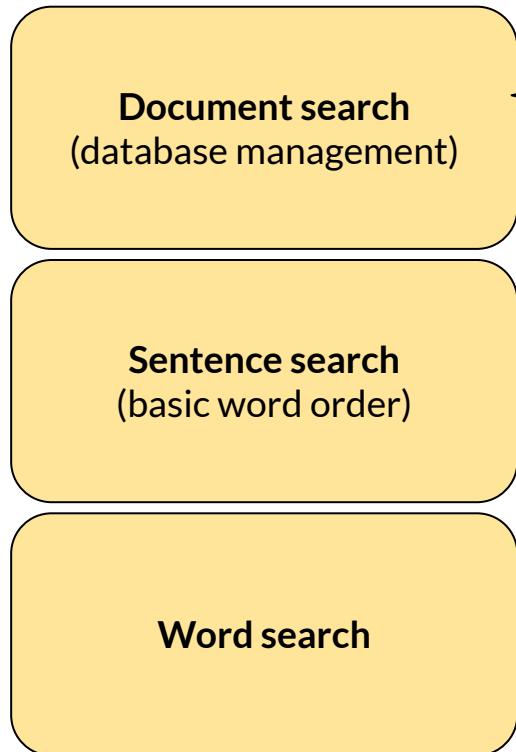
≈ 200 indexed fields

- Metadata:
  - Century, genre, author...
  - Language proficiency
- Sentence:
  - Basic word order
  - Subordinate clauses count/depth...
- Word:
  - Semantic type (PROIEL, AI-assisted analysis)
  - Is in greeting/dating formula
  - Governor pos/relation
  - ...



Fields are “stackable” and the values within each field can be OR-ed, where appropriate

Additional annotation for phenomena, the information for which must be retrieved from several different attributes





# Examples from my research - filtering of results

- Position of the adjective in the NP, but...
- What is an adjective?
  - “Good”, “happy”, “tall”...
  - “One”, “first”... ?
  - “This”, “that”... ?
- + / - quantifiers?
  - Overwhelmingly ubiquitous and stubbornly preposed in the NP
- Are there compounding factors for the position of the adjective?
  - PoS of the NP-head, other modifiers, etc.

## Current filters

Is Attributive: true (x)

Corpus: literature\_verse (x)

First Head PoS: n (x)

Is Quantifier: false (x)

PoS: Adjective (x)

Is Proper Name: false (x)

First Head Has Article: false (x)

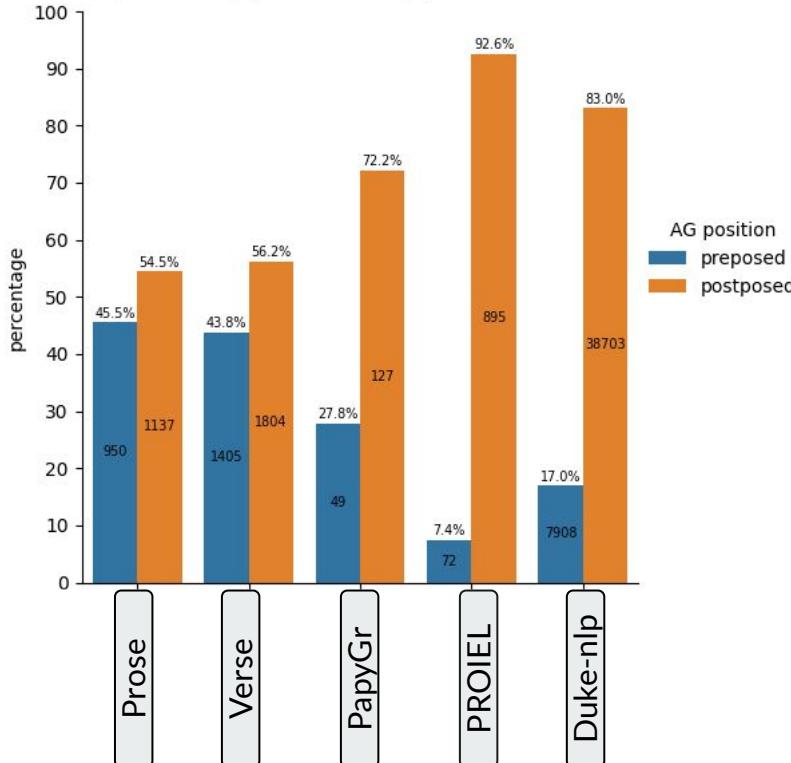
First Head Is Name: false (x)

First Head Has Only Adj Modifier: true (x)

Agrees With Governor: true (x)

### Position of the AG: non-articular NPs by corpus (%)

AG is starting point, only AG modifier, AG is noun, AG is not name, np-head is noun, np-head is not name, np-head is non-articular



# Some results from the NP - adnominal genitives

## Literature

Balanced distribution, a very slight preference for postposition in Prose

## PROIEL (New Testament)

Overwhelming postposition of the modifiers

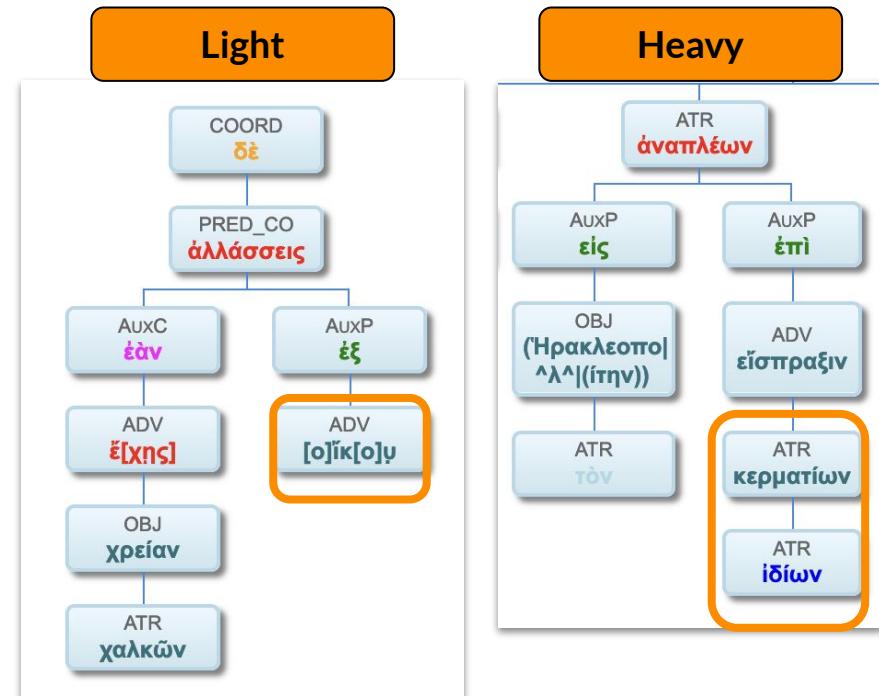
## Documentary corpora

Postposition is clearly preferred, more pronounced in Duke-nlp

\* additional constraints – both head and AG are nouns and not proper names, NP has no other modifiers

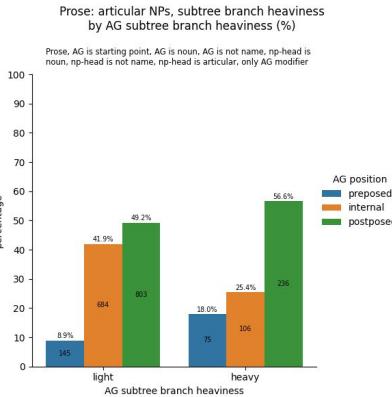
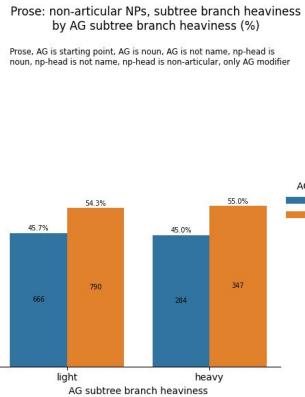
# Some results from the NP - adnominal genitives

- Definiteness, semantic class, and syntactic function all influence the AG's position in the NP in literature;
- In Papyri they don't seem to matter – postposition is universally favoured;
- Heaviness, however, is consistently influential across all corpora!

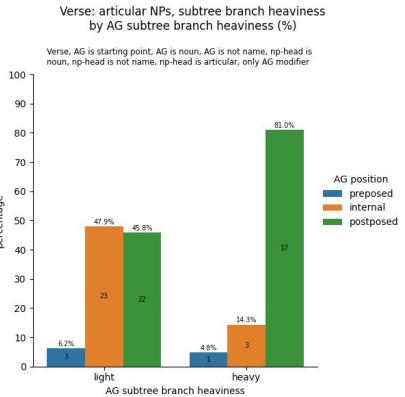
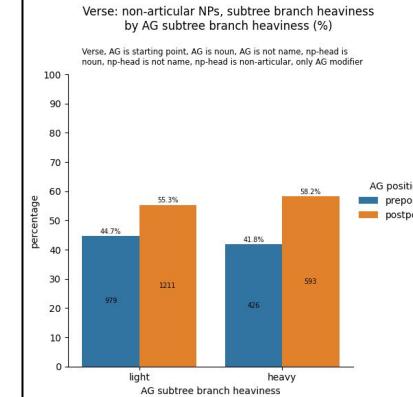


# Heaviness (syntactic constituents): Literary corpora

## Prose



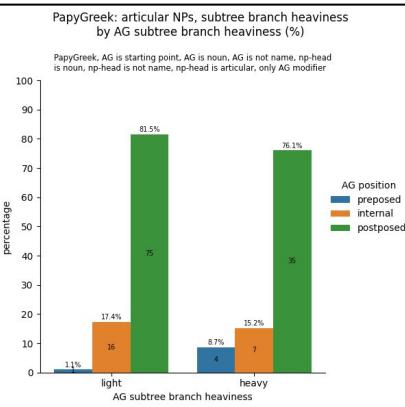
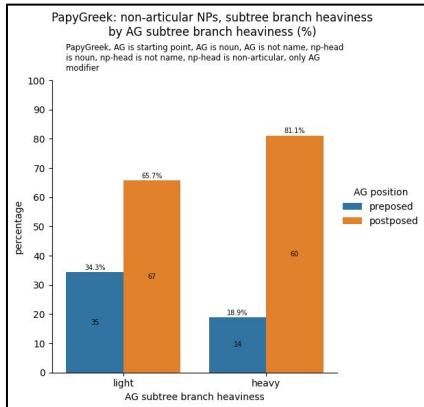
## Verse



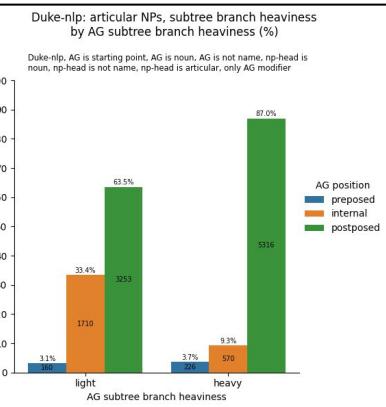
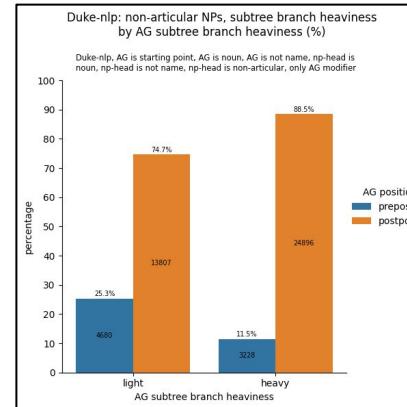
Statistically significant ( $p < 0.05$ ) for articular NPs

# Heaviness: Documentary corpora

## Papygreek



## Duke-nlp



Statistically significant ( $p < 0.05$ ) for both articular and non-articular NPs

---

# Experiments: proficiency annotation

Based on the criteria in Bagnall & Criboire “Women’s Letters”

## “Standard Greek”

- Morphology
  - Correct spelling
- Syntax
  - Subordination (depth and variety)
  - Use of optatives and middle voice
  - Constructions – articular infinitives, genitive absolutes
- Style
  - Competent and variable use of particles
  - Crasis

## “Nonstandard Greek”

- Mistakes in case and conjugation
- Phonetic spelling
- Iotacism
- Irregular iota adscripts
- Poor subordination
- Lack or excessive use of particles and connectives



# Key findings from the experiment:

- It (sort of) works ?!
- Non-standard texts have a higher proportion of postnominal modifiers;
- The phenomenon of polydefinite NPs is more likely to appear in texts from the non-standard group;
- Patterns with verbs which are held back towards the end of the sentence have a higher rate of appearance in standard texts than in non-standard ones in all datasets;
- “Verb-object” is more rigid in non-standard texts, variation is permissible in standard ones.

---

# Recap



# “So, which tool should I use?...”

## PapyGreek’s Search Tool

- All in one!
- User-friendly – no coding!
- Minimal familiarity with tb required

## KTB

- Many existing building blocks for fine-tuned linguistic research
- Possibility for customisation
- Bring in your own texts!



# Major TB projects



The Ancient Greek and Latin  
Dependency Treebank (AGLDT)



LiLa - Linking Latin



Universal Dependencies



Pedalion trees



Opera Graeca Adnotata (OGA)



GLAUx



# Universal Dependencies



- Distinct annotation scheme
- Over 200 treebanks in over 150 languages
- PML Tree Query – available as a web application
- INESS – *Norwegian Infrastructure for the Exploration of Syntax and Semantics*
  - Language independent platform for building, accessing, searching, and visualising treebanks through a web browser



# Pedalion



- **Pedalion trees**
  - Starting from automatically analysed Greek texts, then corrected manually;
  - Genres that are underrepresented in other TB collections, e.g. military texts, biblical tragedies etc.
- **Myria** and **Chilia** – treebank-based vocabulary tools
- **DendroSearch** – standalone search and browsing tool
- **The GLAUx corpus** – automatically annotated Greek literary texts, ca. 20M tokens, browser search





# LiLa: Linking Latin



- Collects and connects Latin linguistic resources (corpora, lexica, dictionaries etc.) and NLP tools (tokenisers, lemmatisers, PoS-taggers, morphological analysers and dependency parsers)
- Knowledge Base:
  - Triple store
  - Lemma Bank
  - Interactive Search Platform

# Opera Graeca Adnotata (OGA)



- The **largest** multilayer corpus of AG texts with automatically generated annotations:
  - 1999 texts;
  - Over 40 mil. tokens.
- Trained on 1.2M+ tokens of AGDT data, so its results are expected to be more reliable than other models;
- Opera Latina Adnotata in preparation!



---

**Thank you!**

---