

# Coptic SCRIptorluM

May 14, 2013 Workshop



Caroline T. Schroeder  
Amir Zeldes



# Research Objectives

Online, open source digital platform:

- For Corpus linguistics in Coptic
- Searchable text for historical and literary research
- Visualizations of Coptic texts and manuscripts
- Interoperable with current XML standards
- Collaborations with and contributions from other scholars

**Sahidic Corpus Research: Internet Platform for Interdisciplinary multilayer Methods**

# Coptic Scriptorium is...

A corpus of Coptic texts:

- a richly-annotated database using ANNIS database technology
- TEI-XML versions of the texts for download
- Visualizations (including diplomatic editions) in HTML and inside the database

Tools for digitally processing texts in Coptic

Documentation

A platform where scholars and students will be able to contribute texts and/or annotations and conduct research

***Completely open source & open access***

# Texts

Proof-of-concept corpora:

Shenoute, Abraham Our Father (Canons vol. 3)

Select Coptic Apophthegmata Patrum

They contain:

- Diplomatic & normalized editions
- Translation
- Annotation

*All of this is available online*

<http://go.pacific.edu/copticscriptorium>

# Tools: Encoding Converters

## What you give: ASCII

- 1) Basic LaserCoptic or CMCL ASCII font
- 2) Exhaustive “Coptic” font (by G. Wurst & D. Van Damme)

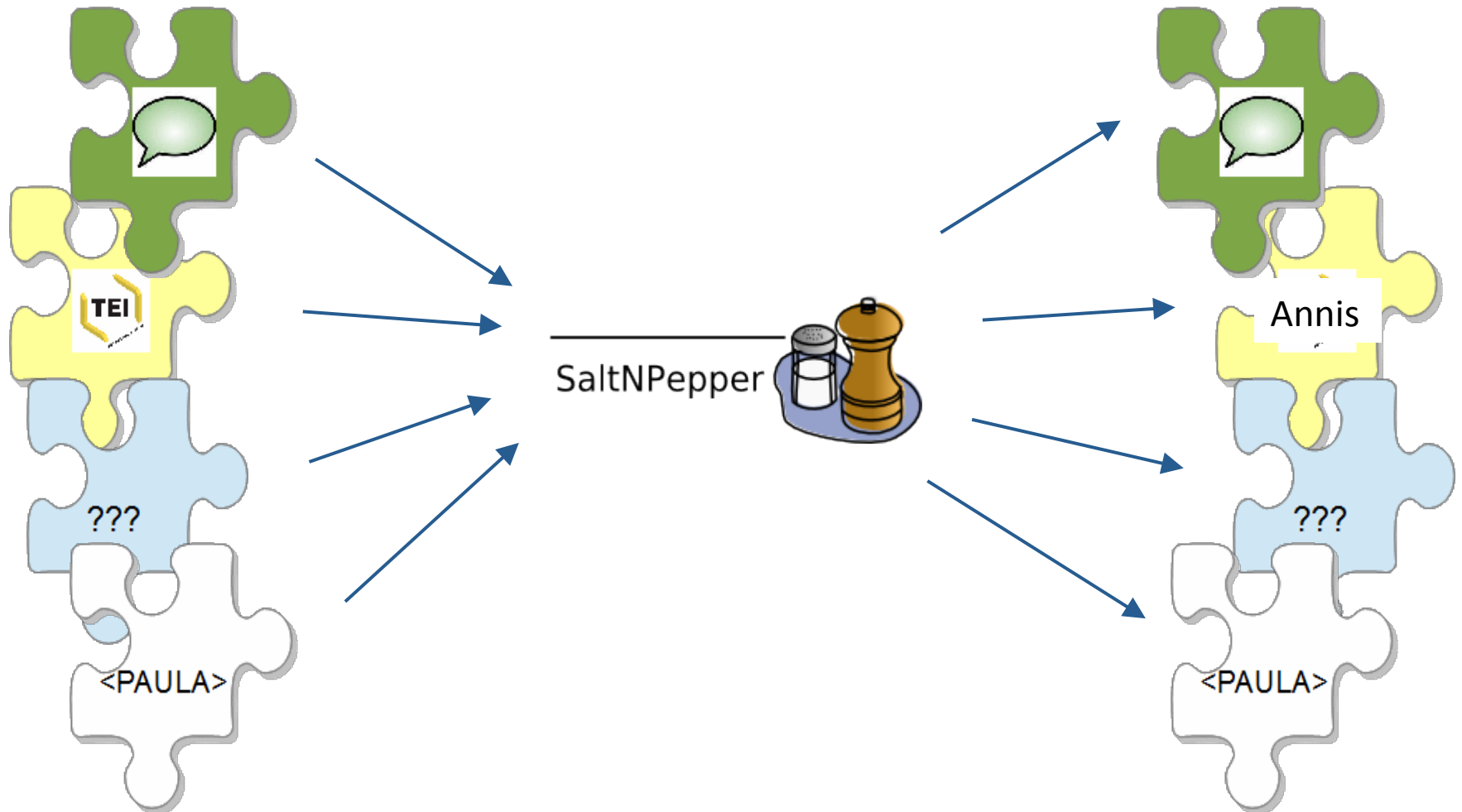
## What you get: UTF-8

Fully compliant Unicode Coptic

etbe néeere de éhm  
ntansvtm etbhhtoy

ΕΤΒΕ ΝΩΕΕΡΕ ΔΕ ΕΗΜ  
ΝΤΑΝCΩΤΜ ΕΤΒΗΗΤΟΥ

# File Format Converter (SaltNPepper)





# Tokenization Scripts & Lexicon

Scripts process files of Coptic text to break the words into morphs

Based on CMCL Lexicon provided by T. Orlandi

Tokenization scheme follows morphemes established in B. Layton's grammar

# Tokenization Script

What you give:  
text as words

ἀρχοὺς ἐτθε ἀπα παπνοῦτε ἄε μεφθεπн ἐσενрп̄ .  
ἐφμοοωε δε νογῶειω ἀφει ἐχῆ οὔκολλῶκιοη ἑλῶςτης  
ἀγω ἀφθεῖτοῦ ἐγσω . ἀφσοῦωνῆ δε ἥσι πετο ἥανχωχ  
ἀγω ἐφσοοῦη ἄε μεφσε ηрп̄ ἀφнаγ ἐροφ ρῆ οὔнос ἥρισε  
ἀφμοῦρ νογαпот ἥηрп̄ ἐρε τεφснφε ρῆ τεφсiχ ἀγω  
πεχαφ ἥπελλο ἄε ектнсω †назотбек . ἀφειне ἥσι  
πελλο ἄε ἐφοῦωω εεire νογῆτολη ἥτε πноῦτε ἐφοῦωω  
ἐ†ρнγ ἥноφ ἀφχιτῆ ἀφσοοφ . πανχωχ δε ἑλῶςτης  
ἀφμετανοι ἐφχω ἥнос ἄε κω ηαι εβολ παειωт ἄε  
αι†ρισε ηак . πεχαφ ηαφ ἥσι πελλο ἄε †πιστεγe  
епноῦτε ἄε φηαεire ηак ἥογηа ἐтθε πειαпот ἥηрп̄ ρῆ  
πειωηῆ ἥῆ πεтннγ . πεχαφ ηαφ ἥσι παρχηλнсτης ἄε  
†πιστεγe епноῦτε ἄε χин तेनोφ ἥ†ηаῤ̄ पेठооῦ ἄη  
ἥааγ . πελλο δε ἀφχпо ηαφ ἥтκολλῶσε тнῤ̄ ἄε  
ἀφκω ἥсωφ ἥπεφοῦωω ἐтθε πноῦτε .

What you get:  
morphemes ready for tagging

|          |          |  |
|----------|----------|--|
| ἀρχοὺς   | α        |  |
|          | γ        |  |
|          | χοο      |  |
|          | с        |  |
| ἐтθε     | ἐтθε     |  |
| ἀπα      | ἀπα      |  |
| παπνοῦτε | παπνοῦτε |  |
| ἄε       | ἄε       |  |
| μεφθεпн  | με       |  |
|          | φ        |  |
|          | θεпн     |  |
| ἐсενрп̄  | ε        |  |
|          | се       |  |
|          | ηрп      |  |
| .        | .        |  |
| ἐφμοοωε  | ε        |  |
|          | φ        |  |
|          | μοοωε    |  |
| δε       | δε       |  |

# Tokenization Script

Give: formatted  
diplomatic text

Get: morphemes for tagging  
Page, column, line breaks, word  
breaks aligned

α τ τ η σ ο τ ω τ ς \_ ' \_  
Ἰακωβ\_δε\_νε|ς|  
ο γ ω ω \_ π ε ` \_ ε |  
τ ρ ε | ρ ε ν | ω η < / c b > < / p b >  
< p b x m l : i d = " Y A 5 2 0 ">  
< c b > ρ ε ` \_ ω ω π ε ` \_ ν α | ς \_  
ε β ο λ \_ ρ ῆ | ρ ρ α χ ι λ \_ ' \_  
ῆ | ο | ε \_ ῆ | ν | ε ν τ α | γ |  
ω ω π ε \_ ν α | ς \_ ε  
β ο λ \_ ρ ῆ \_ λ ι α \_ ' \_  
\_ . . . . .



|       |              |     |     |       |       |       |    |   |
|-------|--------------|-----|-----|-------|-------|-------|----|---|
| cb    |              |     |     |       | cb    |       |    |   |
| πε`   | ε            | τρε | ρεν | ωηρε` | ωωπε` | να    | ς  |   |
| πε`   | ετρερενωηρε` |     |     |       | ωωπε` | νας   |    |   |
| lb    |              | lb  |     |       | lb    |       |    |   |
|       |              |     |     |       |       |       |    |   |
| p     |              |     |     |       |       |       |    |   |
| YA519 |              |     |     |       | YA520 |       |    |   |
|       |              |     |     |       |       |       |    |   |
| πε`   | ε            | τρε | ρεν | ωη    | ρε`   | ωωπε` | να | ς |

# Part of Speech Tagger

**Give: tokenized text**

**Get: POS-tagged text,  
following Layton**

π  
ΝΟΥΤΕ  
πε  
.

|       |       |
|-------|-------|
| π     | ART   |
| ΝΟΥΤΕ | N     |
| πε    | COP   |
| .     | PUNCT |



Coarse grained tagger: 22 tags

Fine grained tagger: 44 tags


All of these layers feed into the ANNIS database

Open source (Apache 2.0)

Developed here at the Institute for German  
Language and Linguistics / SFB632


# ANNIS & Search





AnnisQL:


Show Result History 

Status: Ok

Corpus List

Visible: Scriptorium 

| Name ▲                      | Texts | Tokens |   |
|-----------------------------|-------|--------|---|
| abraham.our.father          | 7     | 7,689  |   |
| AP.172.antonius             | 1     | 64     |  |
| apophthegmata.patrum.5      | 5     | 700    |  |
| apophthegmata.patrum.parall | 2     | 416    |  |

Get help about: Using the ANNIS interface 


## Using the ANNIS Interface

The ANNIS interface is comprised of several windows, the most important being the search form and the results window.

### The Search Form

#### Search Form

AnnisQL:

Show Result History 

Status: 399 matches

Corpus List

# Corpus Information

Corpus information for abraham.our.father (ID: 66)

**Metadata**

Select corpus/document: abraham.our.father

| Name        | Value                           |
|-------------|---------------------------------|
| annotation  | Caroline Schroeder, Amir Zeldes |
| corpus      | Abraham Our Father              |
| languages   | Sahidic Coptic                  |
| project     | Coptic SCRIPTorium              |
| translation | Rebecca Krawiec                 |
| version     | 1.0.0                           |

**Available annotations**

Node annotations

| Name            | Example (click to use query)     | URL               |
|-----------------|----------------------------------|-------------------|
| add_place       | add_place="right"                | <a href="#">↔</a> |
| cb              | cb="cb"                          | <a href="#">↔</a> |
| dipl            | dipl="ā"                         | <a href="#">↔</a> |
| dipl_word       | dipl_word="."                    | <a href="#">↔</a> |
| hi_rend         | hi_rend="superscript"            | <a href="#">↔</a> |
| lb              | lb="lb"                          | <a href="#">↔</a> |
| norm            | norm="α"                         | <a href="#">↔</a> |
| note            | note="o is directly above the γ" | <a href="#">↔</a> |
| note2           | note2="Luke 14:26"               | <a href="#">↔</a> |
| origin          | origin="Hebrew"                  | <a href="#">↔</a> |
| p               | p="p"                            | <a href="#">↔</a> |
| pb_xml_id       | pb_xml_id="ZHfrg.1dr"            | <a href="#">↔</a> |
| pos             | pos="PREP"                       | <a href="#">↔</a> |
| supplied_reason | supplied_reason="lost"           | <a href="#">↔</a> |
| translation     | translation="Then also, "        | <a href="#">↔</a> |

Edge types

Edge annotations

Meta annotations

*Nodes = the layers or fields of available information, e.g. words, line breaks in manuscripts, normalized tex.*

# Example queries

ANNIS interface showing a search query and a detailed view of corpus information.

**Search Form**

AnnisQL: `norm="N"`

example queries | Tutorial | Query Builder

Get help about: Using the ANNIS interface

logged in as "carrie" | Logout

**Corpus information for abraham.our.father (ID: 66)**

| Metadata                                   |                                 | Available annotations |  |
|--|---------------------------------|-----------------------|--|
| Select corpus/document: abraham.our.father |                                 | Node annotations      |  |
| Name                                       | Value                           | Name                  | Example (click to use query)   URL                   |
| annotation                                 | Caroline Schroeder, Amir Zeldes | add_place             | add_place="right"   <a href="#">↔</a>                |
| corpus                                     | Abraham Our Father              | cb                    | cb="cb"   <a href="#">↔</a>                          |
| languages                                  | Sahidic Coptic                  | dipl                  | dipl="ā"   <a href="#">↔</a>                         |
| project                                    | Coptic SCRIPTorium              | dipl_word             | dipl_word="."   <a href="#">↔</a>                    |
| translation                                | Rebecca Krawiec                 | hi_rend               | hi_rend="superscript"   <a href="#">↔</a>            |
| version                                    | 1.0.0                           | lb                    | lb="lb"   <a href="#">↔</a>                          |
|  |                                 | <b>norm</b>           | <b>norm="N"</b>   <a href="#">↔</a>                  |
|  |                                 | note                  | note="o is directly above the γ"   <a href="#">↔</a> |
|  |                                 | note2                 | note2="Luke 14:26"   <a href="#">↔</a>               |
|  |                                 | origin                | origin="Hebrew"   <a href="#">↔</a>                  |
|  |                                 | p                     | p="p"   <a href="#">↔</a>                            |
|  |                                 | pb_xml_id             | pb_xml_id="ZHfrg.1dr"   <a href="#">↔</a>            |
|  |                                 | pos                   | pos="PREP"   <a href="#">↔</a>                       |
|  |                                 | supplied_reason       | supplied_reason="lost"   <a href="#">↔</a>           |
|  |                                 | translation           | translation="Then also, "   <a href="#">↔</a>        |
|  |                                 | Edge types            |  |
|  |                                 | Edge annotations      |  |
|  |                                 | Meta annotations      |  |

Corpus List

Visible: Scriptorium

| Name               |
|--------------------|
| abraham.our.father |
| AP.172.antonius    |
| apophthegmata.pat  |
| apophthegmata.pat  |



# Edited previous query to search for ⲛⲟⲩⲧⲉ in the normalized text

Search Form

AnnisQL: norm="ⲛⲟⲩⲧⲉ"

Status:

Corpus List

Visible: ☐ Scriptorium

Name

|                             | Texts | Tokens |
|-----------------------------|-------|--------|
| apophthegmata.patrum.5      | 5     | 700    |
| apophthegmata.patrum.parall | 2     | 416    |

Example queries

Base text

Token Annotations

1 Path: abraham.our.father > Abraham.XL93-94\_merge

ⲛ ⲧ ⲙⲏⲧⲉⲣⲟⲛ ⲙ ⲡ ⲛⲟⲩ ⲧⲉ ⲙⲁⲗⲓⲥⲧⲁ ⲁⲉ ⲁⲉ ⲉⲛ

2 Path: abraham.our.father > Abraham.XL93-94\_merge

ⲡⲏⲩ ⲁⲉ ⲛⲛⲉ ⲡ ⲛⲟⲩ ⲧⲉ ⲟⲩⲛⲧ ⲉⲣⲟ ⲛ ⲁⲉ ⲙⲡⲁ

3 Path: abraham.our.father > Abraham.YA518-520\_merge

ⲧ ⲁ ⲡ ⲁⲟⲉⲓⲥ ⲡ ⲛⲟⲩⲧⲉ ⲧ ⲛ ⲟⲩ ⲟⲩⲛⲉ ⲛ

4 Path: abraham.our.father > Abraham.YA518-520\_merge

ⲉⲣⲁⲗⲁⲙ ⲉⲃⲟⲗ ⲉⲓⲧⲙ ⲡ ⲛⲟⲩⲧⲉ ⲁⲉ ⲡⲁⲓ ⲉⲧ ⲙⲁⲗ ⲉⲧⲉ

5 Path: abraham.our.father > Abraham.YA518-520\_merge

ⲉⲙ ⲡ ⲥⲟⲟⲩⲛ ⲙ ⲡ ⲛⲟⲩⲧⲉ ⲛⲉⲓ ⲧ ⲙⲁⲗ ⲙ ⲡ

Embedded Unicode keyboard (Antinoou)

# Click on “Base text” to see search results as diplomatic, normalized, or tokens

The screenshot displays the ANN corpus search interface. The search form on the left contains the query `norm="NOYTE"`. The main results pane shows a list of search results, with the first result selected. The result details pane on the right shows the search results for the selected result, with the "Base text" tab selected. A dropdown menu is open, showing the available search results: "tokens (default)", "dipl", "dipl\_word", and "norm". The "tokens (default)" option is selected. The search results are displayed in a table format, showing the path, annotations, diplomatic text, and normalized text for each result.

**Search Form**

AnnisQL: `norm="NOYTE"`

Status:

Corpus List

| Name                        | Texts | Tokens |
|-----------------------------|-------|--------|
| abraham.our.father          | 7     | 7,689  |
| AP.172.antonius             | 1     | 64     |
| apophthegmata.patrum.5      | 5     | 700    |
| apophthegmata.patrum.parall | 2     | 416    |

**Base text** Token Annotation

1 Path: abraham.our.father

2 Path: abraham.our.father

3 Path: abraham.our.father

4 Path: abraham.our.father > Abraham.YA518-520\_merge

5 Path: abraham.our.father > Abraham.YA518-520\_merge

**tokens (default)**

dipl

dipl\_word

norm

our.father >

Path: abraham.our.father >

Path: abraham.our.father > Abraham.YA518-520\_merge

Path: abraham.our.father > Abraham.YA518-520\_merge

Click on “annotations grid” to see all the annotations of the text

The screenshot displays the Coptic Text Explorer application. At the top, the search bar contains the query 'norm="ⲛⲟⲩⲧⲉ"'. Below the search bar, a corpus list table is visible, showing the following data:

| Name                        | Texts | Tokens |
|-----------------------------|-------|--------|
| abraham.our.father          | 7     | 7,689  |
| AP.172.antonius             | 1     | 64     |
| apophthegmata.patrum.5      | 5     | 700    |
| apophthegmata.patrum.parall | 2     | 416    |

Overlaid on the interface is a Coptic keyboard layout. The main window shows search results for the query. The results are displayed in a list, with each entry showing the path (e.g., 'abraham.our.father > Abraham.XL93-94\_merge') and the text. The text is annotated with various tags, including 'norm="ⲛⲟⲩⲧⲉ"', 'ⲛⲟⲩⲧⲉ', and 'ⲛⲟⲩⲧⲉ'. The results are displayed in a grid view, with each entry showing the path, the text, and the annotations.



# Click on sample searches to search metadata

**Search Form**

AnnisQL: `node & meta::msName="MONB.YA"`

example queries Tutorial Query Builder Query Result

Base text Token Annotations

Displaying Results 1 - 10 of 64

Result for query "norm="NO"

Corpus information for abraham.our.father (ID: 66)

| Metadata                                   |                                 |
|--|---------------------------------|
| Select corpus/document: abraham.our.father |                                 |
| Name                                       | Value                           |
| annotation                                 | Caroline Schroeder, Amir Zeldes |
| corpus                                     | Abraham Our Father              |
| languages                                  | Sahidic Coptic                  |
| project                                    | Coptic SCRIPTorium              |
| translation                                | Rebecca Krawiec                 |
| version                                    | 1.0.0                           |

| Available annotations |  |                   |
|-----------------------|--|-------------------|
| Node annotations      |  |                   |
| Edge types            |  |                   |
| Edge annotations      |  |                   |
| Meta annotations      |  |                   |
| Name                  | Example (click to use query)   | URL               |
| translation           | node & meta::translation="Rebecca Krawiec, Heike Behlmer"                  | <a href="#">↔</a> |
| annotation            | node & meta::annotation="Rebecca Krawiec, Caroline Schroeder, Amir Zeldes" | <a href="#">↔</a> |
| <b>msName</b>         | <b>node &amp; meta::msName="MONB.YA"</b>                                   | <a href="#">↔</a> |
| project               | node & meta::project="Coptic SCRIPTorium"                                  | <a href="#">↔</a> |
| language              | node & meta::language="Sahidic Coptic"                                     | <a href="#">↔</a> |
| repository            | node & meta::repository="Bodleian Library, Oxford"                         | <a href="#">↔</a> |
| resp                  | node & meta::resp="Rebecca Krawiec, Caroline Schroeder, Amir Zeldes"       | <a href="#">↔</a> |
| corpus                | node & meta::corpus="Abraham Our Father"                                   | <a href="#">↔</a> |

Corpus List

Visible: Scriptorium

Name

abraham.our.father

AP.172.antonius

apophthegmata.pat

apophthegmata.pat

Status: 64 meta in 7 d

Here we are searching in all White Monastery codex YA fragments

# All instances of $\omega\eta\rho\epsilon$ in YA

Here we've combined a metadata search (YA manuscripts) with a text search.

[About ANNIS](#) [Report Bug](#) [Help us to make ANNIS better!](#)

**Search Form**  
AnnisQL: `norm="ωηρε" & meta::msName="MONB.YA"`

[Show Result](#) [History](#)

Status: 41 matches  
in 3 documents

Corpus List  
Visible: Scriptorium

| Name                        | Texts | Tokens |  |
|-----------------------------|-------|--------|--|
| abraham.our.father          | 7     | 7,689  |  |
| AP.172.antonius             | 1     | 64     |  |
| apophthegmata.patrum.5      | 5     | 700    |  |
| apophthegmata.patrum.parall | 2     | 416    |  |

[example queries](#) [Tutorial](#) [Query Builder](#) [Query Result](#)

**Base text** **Token Annotations**

1 / 5 Displaying Results 1 Result for query "norm:ωηρε"

1 Path: abraham.our.father > Abraham.YA518-520\_merge

ωα' πε ε τρε ζεν **ωηρε** ωα πε να ι εβολ

annotations (grid)  
 diplomatic text (document)  
 normalized text (document)

2 Path: abraham.our.father > Abraham.YA518-520\_merge

πε · ε τρε ζεν **ωηρε** ωα πε' να ι εβολ

annotations (grid)  
 diplomatic text (document)  
 normalized text (document)

3 Path: abraham.our.father > Abraham.YA518-520\_merge

π νουτε' ι̅ ν̅ ογ **ωηρε'** ν̅ α̅ βραζαμ' ζ̅ν̅ τεβ

annotations (grid)  
 diplomatic text (document)  
 normalized text (document)

4 Path: abraham.our.father > Abraham.YA518-520\_merge

ωαπε να γ̅ ν̅ ογ **ωηρε** ν̅ με ριτ' ν̅ ο

annotations (grid)  
 diplomatic text (document)  
 normalized text (document)

5 Path: abraham.our.father > Abraham.YA518-520\_merge

ε ν̅ ογ τβα' ν̅ **ωηρε'** ε ογα' πε · εβ[.....]

annotations (grid)  
 diplomatic text (document)  
 normalized text (document)



# A little more complicated....

translation=/(^.\*)([Ss]on|[Dd]aughter).\*/ & pos="NPROP" & #1\_i\_#2

Show Result History

6 matches  
in 2 documents

|                    | Texts | Tokens |  |
|--------------------|-------|--------|--|
| ur.father          | 7     | 7,696  |  |
| onius              | 1     | 64     |  |
| nata.patrum.5      | 5     | 700    |  |
| nata.patrum.parall | 2     | 416    |  |

Base text Token Annotations

Result for query "translation=/(^.\*)([Ss]on|[Dd]aughter).\*/ & pos="NPROP" & #1\_i\_#2"

Displaying Results 1 - 6 of 6

1 Path: abraham.our.father > Abraham.YA525-530\_merge

⊕ annotations (grid)  
⊕ diplomatic text (document)  
⊕ normalized text (document)

2 Path: abraham.our.father > Abraham.YA525-530\_merge

⊕ annotations (grid)  
⊕ diplomatic text (document)  
⊕ normalized text (document)

3 Path: abraham.our.father > Abraham.YA525-530\_merge

⊕ annotations (grid)  
⊕ diplomatic text (document)  
⊕ normalized text (document)

4 Path: abraham.our.father > Abraham.YA525-530\_merge

⊕ annotations (grid)  
⊕ diplomatic text (document)  
⊕ normalized text (document)

m.YA525-530\_merge

m.YA547-50\_merge

Searching for a proper noun occurring in the same section of the translation as the word “daughter” occurs (searches Part-of-speech layer and translation layers simultaneously)

# Similar query searching Coptic normalized text

(norm=/ⲱ(ⲏ|ⲉⲉ)ⲣⲉ/ & pos="NPROP" & #1 .1,4 #2 | norm=/ⲱ(ⲏ|ⲉⲉ)ⲣⲉ/ & pos="NPROP" & #4 .1,4 #3)

Show Result History

9 matches  
in 3 documents

atorium

|        | Texts | Tokens |
|--------|-------|--------|
| father | 7     | 7,696  |

Search for a proper noun occurring within four words of the Coptic word for “daughter” occurs (searches Part-of-speech layer and normalization layers simultaneously)

Result for query "(norm=/ⲱ(ⲏ|ⲉⲉ)ⲣⲉ/ & pos="NPROP" & #1 .1,4 #2 | norm=/ⲱ(ⲏ|ⲉⲉ)ⲣⲉ/ & pos="NPROP" & #4 .1,4 #3)"  
Displaying Results 1 - 9 of 9

1 Path: abraham.our.father > Abraham.YA518-520\_merge

ⲡⲛⲟⲩⲩⲧⲉⲁⲓ ⲛⲁⲃⲣⲁⲗⲁⲙ ⲉⲛⲧⲉⲩⲙⲛⲧⲉⲗⲗⲟ

⊕ annotations (grid)  
⊕ diplomatic text (document)  
⊕ normalized text (document)

2 Path: abraham.our.father > Abraham.YA518-520\_merge

ⲛⲧⲉⲓⲉⲙⲗⲁⲓ ⲙⲛ ⲡⲁⲩⲱⲛⲣⲉ ⲓⲥⲁⲁⲕ ⲛⲉⲩⲧⲱⲃⲉ

⊕ annotations (grid)  
⊕ diplomatic text (document)  
⊕ normalized text (document)

3 Path: abraham.our.father > Abraham.YA518-520\_merge

ⲛⲧⲉⲓⲉⲙⲗⲁⲓ ⲙⲛ ⲡⲁⲩⲱⲛⲣⲉ ⲓⲥⲁⲁⲕ ⲛⲉⲩⲧⲱⲃⲉ ⲙⲡⲓⲭⲟⲉⲓⲥ

⊕ annotations (grid)  
⊕ diplomatic text (document)  
⊕ normalized text (document)

4 Path: abraham.our.father > Abraham.YA518-520\_merge

ⲁⲕⲱⲃⲉ ⲉⲃⲟⲗ ⲉⲛⲉⲣⲁⲕⲏⲗⲁⲓ ⲓⲱⲥⲏⲩⲥ



# Search for Greek verbs

About ANNIS

Help us to make ANNIS better

## Search Form

AnnisQL: `source_lang="Greek" & pos="V" & #1 _=_ #2`

Show Result History

Status: 2 matches in 1 document

Corpus List

Search Options

Left Context: 5

Right Context: 5

Show context in: tokens (default)

Results Per Page: 25

example queries Tutorial Query Builder Query Result

Base text Token Annotations

1 Path: apophthegmata.patrum.5 > AP.157.papnoute\_merge

ἡλγστῆς ἀμμετανοὶ ἐφῴω ἰμμος

2 Path: apophthegmata.patrum.5 > AP.157.papnoute\_merge

ἡρῶλο σε ἡπιστεγε ἐπνογτε σε ἡμαεῖρε

|             |                                      |      |          |   |         |           |
|-------------|--------------------------------------|------|----------|---|---------|-----------|
| cb          | cb                                   |      |          |   |         |           |
| coptic_sent | S                                    |      |          | S   |         |           |
| dipl        | ἡρῶλο                                | σε   | ἡπιστεγε | ἐπνογτε   |         |           |
| dipl_word   | ἡρῶλο                                | σε   | ἡπιστεγε | ἐπνογτε   |         |           |
| lb_n        | 37                                   | 38   |          |   | 39      |           |
| norm        | π                                    | ρῶλο | σε       | †   | πιστεγε | ε         |
| pb          | pb                                   |      |          |   |         |           |
| pos         | ART                                  | N    | CONJ     | PPERS   | V       | PREP      |
| source_lang |                                      |      |          |   | Greek   |           |
| translation | He said to him, that is the old man: |      |          | "I trust in God that he will do you charity for |         |           |
| word        | ἡρῶλο                                | σε   | ἡπιστεγε | ἐπνογτε   |         |           |
| tok         | ἡ                                    | ρῶ   | λο       | σε  | †       | πιστεγε ἐ |

# Visualization in ANNIS: diplomatic

Base text Token Annotations

Displaying Results 1 - 10 of 64 Result for query "norm="νοϋτ

3 Path: abraham.our.father > Abraham.YA518-520\_merge

ΝΤ Δ Π ΧΟΕΙΣ Π **ΝΟΥΤΕ** † Ν ΟΥ ΩΗΡΕ Ν

☒ annotations (grid)  
☒ diplomatic text (document)

|                    |                 |
|--------------------|-----------------|
| σπ[.]ηροῶ          | ев[.....]       |
| ερω[ο]οπ:-         | ογΔ[.....]      |
| сеноуѳіоуеπιστολῆ: | πε'ε[.....]     |
| Δβραζαν'πεν        | ἡσαπερ[.....]   |
| 5 εἰωτ'νεχοϋ       | 5 ἡταχει'εβο[λ] |
| ωω'πεετρε          | εἰνεμερος'      |
| εεωηρεωω           | εϋ†ἡτοηδε       |
| πенаचेवोलῆ         | οη'ἡτηαδγ'ἡ     |
| сарра'αγω          | ταсхпог'        |
| 10 сарранесοϋ      | 10 αγωἡτεῖρε'   |
| ωωπε'ετρε          | ακληρονομι      |
| εεωηρεωω           | ἡτεκληρονο      |
| πε'нас'εвол        | μια'ἡνεμειο     |

Click on "diplomatic text" to visualize the text as it appears on the manuscript

# Normalized text

Ḳ̄T Δ Π ΔΘΕΙC Π **ΝΟΥΤΕ'** † Ḳ̄ ΟΥ ΨΗΡΕ' Ḳ̄

#### **+** annotations (grid)

**+** diplomatic text (document)

[-] normalized text (document):

CON[.]N N2OCON ENDOOP : -

ΣΕΝΟΥΘΙΟΥ ΕΠΙΣΤΟΛΗ :

Abraham our father wished to have children with Sarah.

Abraham our father wished to have children with Sarah.

εβ[.....]ογΔ[.....]πε ' ε[.....]νσα πεχ[.....] νταφειεβολ ρη νεφμερος ' εφ† ητον δε οη ντηαδγ  
νταсχποч ' αγω ντειρε αчκληροноми ντεκληρονομια нνεφειοτε , κατα πενταγχοοч  
навразам εβολ ριτμηππογτε χε παι ετμηαγ ετειсmanη πε нφνακληρονομει мнок an ' πλιν  
†насμοу δε οη εροч εтвннтк тааγzane ммоу енате нтааач ноугнос нρεθнос енаωωч '

# Coreference

Ἄ γ' αὐτοῦ σ' ἐτβε ἅπα πάπ' νοῦτε· ἔ με γ' σε πη ἐ σε ἡρπ' . - Ἐ γ' μοῶτε  
δε π' οὐ οὐδ' εἶω α' γ' εἰ ἔχ' οὐ κολλυγκίον ἢ λῦς τῆς αὐτῷ α' γ' σὴτ οὐ ἐ γ'  
σῶ . α' γ' σοῦων ὅ Δε ἡσὶ π' ἐτ ο' ν' ἀνδρῶν αὐτῷ ἐ ὅ σοῦγ' ἔ με γ' σε  
ἡρπ' . - Ἄ γ' ναγ' ἐρο γ' γ' οὐ νοσ' ἢ γίσε α' γ' μογ' ν' οὐ ἀποτ' ν' ἡρπ' ἐρε  
τεγ' σηγε γ' τεγ' σίχ' . αὐτῷ πεχᾶ γ' ἡ π' γ' λ' λο· ἔ κ' τῇ σῶ † να γ' οτβε  
κ' . - Ἄ γ' εἰμε ἡσὶ π' γ' λ' λο· ἔ γ' οὐῶν ἐ εἶρε ν' οὐ ἢ τολῆ ἢ τε π' νοῦτε ἐ  
γ' οὐῶν ἐ † γ' ηγ' ν' μο γ' α' γ' χίτ' ὅ α' γ' σῶ ὅ . Π' ἀνδρῶν Δε ἢ λῦς τῆς  
α' γ' μετανοῖ ἐ γ' χῶ ἡμο σ' ἔ κῶ να ἰ ἐβῶ πα εἶωτ' ἔ α' ἰ † γίσε να κ'

Papnoute: referenced by name, the old man, “my father”

We also have the HTML visualizations you saw online

# Wait, there's more...

Download TEI (Text Encoding Initiative) XML  
compatible (text and metadata)

Files validate to EpiDoc TEI schemas to ensure  
interoperability

# Open Source & Open Access

Texts and Tools:

- Freely available
- Can be modified
- Attribution to sources required (cc-by)

# Long term goals

Expand corpus (with collaborators)

Improve existing taggers and tokenizer

Expand tools:

- Named entity recognition
- Language of origin (Greek, Latin, Hebrew...)
- Lemmatizer

Collaborative platform

- Text contributions
- Annotation layers



# What we want from you

Texts

Feedback: what do you want from a database?

Feedback: what visualizations do you want?

Feedback: comments on tools

Funding & Support has been provided by:

- A grant from the Pacific Fund, the University of the Pacific
- A grant from the Faculty Research Council, the University of the Pacific
- The Department of Religious and Classical Studies, The University of the Pacific
- The Institute for German Language and Linguistics, Humboldt University

C. Schroeder and A. Zeldes gratefully acknowledge the following individuals and organizations for their contributions to the project:

- The Corpus dei Manoscritti Copti Letterari (Coptic lexicon and other materials)
- Eric E. Johnson, Tibco Software
- Rebecca S. Krawiec, Ph.D., Canisius College (Abraham Our Father annotations and English translation)
- Tito Orlandi, Ph.D., Sapienza University, Rome; the University of Hamburg
- Janet Timbie, Ph.D., Catholic University of America (Apophthegmata Patrum annotations)
- [The ANNIS team at Humboldt University](#)
- Alex Dickerson, the University of the Pacific, student