

Coptic Scriptorium

Searching Digital Texts and Using Tools

Amir Zeldes
Georgetown University
amir.zeldes@georgetown.edu

Caroline T. Schroeder
University of Oklahoma
ctschroeder@ou.edu

13th International Congress of Coptic Studies
Digital Coptic Tutorial Session
Göttingen, 30 July 2026

Coptic Scriptorium



COPTIC
SCRIPTORIUM

- A collaborative, digital project for Coptic research offering:
 - Richly annotated, searchable Coptic corpora
 - Natural language processing tools:
 - Automatic segmentation
 - Grammatical analysis (identify verbs, their subject nouns, morphology...)
 - Entity recognition and linking (find mentions of John the Baptist...)
 - Manuscript lacuna reconstruction
 - The Coptic Dictionary Online
(collaboration with TLA/DDGLC and others)
 - And more!

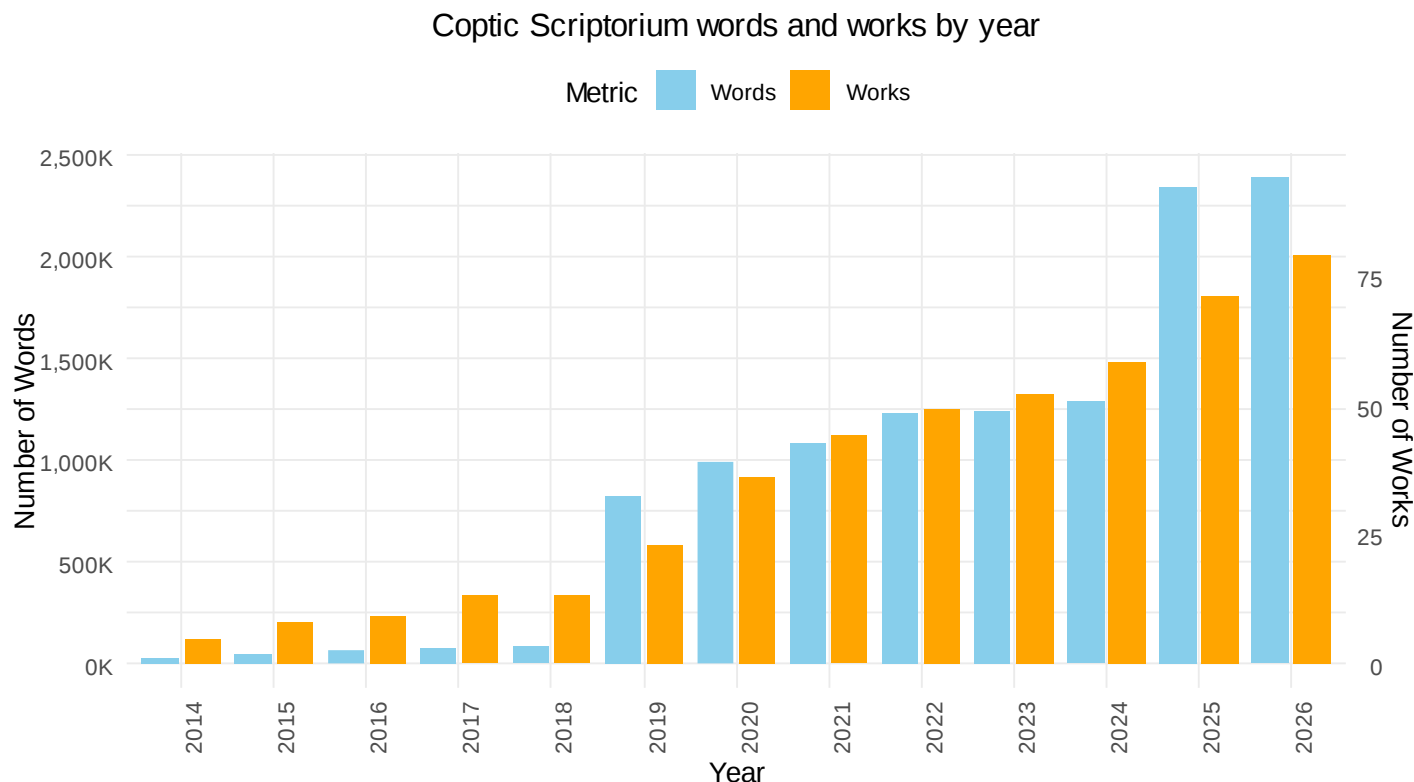
<https://copticSCRIPTORIUM.org/>

Plan

- Reading Coptic texts – the CTS repository
- Using the Coptic Dictionary Online (CDO)
- Linguistic search: ANNIS, parts of speech and syntax trees
- Coptic Natural Language Processing service (Coptic-NLP)
- Lacuna reconstruction tool (experimental)
- Machine readable data (formats on GitHub)

Data – Coptic Scriptorium

- We now have over 75 works with over 2.4M analyzed words!
- More than a decade of data collection & annotation – ongoing



What kind of works do we have?

- Sermons: Shenoute's *Abraham our Father*
 - *"As for us, brethren, let us live by the truth so that we are upstanding in all our works, and so that the prophets, apostles and all the saints might dwell among us, ..."*
- Letters: Besa's *Letter to Aphthonia*
 - *".. you sent to your father and mother: 'they fought with me,' or 'they abused me'. It's you lying, as they didn't fight with you nor did they abuse you..."*
- *Apophthegmata Patrum* (sayings of the desert fathers)
 - *"They said about the blessed Sarah the virgin that she spent sixty years living at the top of the river and she never set foot outside to see the river."*
- Bible, saints' lives, letters, documentary material...
and many other works, all freely available and searchable



Want to read these in
Coptic?

COPTIC
SCRIPTORIUM

<https://copticSCRIPTORIUM.org>

Coptic Scriptorium repository (CTS)

<https://data.copticscriptorium.org/>

- Reading texts in different visualizations
 - Normalized
 - Analytic
 - Diplomatic (where available)
 - Dictionary links
- Searching
 - Basic text search
 - Metadata search
 - Search by entity

Coptic Scriptorium repository (CTS)

<https://data.copticscriptorium.org/>



HOME CORPUS AUTHOR MSNAME PEOPLE PLACES ADVANCED SEARCH besa, ὡληλ, "πेत्रος", εκκλησία -Alexandria or URN...

Apophthegmata Patrum Sahidic 025: Pambo

diplomatic normalized analytic TEI [ANNIS](#) [PAULA](#) [Search ANNIS](#)

Diplomatic Edition

Ντοφόνάχχοοσε
ἀάπαπαμβωχοοσε
χεταίτεθέέτεϝ
25 φρεπιμοναχοσε
Φορείνηεϝοίτε·
ρωστεένεχτεϝ
φτηνήπιβολή
τεϝρίηφρονήτῆ
30 ροογ.ήτετῆλααγ
ταίοςέϝίτς·τοτε
έϝεΦορίμνος·

A. Zeldes & C. T. Schroeder

Coptic Scriptorium tutorial

Coptic Scriptorium repository (CTS)

<https://data.copticscriptorium.org/>

Apophthegmata Patrum Sahidic 025: Pambo

diplomatic

normalized

analytic

TEI 

ANNIS 

PAULA 

Search ANNIS 

Normalized Text

Chapter 1: [\[EG67\]](#) (1) ΝΤΟϢ ΟΝ ΔΕ ΧΟΟC ΔΕΔ ΑΠΑ ΠΑΜΒΩ ΧΟΟC ΧΕΤΑΙ ΤΕ ΤΞΕ

ΕΤΕΦΩΞ ΕΠΜΟΝΑΧΟC ΕΦΟΡΕΙ ΝΝΕΦΖΕC • (2) ΕΦΩCΤΕ ΕΝΕΧ ΤΕΜΑΝΤΙΩ ΜΠΡΟ)

ΝΤΕΦΡΙ ΝΦΟΜΝΤ ΝΞΟΟΥ . ΝΤΕΤΗ ΛΑ

He also said that Abba Pambo said: "This is the way that it is fitting for the monk to wear his clothes:

Coptic Scriptorium repository (CTS)

<https://data.copticscriptorium.org/>

HOME CORPUS AUTHOR MSNAME PEOPLE PLACES **ADVANCED SEARCH** besa, ψληλ, "πέτρος", εκκλησία -Alexandria or URN... 🔍 ☐ Keyboard ☒ Exact Search

23 results ⓘ Arabic_Translation = Philippe Zaher (23) ✕

Page 1 of 2 Next »

Abraham Our Father XL 93-94
Shenoute urn:cts:copticLit:shenoute.abraham.monbxl:23-24

Abraham Our Father: YA 518-20
Shenoute urn:cts:copticLit:shenoute.abraham.monbya:1-4

Abraham Our Father: YA 525-30
Shenoute urn:cts:copticLit:shenoute.abraham.monbya:10-18

Abraham Our Father: YA 535-40

Refine your search

Annotation
Caroline T. Schroeder (21)
Amir Zeldes (17)
Christine Luckritz Marquis (15)
Marina Ghaly (10)
David Sriboonreuang (6)
Elizabeth Davidson (5)
Rebecca Krawiec (5)
Elizabeth Platte (4)
Philippe Zaher (2)
Show More ▼

Arabic translation
Philippe Zaher (23) ✕

Author
anonymous (18)
Shenoute (5)

Corpus
apophthegmata.patrum (18)
shenoute.abraham (5)

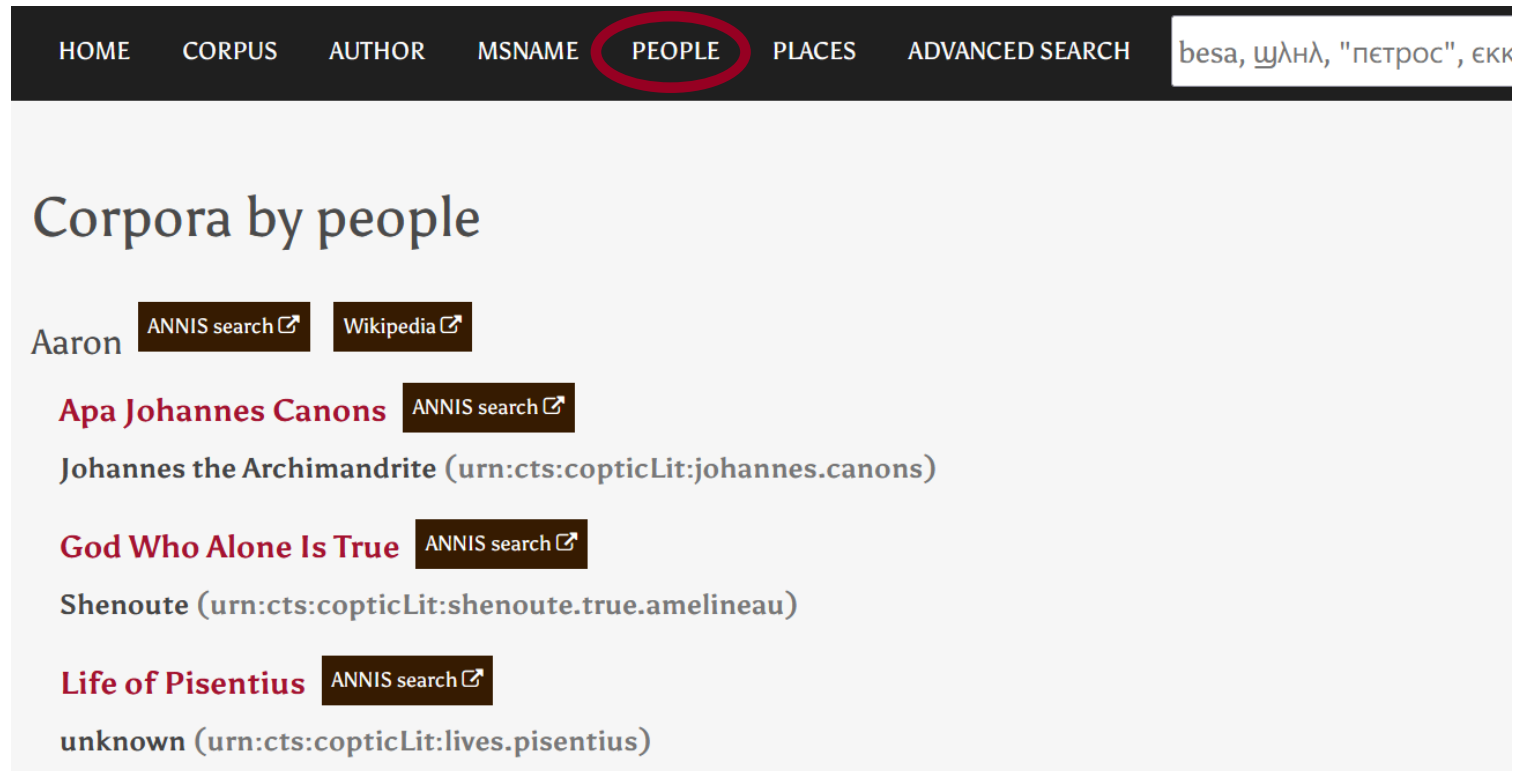
Msnames
MONB.EG (18)
MONB.YA (3)

A. Zeldes & C. T. Schroeder

Coptic Scriptorium tutorial

Coptic Scriptorium repository (CTS)

<https://data.copticscriptorium.org/>



HOME CORPUS AUTHOR MSNAME **PEOPLE** PLACES ADVANCED SEARCH besa, ὡληλ, "петрос", екс

Corpora by people

Aaron [ANNIS search](#) [Wikipedia](#)

Apa Johannes Canons [ANNIS search](#)
Johannes the Archimandrite (urn:cts:copticLit:johannes.canons)

God Who Alone Is True [ANNIS search](#)
Shenoute (urn:cts:copticLit:shenoute.true.amelineau)

Life of Pisentius [ANNIS search](#)
unknown (urn:cts:copticLit:lives.pisentius)

A. Zeldes & C. T. Schroeder

Coptic Scriptorium tutorial

Coptic Dictionary Online

<https://coptic-dictionary.org/>

- Basic search
- Advanced search:
 - Search by definition (trilingual)
 - Part of speech and dialect
 - Quick search
- Detailed information
 - Frequency information
 - Linking to other dictionaries
 - Collocation analysis
 - Example usage and entity matching
 - Network graphs

Coptic Dictionary Online

<https://coptic-dictionary.org/>

Coptic Dictionary Online

[Home](#) [How to](#) [About](#)

DH 2013 Awards Winner: Best DH tool or suite of tools

Search [?]

Coptic Word:
 (install Coptic keyboard)

Dialect:

Scriptorium tag:

Definition:

Definition text:
☒ exact sequence
☐ contains all words

Definition language:

Need help segmenting Coptic text?
Try the [Coptic NLP service](#)

Lexicon data released under [CC BY-SA License](#). Search interface code is released under the [Apache 2.0 license](#). CSS based on [Bootstrap](#). Icons from [Font Awesome](#). Web fonts from [Google](#). Theme adapted from [Thomas Park](#) under the [MIT License](#).



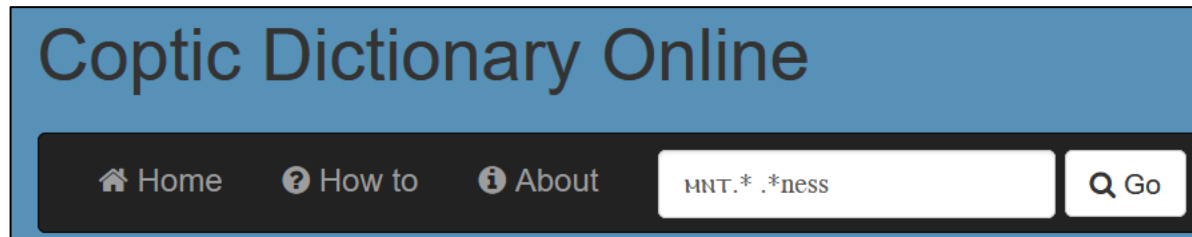
A. Zeldes & C. T. Schroeder

Coptic Scriptorium tutorial

Coptic Dictionary Online

<https://coptic-dictionary.org/>

- Quick search:
 - Enter a Coptic word
 - Definition word (English/French/German)
 - TLA ID (C3815)
 - Regular expression: MNT.*
 - Any combination

A screenshot of the Coptic Dictionary Online website. The header is a blue bar with the text "Coptic Dictionary Online" in white. Below the header is a dark navigation bar with three links: "Home" with a house icon, "How to" with a question mark icon, and "About" with an information icon. To the right of these links is a search input field containing the text "MNT.*.*ness". To the right of the input field is a "Go" button with a magnifying glass icon.

Coptic Dictionary Online

Home How to About

MNT.*.*ness

Go

Coptic Dictionary Online

<https://coptic-dictionary.org/>

Coptic Dictionary Online

[Home](#) [How to](#) [About](#)

Search results

You searched for 'ⲙⲏⲧ.ⲥ' definitions matching .ⲥⲏⲥ in language *any*
67 Results

ⲙⲏⲧⲁⲗⲟϥ	--	time of youth, youthfulness
ⲙⲏⲧⲁⲥ	ⲙⲏⲧⲁⲡⲁⲥ	oldness
ⲙⲏⲧⲁⲧⲁⲙⲁⲗⲧⲉ	ⲙⲉⲧⲁⲧⲁⲙⲁⲗⲧⲉ	uncontrolledness, incontinence
ⲙⲏⲧⲁⲧⲁⲣⲏⲭⲥ	--	boundlessness
ⲙⲏⲧⲁⲧⲏⲟϥ	--	deathlessness, immortality
ⲙⲏⲧⲁⲧⲏⲡⲱⲗⲁ	--	unworthiness
ⲙⲏⲧⲁⲧⲏⲁ	--	pitilessness
ⲙⲏⲧⲁⲧⲏⲟϥⲧⲉ	--	godlessness
ⲙⲏⲧⲁⲧⲏⲟϥⲣⲉ	--	unprofitableness
ⲙⲏⲧⲁⲧⲟϥⲭⲁⲓ	--	unsoundness
ⲙⲏⲧⲁⲧⲡⲱⲱⲛⲉ	--	unchangeableness

A. Zeldes & C. T. Schroeder

Coptic Scriptorium tutorial

Coptic Dictionary Online

<https://coptic-dictionary.org/>

Form	Dial.	Form ID	POS	Attestation
ⲙⲏⲧⲣⲣⲟ	S	CF8163	Subst. f.	ⲉ ⲓⲁ ⲛⲓ
ⲙⲉⲧⲟⲩⲣⲟ	B	(from Coptic Scriptorium)		ⲉ ⲓⲁ ⲛⲓ

Scriptorium tag: N

Known entity types:   

1. (En) kingdom, reign
(Fr) royauté, règne, royaume
(De) Königtum, Herrschaft
Bibliography: [CD 299b](#)[Ⓢ]; [KoptHWb 165](#)[Ⓢ]; [ChLCS 42a](#)[Ⓢ]

Example usage: (automatically extracted, use with caution and report bad examples)

Sahidic:

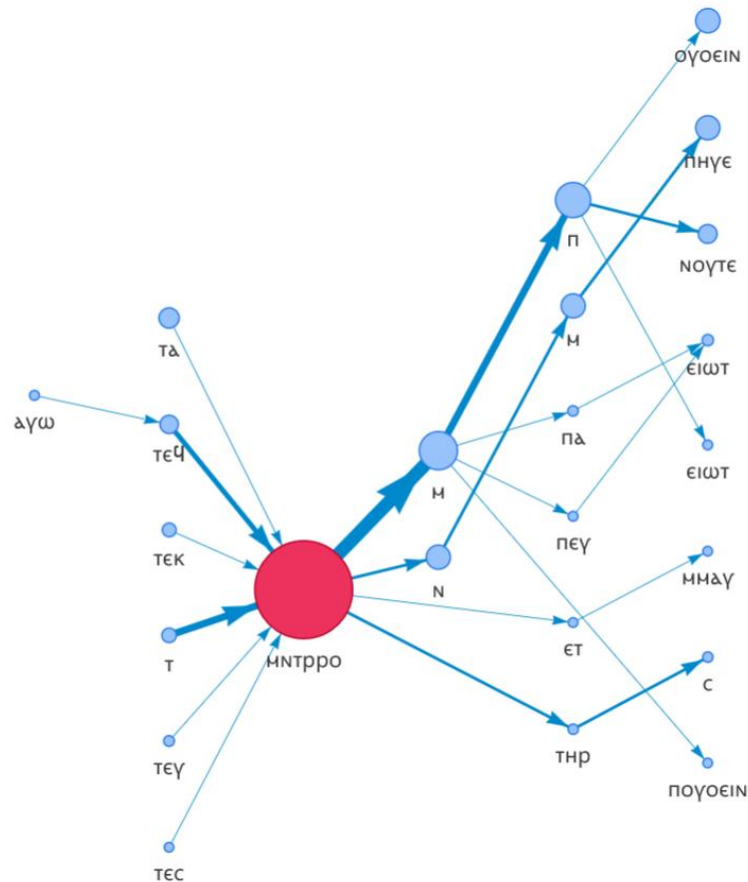
- ⲉⲩⲏⲧⲏⲧⲏ ⲙⲙⲁⲩ ⲙⲡⲁⲣⲏⲃ ⲛⲧⲉⲩⲙⲏⲧⲣⲣⲟ ·
There is for you the pledge of His kingdom
Dormition of John **MERC.AD** in Budge ed. ([urn:cts:copticLit:misc.dormition_john.budge](#))
- ⲩⲏⲛⲉ ⲛⲧⲟⲩ ⲛⲁⲩⲉⲩⲙⲏⲧⲣⲣⲟ ⲙⲏⲧⲉⲩⲁⲓⲕⲁⲓⲟⲩⲩⲏⲏ ·
Seek ye His kingdom and His righteousness,
Life of Onnophrius part 2 ([urn:cts:copticLit:lives.onnophrius.budge:9-16](#))
- ⲁⲗⲗⲁ ⲉⲛⲙⲟⲕⲉⲣ ⲛⲉⲣⲏⲧ ⲁⲩⲱ ⲉⲛⲗⲩⲡⲉⲓ ⲙⲡⲉⲣⲟⲩⲩ ⲙⲏⲧⲟⲩⲩⲱⲛ · ⲉⲧⲃⲉⲛⲉⲧⲟ ⲛⲁⲧⲥⲱⲧⲏ ⲛⲁⲡⲉⲩⲁⲓⲧⲉⲓⲟⲩ ⲙⲡⲉⲛⲭⲟⲉⲓⲥ ⲓⲛⲥⲟⲩⲥ
ⲛⲛⲉⲓⲱⲛⲉ ⲩⲱⲛ ⲛⲛⲉⲧⲏⲃⲟⲕ ⲉⲣⲟⲩⲛ ⲉⲧⲙⲏⲧⲣⲣⲟ ⲛⲙⲡⲏⲩⲉ

A. Zeldes & C. T. Schroeder

Coptic Scriptorium tutorial

Coptic Dictionary Online

<https://coptic-dictionary.org/>



ANNIS – linguistic searches

<https://annis.copticscriptorium.org/annis/scriptorium>

- Words and word sequences
- Part of speech tags
- Manuscript properties
- Metadata
- Syntax trees and functions
- Entity types and named entity identities
- Quantitative breakdowns and export

➤ Tutorial cheat sheet: [https://copticscriptorium.org/ANNIS tutorial](https://copticscriptorium.org/ANNIS_tutorial)

Annotations: The grid-based model

ΚΑΤΑΤΖΕ **ΕΝΤΑΤΜΝΤΜΝΤΡΕ** ΜΠΝΟΥΤΕ ΤΑΧΡΟ

☐ annotations (grid)

lemma	π	ζε	ετερε	α	π	μντμντρε		ν	π	νουτε	ταχρο
morph						μντ	μντρε				
norm	τ	ζε	εντ	α	τ	μντμντρε		μ	π	νουτε	ταχρο
norm_group	κατατζε		εντατμντμντρε					μπνουτε			ταχρο
orig	τ	ζε	εντ	α	τ	μντμντρε		μ	π	νουτε	ταχρο
orig_group	κατατζε		εντατμντμντρε					μπνουτε			ταχρο
pos	ART	N	CREL	APST	ART	N		PREP	ART	N	V
func	det	root	mark	aux	det	nsubj		case	det	nmod	acl:relcl
entity					abstract						
entity									person		
translation	even as the testimony of Christ was confirmed in you:										
chapter_n	1										
verse_n	6										
vid_n	urn:cts:copticLit:nt.1cor.sahidica_ed:1.6										

Annotations: keys, values, regular expressions

Query	Result
norm="ⲧⲉⲓ"	standard word form
orig="ⲧⲉⲉⲓ"	form as spelled in manuscript
lemma="ⲡⲉⲓ"	dictionary entry
norm=/. *oc/	form ends in -oc
lang	item with a language annotation (Greek, Latin, Hebrew, etc.)
lang="Greek"	exactly all Greek origin terms
pos="V" _o_ lang="Greek"	a verb overlapping Greek material (part or all of word)
pos="VSTAT"	all stative verbs
pos=/V. */	any kind of verb
pos=/CCIRC CFOC/	a circumstantial or a focalizing converter
morph="ⲙⲛⲧ"	part of a compound unit
norm_group="ⲙⲡⲉⲭⲣⲓⲥⲟⲥ"	whole bound group
orig_group="ⲙⲡⲉⲭⲥ̅"	whole group as spelled
norm="ⲧⲉⲓ"	standard form

The POS tag set 1/2

Tag	Name	Examples
A.*	Auxiliary	α[ϣ], με[ϣ], τρε[ϣ], ...
ADV	Adverb	εβολ, ον, πως
ART	Article	π(ε), τ(ε), ν(ε), zen, κε
C.*	Converter	ε, ετε, νε, ...
CONJ	Conjunction	αγω, δε, η, μη, ειτε, ...
COP	Copula	πε/τε/νε
EXIST	Existential/possessive	ουν/μν
FM	Foreign material	παρα τογτο
FUT	Future	να
IMOD	Inflected modifier	τηρ[ϣ], ζωω[τ], ...
N.*	Noun	αθτ, ρωμε, αρχη, ...
NEG	Negation	ν, αν, τμ[σωτμ]
NUM	Numeral	ογα, cnaγ, ...

➤ And subtypes: APST, ANEGPST, AAOR... N, NPROP CCIRC, CFOC, ...

The POS tag set 2/2

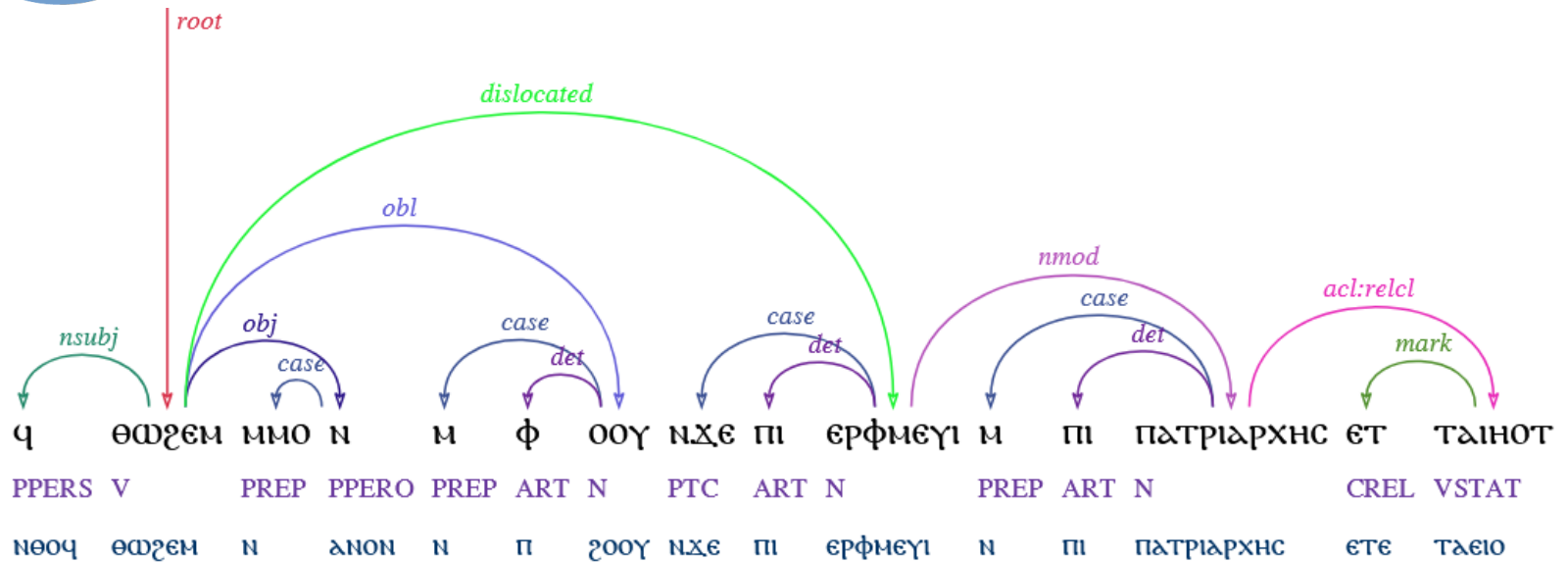
Tag	Name	Examples
PDEM	Pronoun, demonstrative	πει/παι, τει/ται, νει/ναι
PINT	Pronoun, interrogative	οϋ, νιμ
PPER.*	Pronoun, personal	ϣ,ϣ,ι,ϣ,ν,ανοκ,ανῖ,...
PPOS	Pronoun, possessive	πεϣ,τετῖ,ποϋ,πα,πωι,...
PREP	Preposition	ετβε, ζῖ, ν, ῖμο[ϣ], ...
PTC	Particle	δε, ῖβι, ...
PUNCT	Punctuation	. , ' ...
UNKNOWN	Unknown, lacuna	В_ _ _ , _ _ _ _ _ , _ _ _ , ...
V.*	Verb	сωтм, сωтп, сотп, еире, о, ари, ...
VBD	Verboid	наноϣ[ϣ], πεχα[ϣ], πεχε,...

- And subtypes: PPERS, PPERO, PPERI V, VSTAT, VIMP
- Where necessary we have fused tags:
 - τοοτϣ “his hand” = two words, two tags: N PPERO
 - τοοτ “my hand” = one word, tag: N_PPERO

Multiple words

Query	Result
norm="Νῆι" . norm="Τ"	Νῆι followed by feminine article
lang="Greek" . lang="Greek"	two adjacent Greek words
norm="/ΝΤΟΚ?/" .1,3 lemma="πε"	ΝΤΟ or ΝΤΟΚ followed by lemma πε within 3 tokens
lemma="αγω" ^1,5 lemma="ον"	αγω and ον within 5 tokens (any direction)
pos="N" . pos="PPER0"	noun followed by clitic pronoun
norm="α" . norm="γ" . norm="σωτη"	the sequence α-γ-σωτη
lemma="ογ" . norm="εβολ" . lemma="ζν"	ογ-εβολ followed by lemma ζν
norm="ρ" . lemma="π" . pos="N"	ρ, definite article, noun (ρημεεγε...)
norm="ε" . pos="V" . pos="PPER0"	to-infinitive followed by clitic pronoun

Syntactic structure – Universal Dependencies



Syntactic queries

Query	Result
<code>func="obj"</code>	direct object of a verb
<code>func="vocative"</code>	vocative appellation
<code>func=/.subj/</code>	any kind of subject (nominal or clause)
<code>lemma="j" ->dep func="obj"</code>	the verb j and the object it governs
<code>pos="V" ->dep func="dislocated" & #2 .* #1</code>	verb with preceding dislocated participant
<code>entity="person" ->head pos="NPROP"</code>	person entity headed by a proper noun

- See the cheat sheet for the complete list of syntactic function labels!

Metadata

Search by MS:

- `norm="ⲙⲉⲣⲉ" & meta::msName="MONB.YA"`
- `norm="ⲙⲉⲣⲉ" & meta::msName="/MONB.*/"`

Path: coptic.treebank > 1Cor_01 (tokens 140 - 150)

ⓘ

⊞ annotati

ⲕⲁⲧⲁⲧⲉ

Info for salt:/coptic.treebank/1Cor_01 + x

Metadata

document: 1Cor_01

Name	Value
Greek_source	Greek text annotated as sbl_greek and sbl_apparatus are from the SBL Greek New Testament (c)2010 Society of Biblical Literature and Logos Software.
annotation	Elizabeth Platte, David Sriboonreuang, Christine Luckritz Marquis, Mitchell Abrams, Lance Martin, Caroline T. Schroeder, Amir Zeldes
book	1Cor
chapter	1
corpus	sahidica.1corinthians
document_cts_urn	urn:cts:copticLit:nt.1cor.sahidica_ed:1
entities	gold
identities	gold

corpus: coptic.treebank

lemma

morph

norm

norm_group

orig

orig_group

pos

func

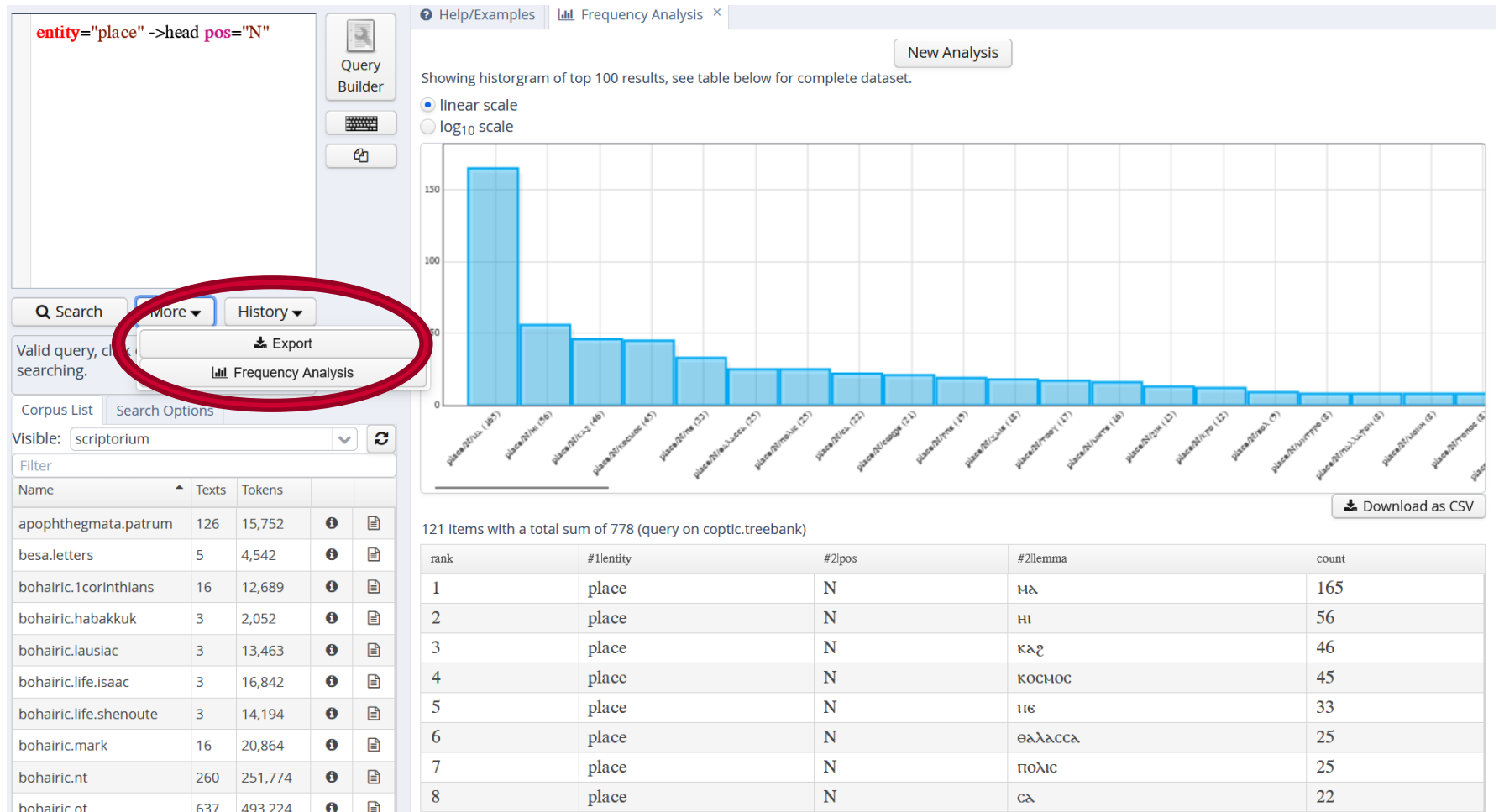
entity

entity

translation

chapter_1

Aggregate statistics



A. Zeldes & C. T. Schroeder

Coptic Scriptorium tutorial

Coptic NLP

<https://tools.copticscriptorium.org/coptic-nlp/>

- Automatic word segmentation
- Part of speech tagging
- Entity recognition
- Syntax parsing

Coptic NLP

<https://tools.copticscriptorium.org/coptic-nlp/>

Coptic NLP Service

Enter Coptic text in UTF-8 (XML markup is also allowed, 10,000 characters max).

Bound groups should be separated by **spaces** or **underscores**.

If you need to analyze longer texts or multiple texts automatically, you can log in to the **secure** area or use the **API**. For a login please contact **Amir Zeldes**.

Dialect:

- ☒ Auto-detect
- ☐ Sahidic
- ☐ Bohairic

Input:

- ☐ My data contains meaningful linebreaks ⓘ
- ☒ Ignore linebreaks in my data

Output:

- ☐ Use old finite state tokenizer ⓘ
- ☐ Re-merge bound groups ⓘ
 - ☐ Conservative merging ⓘ
 - ☐ Aggressive merging ⓘ
 - ☐ Smart merging ⓘ
 - ☐ Segment at merge point ⓘ

Coptic NLP

<https://tools.copticscriptorium.org/coptic-nlp/>

☐ SGML pipeline ☒ Just piped and dashed morphemes

☒ Stretch milestones ⓘ
☒ Tokenize [stk-6.0.0]

☒ Automatic
☐ From pipes in input

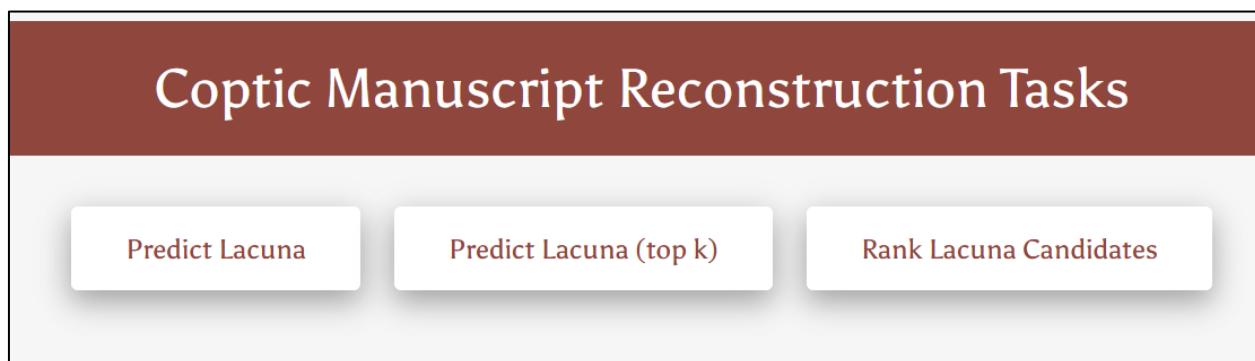
☒ Normalize ⓘ
☒ Tag[fairbert-6.0.0]
☒ Lemmatize
☒ Language of origin
☒ MWE recognition ⓘ
☒ Parse[diabert-UD2.15]
☒ Entity recognition ⓘ

λ<hi rend="red">ϣ</hi>ϥⲧⲏ ⲛϥⲱ ϣⲏⲥ
παγγελος ⲙⲏⲛⲧⲱⲃⲏⲣ .
ⲱⲧⲧⲏⲣⲱϥⲉ ⲛⲧⲱⲧⲏ ⲛⲧⲕⲁⲛⲁⲧⲏ ⲛⲡⲥⲁⲧⲁⲛⲁⲥ
ⲉⲧⲣⲉϥⲉⲓ ⲉⲣⲟϥⲏ ⲛⲧⲕⲱⲧⲏ ⲛⲏⲛⲁ ⲉⲧⲟϥⲁⲁⲃ
ⲛⲡⲛⲟϥⲧⲉ ⲁϥⲱ ⲛⲧⲉⲧⲏⲥⲱⲱϥ ⲛⲡⲉϥⲣⲏⲉ
ⲉⲧⲉⲏⲧⲱⲧⲏ ⲛⲉ ⲛⲙⲏⲛⲙⲱⲧⲏ ⲛⲡⲉⲧⲏⲣⲏⲛⲉⲉϥⲉ
ⲛⲧⲁⲡⲟⲫⲁⲥⲏⲥ ⲛⲡⲁⲡⲟⲥⲧⲟⲗⲟⲥ
ⲕⲉⲡⲉⲧⲛⲁⲥⲱⲱϥ ⲛⲡⲉⲣⲏⲉ ⲛⲡⲛⲟϥⲧⲉ .

Lacuna reconstruction (experimental)

<https://gucorpling.org/lacuna-demo>

- Reconstructing individual letters
- Top-K candidates
- Ranking proposed candidates
- Comparing models



➤ More details in the paper:

<https://aclanthology.org/2024.ml4al-1.8/>

Machine Readable Data (GitHub)

<https://github.com/CopticScriptorium/corpora>

- Downloading datasets
- Overview of formats
 - TEI XML
 - TT SGML
 - CoNLLU
 - ANNIS & PAULA

Thanks to our team, our supporters and you!

**COPTIC
SCRIPTORIUM**



- Coptic Scriptorium PIs:
 - Amir Zeldes, Georgetown University
 - Caroline T. Schroeder, University of Oklahoma
 - DH Specialist
 - Nicholas Wagner
 - Senior Editors & Consultants:
 - Rebecca Krawiec, Canisius University
 - Christine Luckritz-Marquis, Union Presbyterian Seminary
 - Hany Takla, St. Shenouda Society
 - Nina Speransky, Hebrew University of Jerusalem
 - And many more volunteers! (<https://copticcriptorium.org/about>)
- We thank the St. Shenouda Society for their generous support!
- Support us at: <https://copticcriptorium.org/donate>