

# Computational Analysis on Text Re-use of Shenoute and Besa

So Miyagawa & Dr. Marco Büchler

Georg-August-Universität Göttingen

August 27, 2016

**GCDH**



## Introduction

### Preprocessing1

OCR

Converter

Virtual Manuscript Room

### Preprocessing2

Natural Language Processing

### Texts of Shenoute and Besa

### Processing

Text Reuse Detection by Tracer

### What I found

## Text reuse example

### Shenoute Canon 6, Codex MONB.XF, p. 1

π-                    ετ-    ρμοοϑ   ρι-   πεϑ-                    θρονοϑ   ἰϑ̄  
 DEF.SG.M-   REL-   sit            on   POSS.SG.M:3SG.M-   throne.M   Jesus  
 π-                    ωηρε   ᾱ-            π-                    νογτε    π-                    ετ-  
 DEF.SG.M-   son.M   ATTR-   DEF.SG.M-   god.SG.M   DEF.SG.M-   REL-  
 κρινε   ᾱ-            τ-                    δικαιοϑνη  
 judge   ADVZ-   DEF.SG.F-   justice.F

“He who sits upon his throne, Jesus, the Son of God, he who judges justly”.

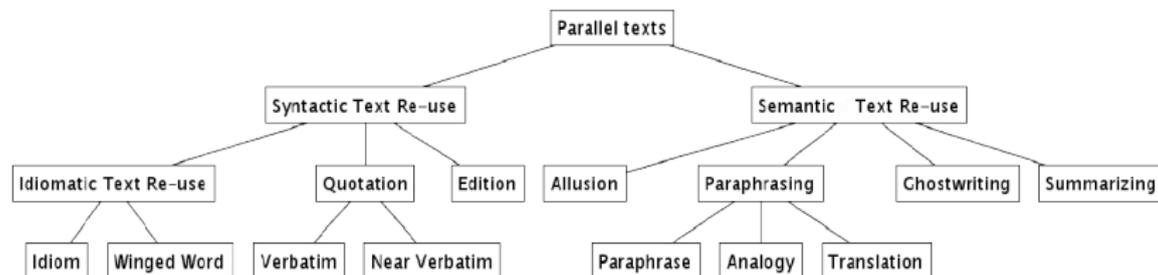
# The source

## Psalm 9:4b

δ- κ- ρΜΟΟC ρι- ΠΕΚ- ΘΡΟΝΟC Π-  
 PST- 2SG.M- sit on- POSS.SG.M:2SG.M- throne.M DEF.SG.M-  
 ΕΤ- κΡΙΝΕ ἄ- Τ- ΔΙΚΑΙΟCΥΝΗ  
 REL- judge ADVZ- DEF.SG.F- justice.F

“You sat upon your throne, you who judge justly”.

# Reuse styles



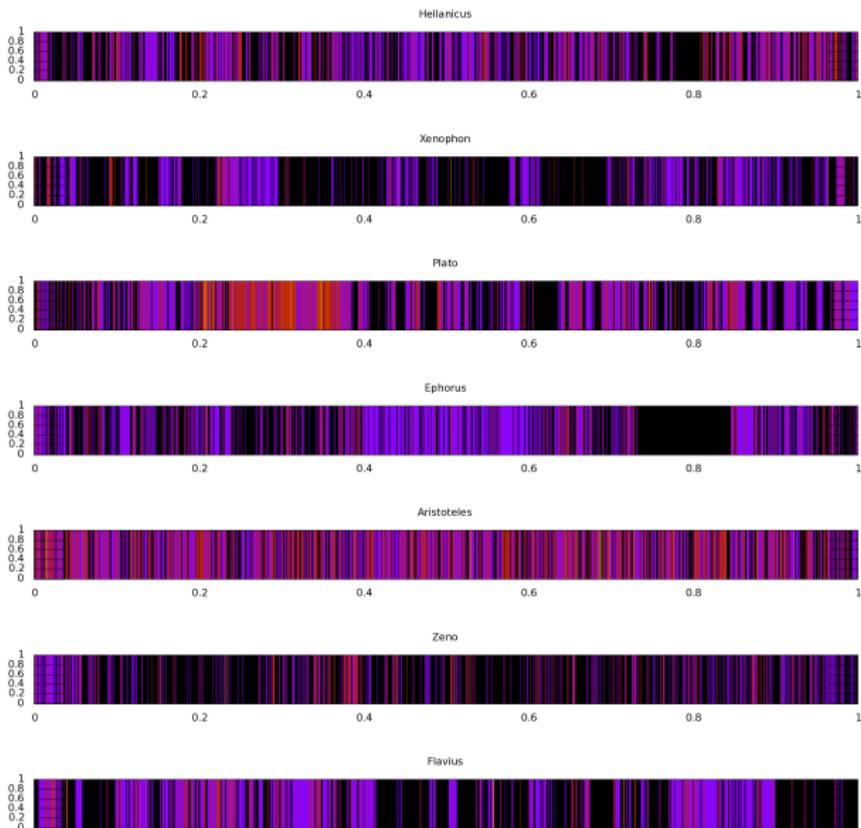
# Tracer

- Text Reuse Detection Software made by Marco Büchler and his eTrap team (incl. Greta Franzini, Emily Franzini, Maria Moritz and students).
- It can manage with Big Data.
- Much faster than human readers/researchers.
- Successful results in English, German, Latin, Greek, Arabic and Pali.



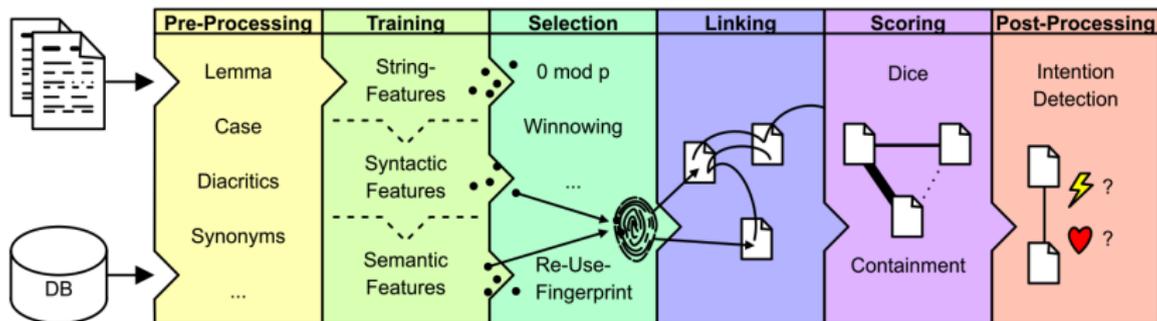


# Text reuses detected by Tracer





# Current approach: TRACER



# Goal

- Text reuse detection in Shenoute and Besa is not complete. Especially, in Shenoute.
- But the text corpus is huge.
- Tracer has successful results in other languages. And it detects text reuses far more quickly than human readers/researchers.

Why not using this Tracer technology on Shenoute and Besa??

# Project on the Coptological side

## SFB 1136 “Bildung und Religion”

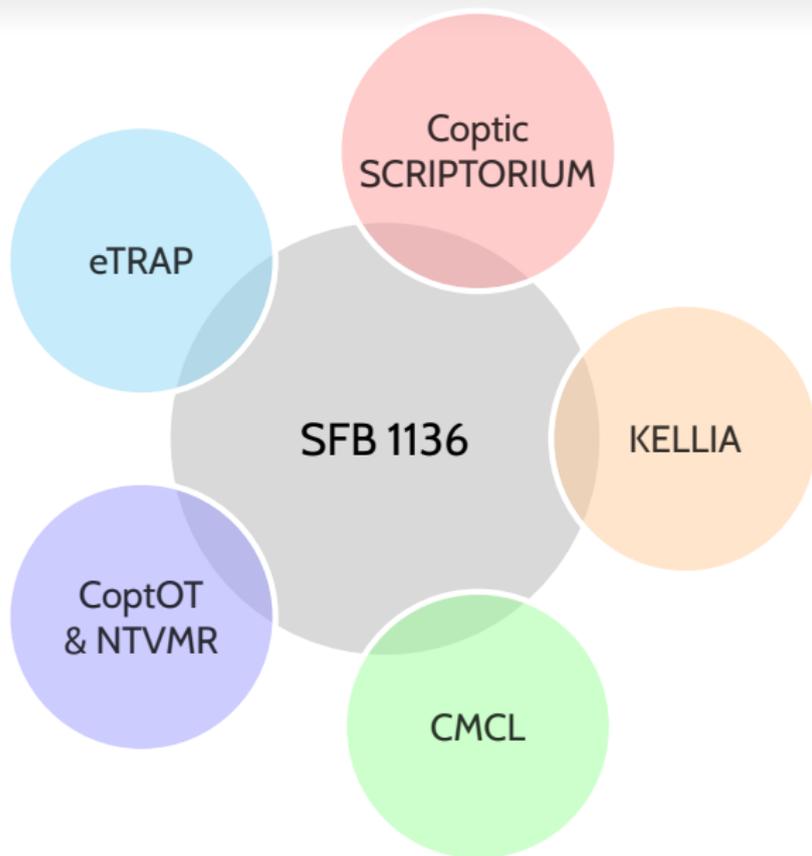
- “Bildung und Religion in Kulturen des Mittelmeerraums und seiner Umwelt von der Antike bis zum Mittelalter und zum Klassischen Islam”

## Teilprojekt BO5: Koptologie

- “Schriftauslegung und Bildungstraditionen im koptischsprachigen ägyptischen Christentum der Spätantike: Shenute, Kanon 6”

## Team member of the Coptology side

- Prof. Dr. Heike Behlmer
- So Miyagawa, M.A.
- Julien Delhez, M.A. M.A.



### Partner projects of KELLIA

- DDGLC
- Thesaurus  
Linguae  
Aegyptiae



**Nationalities of Involved  
People (as far as I know)**



<http://www.cmcl.it/>

Unione Accademica Nazionale — Hiob Ludolf Zentrum, Universität Hamburg — Istituto Patristico Augustinianum

## CMCL

### Corpus dei Manoscritti Copti Letterari

Projet patronné par l'Union Académique International  
CIM server

Direttore: [Tito Orlandi](#)

Commissione scientifica: Edda Bresciani (Pisa), Paola Buzi (Roma), Alberto Camplani (Roma), Stephen Emmel (Münster), Bentley Layton (Yale), Francesco S. Pericoli Ridolfini (Roma), Manlio Simonetti (Roma)

Collaboratori scientifici: [Alin Suciú](#) (Hamburg HLZ), Diliانا Atanassova (Salzburg)

The [CMCL](#) is an enterprise dedicated to the study of Egyptian Christian culture in the I-XII centuries C.E., especially for the documents in Coptic language.

#### *New publications*

**2008** Tito Orlandi, *Coptic Texts Relating to the Virgin Mary. An Overview* [printed and on-line](#)

**2004** Davide Righi, Severiano di Gabala *In Apostolos* cc0331 = cpg4281. Introduzione, testo copto,

#### *Clavis Coptica*

(List of the Coptic Literature)

[Enter](#)

---

***Bibliography Update (experimental)***

# http://copticSCRIPTORIUM.org/

HOME

TOOLS

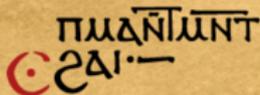
CORPORA

DOCUMENTATION

NEWS

KELLIA

ABOUT



COPTIC  
SCRIPTORIUM

Coptic SCRIPTORIUM is a platform for interdisciplinary and computational research in texts in the Coptic language, particularly the Sahidic dialect. As an open-source, open-access initiative, our technologies and corpus facilitate a collaborative environment for digital research for all scholars working in Coptic. We provide:

- tools to process Coptic texts
- a searchable, richly-annotated corpus of texts using the **ANNIS** search and visualization architecture
- visualizations of Coptic texts
- a collaborative platform for scholars to use and contribute to the project
- research results generated from the tools and corpus

Coptic SCRIPTORIUM is a collaborative, digital

**Latest news:** [\[more\]](#)

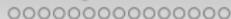
## New feature + texts in our corpora: Apophtegmata, I See Your Eagerness

2016-7-26

We are very excited to release new versions of two of our corpora in time for the Coptic Congress. And keep reading to learn about a new feature on our website. As usually, we provide a diplomatic transcription of the texts' manuscripts, normalized text for ease of reading, and an analytic visualization with the normalized text [...]

[\[link\]](#)

**Coptic SCRIPTORIUM at the Coptic Congress**



# http://coptot.manuscriptroom.com/



## Digital Edition and Translation of the Coptic-Sahidic Old Testament [Sign In](#)



### Welcome

Staff

Project Partners

Steering  
Committee

Manuscript  
Catalog

Workspace

Indexing

Transcribing

Forum

Alin's Blog

About

### CoptOT Bloggers

#### Papyrus Workshop

By **Frank** On 6/24/16 9:58 AM

**Feder,**



On June 2, Myriam Krutzsch, the papyrus conservator of the Berlin Papyrus Collection (Staatliche Museen zu Berlin SPK, Ägyptisches Museum und Papyrussammlung), visited the CoptOT team to give two inspiring lectures on the material aspects of papyrus manuscripts and methods of dating the papyrus material, followed by a practical guide how to use these observations on two examples, the early papyrus codices: the so-called 'Berolinus Gnosticus' (ÄMP P. 8502), and the Akhmimic Proverbs (Berlin, Staatsbibliothek SPK, Berol. orient. oct. 987).



0 Comments

#### Report on the International Conference "Shenoute and the Bible" (Göttingen, May 17-21, 2016)

By **Alin Suciuc**, On 6/21/16 1:56 PM



On May 17-21, 2016, the Coptic research group coordinated by **Heike Behlmer** ([Seminar für Ägyptologie und Koptologie, Universität Göttingen](#)) hosted in Göttingen the international conference "Shenoute and the Bible." The conference was organized on the occasion of the annual meeting of the team that is producing the first critical edition of the works of Shenoute, which this year took place in Göttingen. This international team of scholars is coordinated by **Stephen Emmel** ([Institut für Ägyptologie und Koptologie, Universität Münster](#)). "Shenoute and the Bible" was sponsored and funded by the DFG-Sonderforschungsbereich 1136 "Bildung und Religion," one of whose projects examines the re-use of the Bible in

### Partner Blogs

#### Alin Suciuc

##### An Unusual Sahidic Lectionary Manuscript

7/17/16 7:13 PM

Under the siglum sa 297L, Schmitz and Mink's list of the Sahidic New Testament manuscripts mentions a number of fragments from a lectionary which belonged to the Monastery of Shenoute, i.e. the White Monastery. Schüssler's *Biblia Coptica* designates the same manuscript ... [Continue reading](#) →

##### The Vossen Collection of Coptic Manuscripts

7/16/16 3:41 PM

On June 21, 2016, Tom Vossen, an ancient coins dealer based in Kerkrade, the Netherlands, sent me photographs of five Coptic manuscript fragments which are in his possession. Vossen said that he purchased the fragments ten years ago from a ... [Continue reading](#) →

#### Coptic Scriptorium

##### New publication: Raiders of the Lost

http://ntvmr.uni-muenster.de/

[Sign In](#)

**Welcome**

Manuscript Workspace

NT Transcripts

Forum

Liste

Transcribing

Indexing

Status

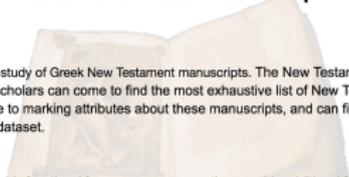
Help

NA28

About

Welcome to the...

## New Testament Virtual Manuscript Room



This site is devoted to the study of Greek New Testament manuscripts. The New Testament Virtual Manuscript Room is a place where scholars can come to find the most exhaustive list of New Testament manuscript resources, can contribute to marking attributes about these manuscripts, and can find state of the art tools for researching this rich dataset.

While our tools are reasonably functional for anonymous users, they provide additional features and save options once a user has created an account and is logged in on the site. For example, registered users can save transcribed pages to their personal account and create personalized annotations to images.

[License Agreement](#)

NTVMR Bloggers

**New entry to the "Liste"**  
By [marie-luise lakmann](#), On 4/8/16 9:44 AM

'GA 2932 (ID 32932) 10th century, greek minuscule of John 10,18-31 1 folio, 1 column, 19... \* [Read More >](#)

---

**Broken CSNTM links**  
By [Klaus Wachtel](#), On 3/8/16 10:42 AM

"We have received several queries recently referring to images hosted by CSNTM. CSNTM changed their... \* [Read More >](#)

---

**Call for Papers: Divine Sonship in Jewish a...**  
By [Garrick Allen](#), On 9/15/15 10:49 AM

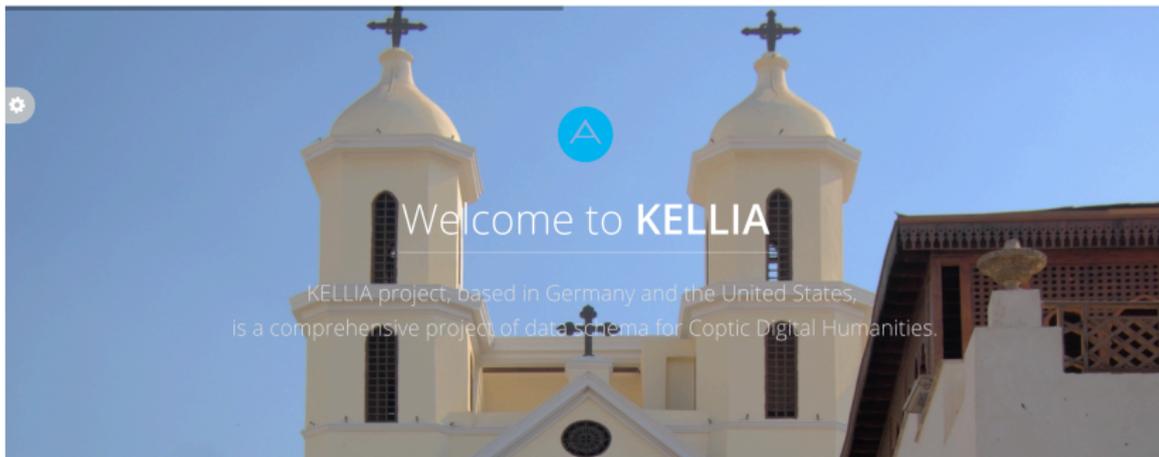
"For the those interested, the St Andrews Symposium for Biblical and Early Christian Studies... \* [Read More >](#)

---

**The ECM at the British New Testament Con...**  
By [Garrick Allen](#), On 9/2/15 10:46 AM

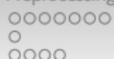


# homepage is under construction by me



## Welcome to KELLIA

KELLIA project, based in Germany and the United States,  
is a comprehensive project of data schema for Coptic Digital Humanities.



<http://research.uni-leipzig.de/ddglc/>



# DDGLC

Database and Dictionary of Greek Loanwords in Coptic

[HOME](#) [Project](#) [Staff](#) [Co-operation](#) [Activities](#) [Contact](#) [Resources](#) [Announcements](#)

## Database and Dictionary of Greek Loanwords in Coptic (DDGLC)



Database and Dictionary of  
Greek Loanwords in Coptic

The results of the project shall be made available in an online database and in a printed dictionary. The core tool of the DDGLC project is currently a relational database designed to connect linguistic and extra-linguistic data concerning types and tokens of all identifiable Greek loanwords in Coptic.

[read more ...](#)

The **DDGLC** project is hosted by the Egyptological Institute -Georg Steindorff- of the University of Leipzig and funded by the Deutsche Forschungsgemeinschaft as a long-term project with a projected lifespan ranging through 2024. The DDGLC project seeks to produce a systematic, comprehensive and detailed lexicographical compilation and description of Greek loanwords as attested in the entire Coptic corpus through every dialect, each kind of text, and in pre-Coptic Egyptian. The project will thus eventually record 1500 years of contact-induced language change of the Egyptian-Coptic language.

hosted by

**UNIVERSITÄT LEIPZIG**

Ägyptologisches Institut / Ägyptisches Museum  
– Georg Steindorff –

funded by

**DFG**

Deutsche  
Forschungsgemeinschaft

supported by

**Gertrud-und-Alexander-  
Böhlig-Stiftung**

**Advisory Board**

Roger Bagnall  
Heike Behlmer  
Anne Boud'hors  
Stephen Emmel  
Wolf-Peter Funk  
Eitan Grossmann  
Martin Haspelmath  
Mark Janse  
Bentley Layton  
Pieter Muysken  
Peter Nagel  
Wolfgang Schenkel  
Ariel Shisha-Halevy

<http://aaew.bbaw.de/>



Strukturen und Transformationen des Wortschatzes der ägyptischen Sprache

## Thesaurus Linguae Aegyptiae



Einführung: Das virtuelle Wörterbuch und die Textdatenbank im Internet \* Kontakt und Abonnement der mailinglist \* Impressum

Im Aufbau des digitalen Corpus altägyptischer Texte kooperieren die folgenden Institutionen und Projekte:

Berlin-Brandenburgische Akademie der Wissenschaften, Strukturen und Transformationen des Wortschatzes der ägyptischen Sprache \* Sächsische Akademie der Wissenschaften zu Leipzig, Strukturen und Transformationen des Wortschatzes der ägyptischen Sprache \* Akademie der Wissenschaften und der Literatur - Mainz, Demotische Textdatenbank \* Nordrhein-Westfälische Akademie der Wissenschaften, Totenbuchprojekt \* Katholieke Universiteit Leuven, Leuven online Index of Ptolemaic and Roman Hieroglyphic Texts \* Akademie der Wissenschaften zu Göttingen, EdFu-Projekt



Der Thesaurus Linguae Aegyptiae wirkt im philologisch-linguistischen Forschungsbereich des Exzellenzclusters TOPOI mit und dankt für vielfache Unterstützung in diesem Kontext.

für die Unterstützung der Arbeit am Thesaurus Linguae Aegyptiae danken wir den folgenden Institutionen:

Deutsche Forschungsgemeinschaft \* Hermann und Elise geborene Heckmann Wentzel-Stiftung \* Fritz Thyssen Stiftung

für die Erlaubnis zur Benutzung von Bildmaterial danken wir den folgenden Institutionen:

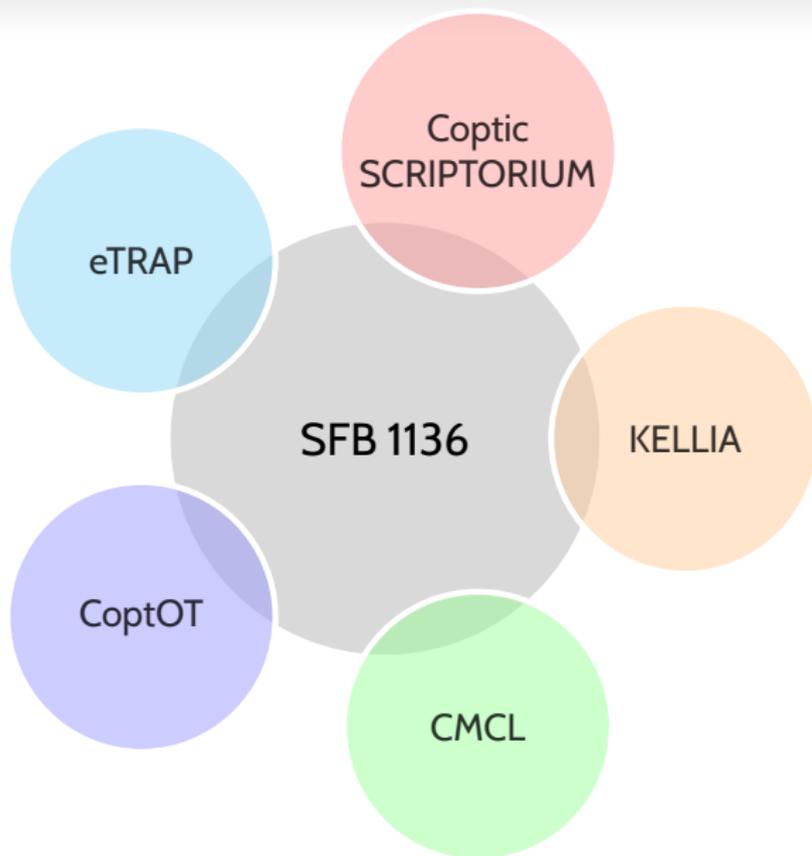
Österreichische Akademie der Wissenschaften (Wien) \* Deutsches Archäologisches Institut (Abteilung Kairo)

weiter danken wir für vielfältige Unterstützung

Jochen Hallof \* Daniel Werning

Stand der Datenbanken: 31. Oktober 2014, Version des Anzeigeprogramms: 31. Oktober 2014

BBAW

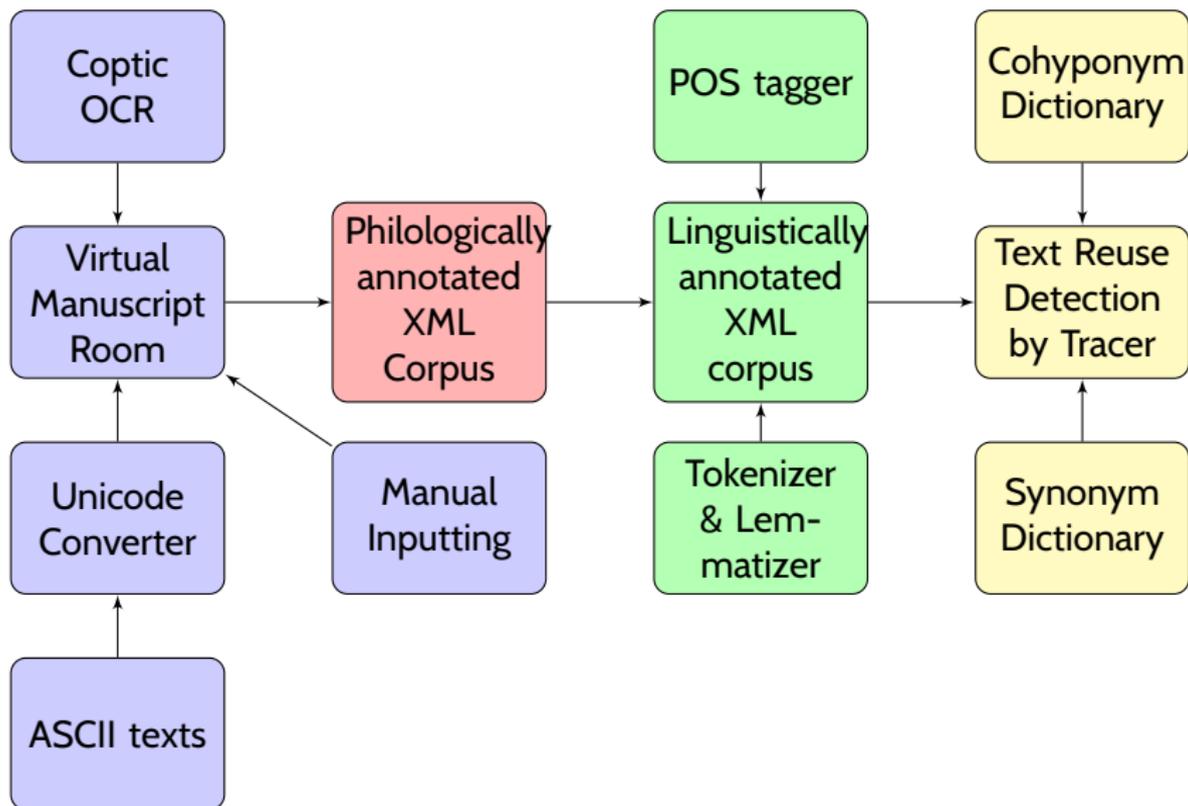


## Partner Projects of KELLIA

- DDGLC
- Thesaurus Linguae Aegyptiae
- Ancient Lives
- and so on...



## workflow



## Introduction

## Preprocessing1

OCR

Converter

Virtual Manuscript Room

## Preprocessing2

Natural Language Processing

## Texts of Shenoute and Besa

## Processing

Text Reuse Detection by Tracer

## What I found

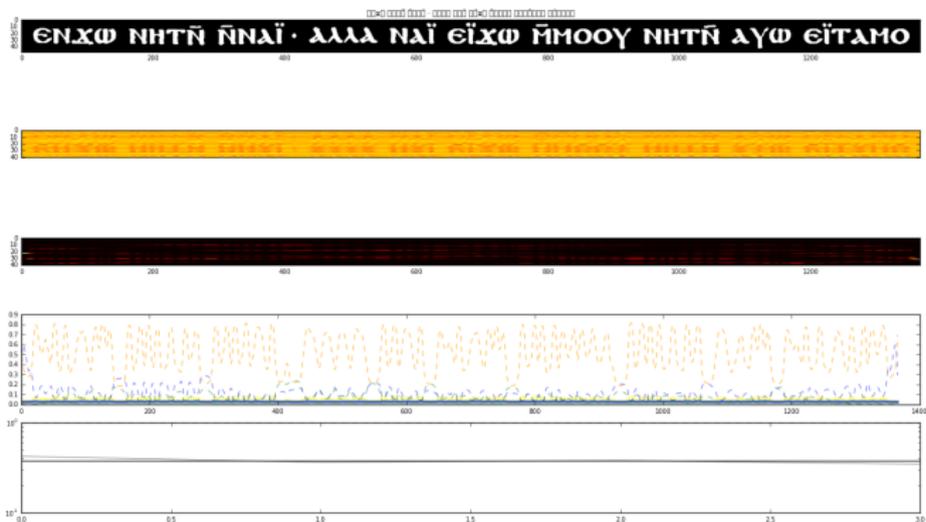
# OCROpy

- <https://github.com/tmbdev/ocropy>
- Massive contribution of Kirill Bulert (Max Planck Institute for Biophysical Chemistry & Göttingen Centre for Digital Humanities)
- Neuronal network model, artificial intelligence
- Deep learning
- Automation already developed
- Training creates models, which can be extended
- Good results, depending on ground truth and image quality

Using GCDH's high-performance computer, 'roedel'



# OCRopy training



# OCRopy result photo

output/0001/010001.bin.png

ΔΟΥΣΜΗ ΨΩΠΙ ΨΑΡΩΟΥ ΕΒΟΛ ΖΕΝΤΦΕ' ΕΣΧΩ ΜΜΟΣ

ΔΟΥΣΜΗ ΨΩΤΠΙ ΨΑΡΩΟΥ ΕΒΟΛ ΖΕΝΤΦΕ ΕΣΧΩ ΜΜΟΣ

output/0001/010002.bin.png

ΧΕΛΥΘΑΨΥΕΝΟΥΤ ΝΑΡΧΗΜΑΝΔΡΙΤΗΣ ΕΠΙΚΟΣΜΟΣ ΤΗΡΥ

ΧΕΛΥΘΑΨΥΕΝΟΥΤ ΝΑΡΧΗΜΑΝΔΡΙΤΗΣ ΕΠΙΚΟΣΜΟΣ ΤΗΡΥ

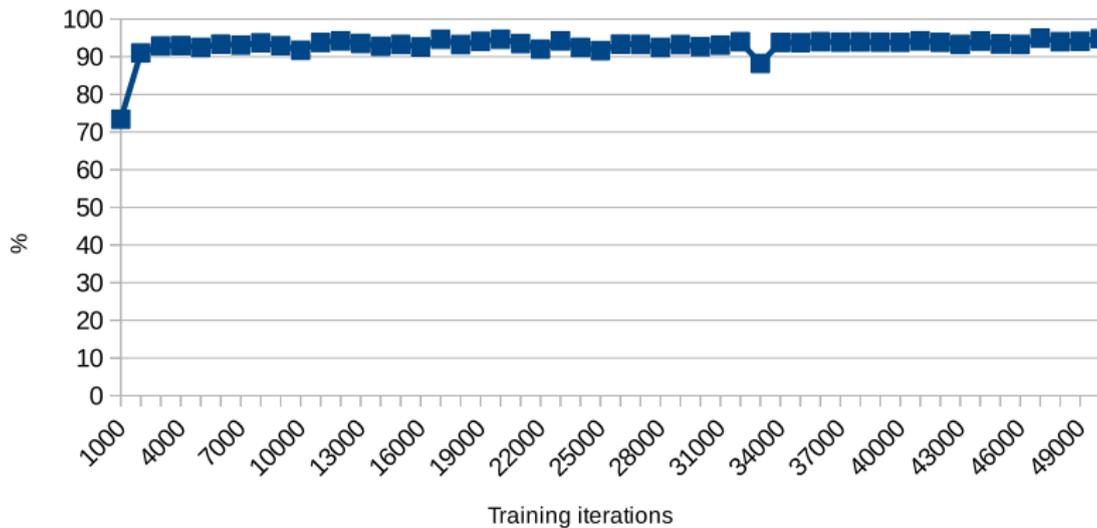
output/0001/010003.bin.png

ΜΦΟΥΥ. ΠΕΧΕΛΠΑ ΠΧΩΛ ΝΑΠΑ ΠΨΟΙ ΧΕΠΑΣΟΝ ΠΨΟΙ

ΜΦΟΥΥ ΠΕΧΕΛΠΑ ΠΧΩΛ ΝΑΠΑ ΠΨΟΙ ΧΕΠΑΣΟΝ ΠΨΟΙ

# Accuracy

## OCR Accuracy

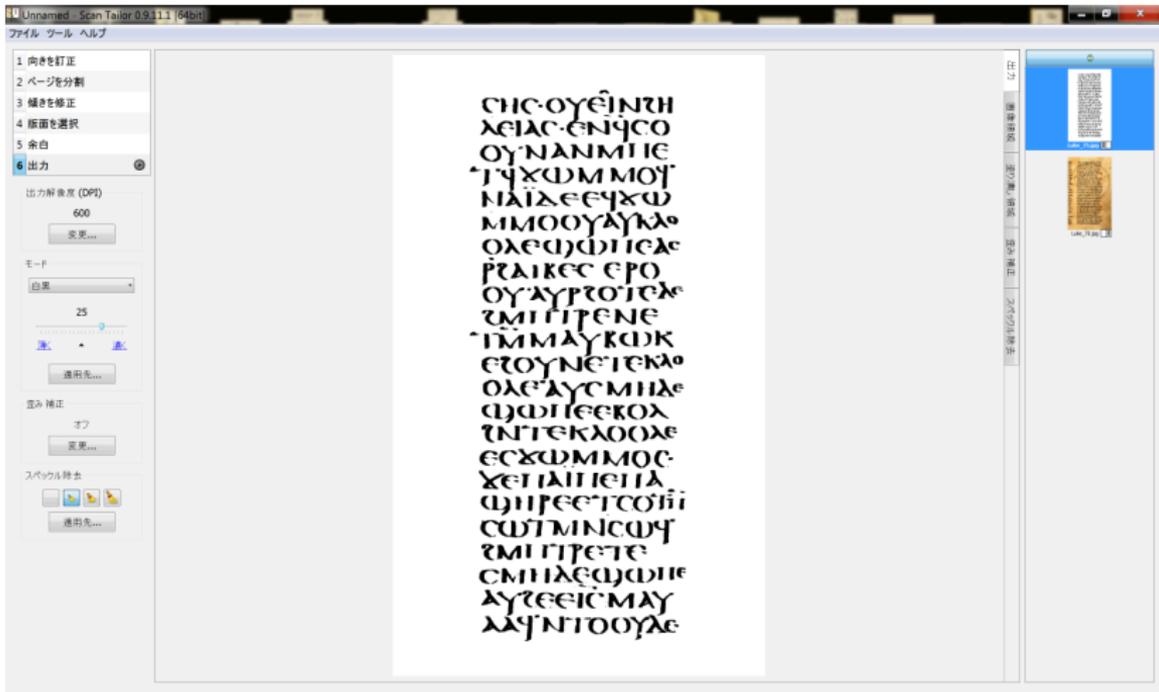


# Preprocessing 1: OCR for handwritten texts



余白の幅を変えるには実線のどれかをドラッグします。

# Preprocessing 2: OCR for handwritten texts



## Teaching: OCR for handwritten texts

output/Luke\_71\_1L/010001.bin.png



ZEEZPEXECE

output/Luke\_71\_1L/010002.bin.png



ZNOYMANXA



# Converter

```
--  
43 Copyright 2013-2015, Amir Zeldes, Caroline T. Schroeder and So Miyagawa  
44  
45 This program is free software. You may copy or redistribute it under  
46 the same terms as Perl itself.  
47 _USAGE_  
48 }  
49  
50 ### OPTIONS BEGIN ###  
51 %opts = ();  
52 getopt('hf:', \%opts) or die $usage;  
53  
54 #help  
55 if ($opts{h} || (@ARGV == 0)) {  
56     print $usage;  
57     exit;  
58 }  
59  
60 #format  
61 if (!$format = $opts{f})  
62     {$format = "Coptic";}   
63 elseif($format eq "CopticLS")  
64     {$format="CopticLS"}
```



# Virtual Manuscript Room

Dashboard   Contacts Center   Microblogs   Messages   My Documents   Tasks   Go To   Admin   Go MySpace

**SFB 1136 'Bildung und Religion' and KELLIA**

Home   Calendar   Documents   Forums   Blogs   Wiki   Members   **Workspace**   Transcribing   Indexing   NLP-Pipe   Colate

Lists Catalog

Manuscripts   Pages   Favorites

Doc ID: 690009   1640

Year	Count	Thumbnail
1650		
1660		
2920 (292)	IV	
2830 (293)		
2940 (294)		
2950 (295)		
2960 (296)		

Image Viewer

Manuscript Transcript

Col 1

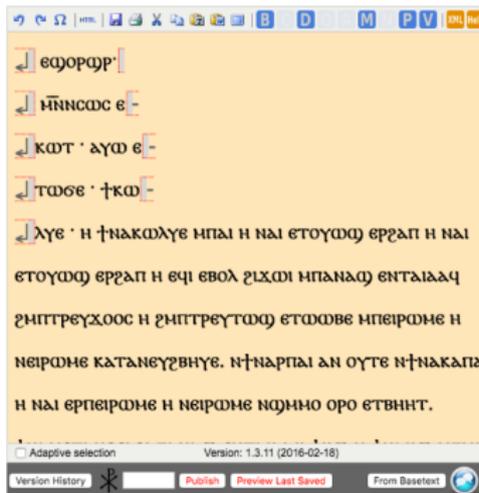
- 1 ΜΗΝ ΟΥΤΩΣΤΙ-
- 2 ΠΑΝΟΛΟΝ ΑΝ
- 3 ΟΥΚ ΕΙΣΙΝ ΟΥΤΟ
- 4 ΟΥΤΕ ΕΙΣΙΝ
- 5 ΟΥΤΩΣ ΕΡΕΘΙ
- 6 ΑΝΤΙΡΕΝΑΡΕ-
- 7 ΔΟΥΝ ΟΥΤΩΣ
- 8 ΗΝ ΠΑΡΕΘΟ-
- 9 ΜΗΤ ΗΡΕΟΥ ΟΥ
- 10 ΚΑΙΕ ΟΥΤΩΣ
- 11 ΤΙΜΕΡΕ ΕΙΣΤ
- 12 ΕΡΕΘΙ ΟΥΤ
- 13 ΤΩ ΒΙΟΥΤΩ
- 14 ΡΑΜΑΚΕ ΕΡΕ
- 15 ΔΕΥΤΕΡΑ ΟΥ
- 16 ΤΩ ΜΕΤΩΝΟ-
- 17 ΤΩ ΕΡΕΘΙ
- 18 ΤΩ ΕΡΕΘΙ
- 19 ΟΥΤΩ ΕΡΕΘΙ

Created by Troy Griffiths & Ulrich Schmid

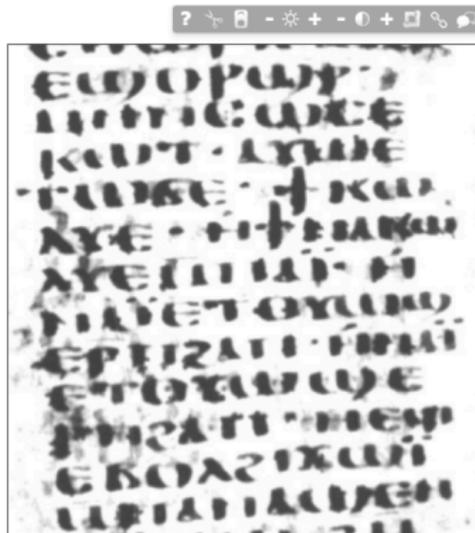
# Transcribe and annotate texts from photos

Easily annotates page/column/line numbers, decoration, size of gaps etc.  
Saved in TEI XML format

## Transcription Editor



## Image Viewer



XM294

# Also it has a collation tool for critical edition

The screenshot displays the Collation tool interface in a web browser. The main area shows a graph of textual variants starting from a root node '#START#'. The graph includes nodes for 'εκληθη', 'δε', 'και', 'null', and 'Δε', with various witness identifiers and scores (e.g., P75 01 02 011 NA28 NA28NS, 02, 03 09 013 017 021 028 030 031 ... (185), sa 1 sa 136 sa 5, sa 9, sa 4, sa 10) indicating the source and weight of each variant. A legend on the right identifies the types of variants: Spelling (green), Itacism (yellow), Abbreviation (orange), Unclear Resolved (blue), and Parallel Segmentation (red).

The control panel at the bottom includes:

- Witnesses:** Ignore Markers (Supplied, Unclear, Punctuation, Normalize Final Sigma), Clear Witnesses.
- Regularization Rules:** Collate (Graph, TEI, Text, Alignment Table, Apperatus Editor).
- Witness Lists:** Algorithm (Dekker, Needleman-Wunsch).
- Context:** John. 2.2, 242 witnesses.

At the bottom, a table shows the collation results for witnesses NA2B and NA28NS:

Witness	Text
NA2B	εκληθη δε και ο ιε και οι μαθηται αυτου εις τον γαμον
NA28NS	εκληθη δε και ο ιε και οι μαθηται αυτου εις τον γαμον

# Photos

- The photos are from CMCL and Biblia Coptica.
- We arranged those photos according Stephen Emmel's reconstruction of Canon 6 and uploaded almost all of them to our server.
- It was done with massive help from Dr. Diliانا Atanassova.

Introduction

Preprocessing1

OCR

Converter

Virtual Manuscript Room

**Preprocessing2**

**Natural Language Processing**

Texts of Shenoute and Besa

Processing

Text Reuse Detection by Tracer

What I found



# Coptic NLP Service

**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](#).

**Input:**

- My data contains meaningful linebreaks
- Ignore linebreaks in my data

**Output:**

- SGML pipeline  Just piped and dashed morphemes
- Stretch milestones
- Tokenize
  - Automatic
  - From pipes in input
- Normalize
- Tag
- Lemmatize
- Language of origin
- Parse

ΑΧΡΩΤΗ ΤΕΤΡΑΔΟΧΗΤΙΖΕ ΖΗΚΟΧΟΣ ΗΜ ΝΗΝΤΩΣ ΟΥΤΕ ΗΠΡΩΤΗ ΟΥΤΕ ΗΠΡΩΤΗ ΕΡΩΤΗ ΕΤΕΜΑ ΗΕ ΔΟΛΩΣ ΕΘΩ ΡΗΝΗΕ ΗΕΚΟΧΟΣ ΕΛΛΗΚΟΥ ΝΟΜΟ ΑΖΩΝ ΟΝ ΕΚΕΤΟ ΗΜΕΝ ΕΠΙΔΟΛΩΝ ΖΥΑ ΠΕΥΤΟΥ ΡΥΛΩΣ ΟΝ ΕΠΙΤΥΧΩΣ ΔΕΜΠΛΟΥΘΩ ΓΑΡ ΝΕΤΕΤΟ ΝΗΛΑ ΠΕ ΤΕΝΟΥ ΔΕ ΝΟΥΩΝ ΖΗΚΟΧΟΣ ΝΟΧΩ ΖΩΣ ΟΑΡΕ ΝΕΖΟΥΩΝ ΕΡΕΠΚΑΡΟΣ ΓΑΡ ΝΠΟΥΩΝ ΕΠΕΤΗΝΟΥΤΩ ΗΝ ΗΠΤΑΚΑΛΩΣΤΗ ΝΗΤΕ ΕΤΕΤΗΛΟΚΑΖΕ ΔΟΛΩ ΠΕΤΡΑΥΩ ΗΕΚΟΧΟΣ ΑΥΟ ΝΕΚΡΩΝΗΕΝ ΕΠΕΜΕΝΕ ΗΠΚΑΚ ΕΤΑΝΕΤΟΥ ΚΑΚΩΣ ΗΤΕΤΗΛΩΘΩ ΔΕ ΗΤΩΚ ΗΡΩΤΟ ΗΕΥΕΡΩΣ ΓΑΡ ΗΝΟΥΟ ΝΑΟΥΕ ΡΕΠΩΛΩ ΗΕ ΕΡΕΠΚΑΩΣ ΟΑΡΩΣΩ ΔΕ ΗΝ ΟΥΤΩΣ ΕΘΩ ΕΤΟΧΑΠΩ ΗΝΟΥΟ ΕΤΗΝΟΥΩΝ ΖΩΑ ΓΑΡ ΗΝ ΕΤΟΥΤΩΣ ΕΘΩ ΖΕΠΟΥΩΝ ΗΕ ΕΤΕΜΑ ΖΩΩΝ ΗΑΡΧΑΠΩ ΗΜΗΝΗΩΤΩΣ ΖΥΑ ΗΡΩΤΗ ΗΜΗΝΕΥΕ ΤΗΡΟΥ ΗΕΛΛΑΘΩΣ ΕΤΡΑΥΩΣΕ ΗΝΟΥΟ ΕΡΩΤΗ ΕΡΩΝ ΗΜΗΝΕΤΑΛΩΣΤΕ ΗΜΗΝΩΣΤΕ ΝΩΕ ΕΚΑΤΤΕΡΗΝΟΥΕ ΕΠΠΩΟΥ ΕΡΩΤΗ ΕΠΗΝΟΥ ΕΠΤΕΤΑΝΗΜΕΡΟΥ ΑΝ ΕΠΠΕΡΩ ΗΑΤΕ ΑΥΟ ΕΠΗΝΕΥΚΕ ΕΠΗΝΟΥ ΑΥΟ ΕΡΩΝ ΔΕ ΗΜΕΝ ΗΜΕΝ ΖΩΣ ΕΡΕΠΚΑΡΟΣ ΕΡΩΠ ΕΡΩΤΗ ΕΠΗΝΟΥ ΑΥΟ

Web-Interface

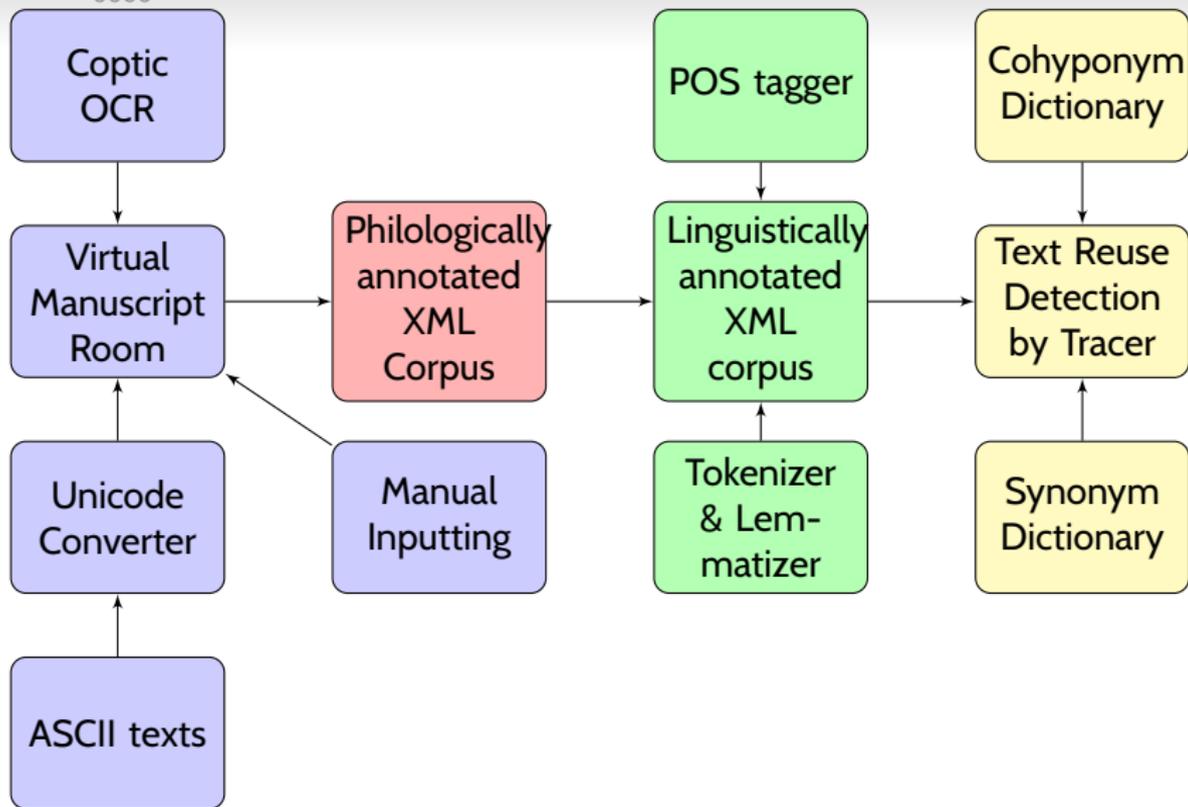
```

1 <nom_group nom_group="λεπρωτ" orig_group="λεπρωτ">
2 <nom.xml.id="u1" pos="VBD" lemma="λερω" func="root" norm="λερω">
3 λερω
4 </nom>
5 <nom.xml.id="u2" pos="PPERS" lemma="λερω" func="subj" norm="τι">
6 τι
7 </nom_group>
8 <nom_group nom_group="τετρηδοχητιζε" orig_group="τετρηδοχητιζε">
9 <nom.xml.id="u3" pos="PPERS" lemma="τετρηδοχητιζε" func="subj" norm="τι">
10 τετι
11 </nom>
12 </nom_group>
13 <nom.xml.id="u4" pos="V" lemma="αοχητιζε" xml:lang="Greek" func="root" norm="αοχητιζε">
14 αοχητιζε
15 </nom>
16 </nom_group>
17 <nom_group nom_group="ζηκωχος" orig_group="ζηκωχος">
18 <nom.xml.id="u5" pos="PREP" lemma="ζη" func="case" head="#u7" norm="τι">
19 η
20 </nom>
21 <nom.xml.id="u6" pos="ART" lemma="τι" func="det" head="#u7" norm="τι">
22 η
23 </nom>
24 <nom.xml.id="u7" pos="N" lemma="κωχος" xml:lang="Greek" func="nmod" head="#u4" norm="κωχος">
25 κωχος
26 </nom>
27 </nom_group>
28 <nom_group orig_group="ημε" norm_group="ημε">
29 <nom.xml.id="u8" pos="PREP" lemma="η" func="case" head="#u10" norm="τι">
30 η

```

Annotated SGML

Coptic NLP Service by Amir Zeldes & Caroline Schroeder



# Coptic Natural Language Processing

## Lemmatizer

Lemmatize words

## POS-tagger

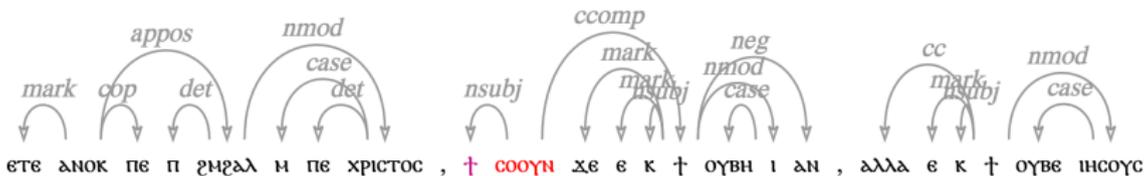
Tagging Part-of-Speech

## TreeBank

Analyzing syntax and making syntactic trees depending on Universal Dependency

Made by our project partners, Amir Zeldes and Caroline Schroeder

# ANNIS View (Cf. <http://corpus-tools.org/annis/>)



ANNIS interface showing search results for the query `norm="τοπος"`. The interface includes a search bar, a list of corpus documents, and a detailed view of the search results.

Query: `norm="τοπος"`

Results: 1 match in 1 document

Corpus List:

Name	Texts	Tokens
apophthegmata.patrum	27	3,011
besa.letters	3	2,297
doc.papyri	3	290
sahidica.1corinthians	12	8,849
sahidica.mark	16	20,190
sahidica.nt	260	218,486

Search Results:

Path: besa.letters > on\_lack\_of\_food (norm\_group 31 - 40)

Annotations (grid):

norm_group	εβωλ	εραϊ	ζηηηεφγνμαγωγη	,	αγω	ηεφτκαηεφτοπος
norm	εβωλ	εραϊ	ζηη	ηεφ	αγωμαγωγη	,
pos	ADV	ADV	PREP	PPOS	N	PUNCT
lemma	εβωλ	εραϊ	ζηη	ηεφ	αγωμαγωγη	,
orig_group	εβωλ	εραϊ	ζηηηεφγνμαγωγη	,	αγω	ηεφτκαηεφτοπος
orig	εβωλ	εραϊ	ζηη	ηεφ	αγωμαγωγη	,
lang					Greek	
translation	Hence, let us all pray, everyone who fears God, that Christ may put his blessing and that it may remain in his congregations (places) lack anything.					
p_n	L2					
pb_xml_id	BB553					
cb_n	1					
lb_n	24	25	26	27	28	
tok	εβωλ	εραϊ	ζηη	ηεφ	αγωμαγωγη	,

Additional annotations: αγω ηεφτκαηεφτοπος εφωδωτ ηδααγ :

Annotations (grid):

norm_group	εβωλ	εραϊ	ζηηηεφγνμαγωγη	,	αγω	ηεφτκαηεφτοπος
norm	εβωλ	εραϊ	ζηη	ηεφ	αγω	ηεφτκαηεφτοπος
pos	ADV	ADV	PREP	PPOS	N	PUNCT
lemma	εβωλ	εραϊ	ζηη	ηεφ	αγω	ηεφτκαηεφτοπος
orig_group	εβωλ	εραϊ	ζηηηεφγνμαγωγη	,	αγω	ηεφτκαηεφτοπος
orig	εβωλ	εραϊ	ζηη	ηεφ	αγω	ηεφτκαηεφτοπος
lang					Greek	Greek

Annotations (grid):

norm_group	εβωλ	εραϊ	ζηηηεφγνμαγωγη	,	αγω	ηεφτκαηεφτοπος
norm	εβωλ	εραϊ	ζηηηεφγνμαγωγη	,	αγω	ηεφτκαηεφτοπος
pos	ADV	ADV	PREP	PPOS	N	PUNCT
lemma	εβωλ	εραϊ	ζηηηεφγνμαγωγη	,	αγω	ηεφτκαηεφτοπος
orig_group	εβωλ	εραϊ	ζηηηεφγνμαγωγη	,	αγω	ηεφτκαηεφτοπος
orig	εβωλ	εραϊ	ζηηηεφγνμαγωγη	,	αγω	ηεφτκαηεφτοπος
lang					Greek	Greek

Annotations (grid):

norm_group	εβωλ	εραϊ	ζηηηεφγνμαγωγη	,	αγω	ηεφτκαηεφτοπος
norm	εβωλ	εραϊ	ζηηηεφγνμαγωγη	,	αγω	ηεφτκαηεφτοπος
pos	ADV	ADV	PREP	PPOS	N	PUNCT
lemma	εβωλ	εραϊ	ζηηηεφγνμαγωγη	,	αγω	ηεφτκαηεφτοπος
orig_group	εβωλ	εραϊ	ζηηηεφγνμαγωγη	,	αγω	ηεφτκαηεφτοπος
orig	εβωλ	εραϊ	ζηηηεφγνμαγωγη	,	αγω	ηεφτκαηεφτοπος
lang					Greek	Greek

Introduction

Preprocessing1

OCR

Converter

Virtual Manuscript Room

Preprocessing2

Natural Language Processing

**Texts of Shenoute and Besa**

Processing

Text Reuse Detection by Tracer

What I found

# Texts

## Main Text at the Current Stage

- Shenoute's Canon 6
- Kuhn's edition of Besa
- Acephalous Work A26 by Shenoute
- → next stage, other texts from White Monastery, eg. other Shenoutean texts, Apa John etc.

## Compared text

- Sahidic Bible (OT & NT)
- → next stage, Greek authors and other Coptic texts.



## Main topics of Canon 6

Accusations against Shenoute of excessive force (and his defense) (He Who Sits upon His Throne; Remember, O Brethren; Is It Not Written)

Shenoute's illness (Remember, O Brethren; Is It Not Written; Then I Am Not Obligated)

Affairs of the female community (He Who Sits upon His Throne; Then I Am Not Obligated; People Have Not Understood)

Monastic rules

Behlmer (2008)

## Introduction

## Preprocessing1

OCR

Converter

Virtual Manuscript Room

## Preprocessing2

Natural Language Processing

## Texts of Shenoute and Besa

## **Processing**

**Text Reuse Detection by Tracer**

## What I found

## Introduction

## Preprocessing1

OCR

Converter

Virtual Manuscript Room

## Preprocessing2

Natural Language Processing

## Texts of Shenoute and Besa

## **Processing**

**Text Reuse Detection by Tracer**

## What I found

# Corpus size

ca. 500,000 words (tokens)

## Abbots' Works

- Canon 6
- Kuhn's Besa
- Acephalous Work A26

## Bible Texts

- Sahidic Bible (OT & NT)

# Corpus size

texts	tokens
Canon6	48,394
Besa	61,397
Sahidic Bible	421,344



# Starting TRACER

**Start the tool with the command:**

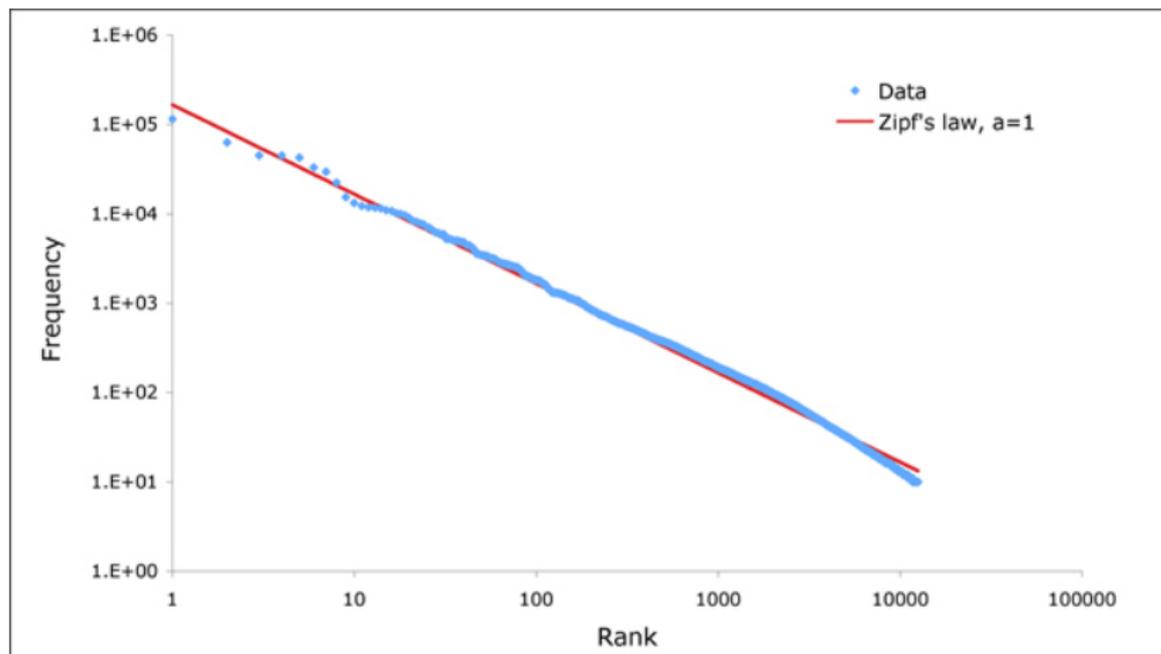
```
java -Xmx600m  
-Dde.gcdh.medusa.config.ClassConfig=conf/tracer_config.xml  
-jar tracer.jar
```

## Explanation:

- `-Xmx600m` (up to 600 MB memory);
- `-Dfile.encoding` sets the encoding of your input file (optionally);
- `-Dde.gcdh.medusa.config.ClassConfig` (configuration file).



# Foundations for preprocessing: Zipfian Law



## Implications of the Zipfian Law

- **Approx. 50%** of all words occur **only once**
- **Approx. 16%** of all words occur **only twice**
- **Approx. 8%** of all words occur **three times**
- ...
- **Approx. 90%** of all words in a corpus occur **10 times or less**

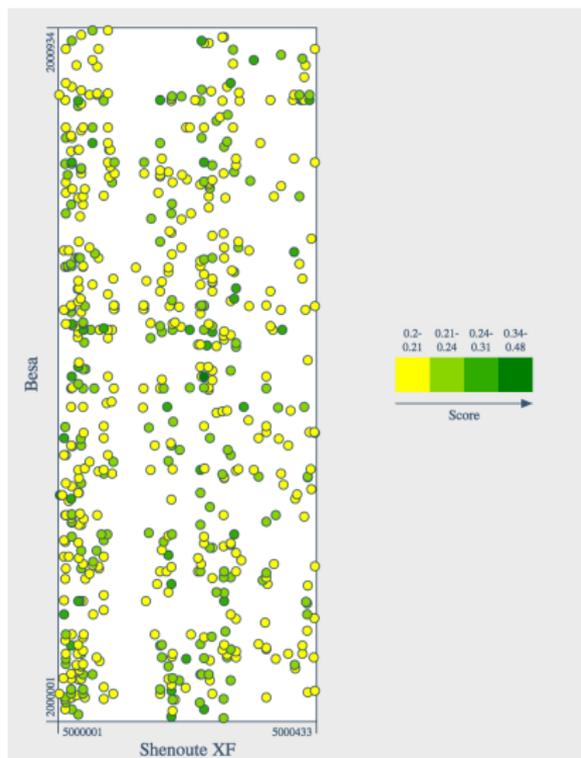
$$s(f) = \frac{1}{f * (f + 1)}$$

$$s^n(f) = \sum_{f=1}^n \frac{1}{f * (f + 1)}$$

- The **top 300 - 700 most** frequent words cover already about **50%** of all tokens (depending language)



# TRAViz by Stefan Jänicke



- x-axis = Shenoute XF
- y-axis = Besa
- left and bottom = the start of the texts
- right and top = the end of the texts
- circle = similar texts
- green = more probable text reuse candidate
- yellow = less probable text reuse candidate



# TRAViz by Stefan Jänicke

13835 Text Re-uses



## Sahidic Bible

<p>ϣⲁⲣⲉ ⲗⲁⲃ ⲁⲛⲏ ⲟⲩ ⲁⲛⲁⲗⲉ ⲉⲃⲟⲗ ⲉⲧ ⲟⲩ ⲧⲣⲓⲛⲟ ⲛⲁⲟ ⲟⲩ ⲗⲓⲧⲛⲓ ⲡⲓ ⲟⲩⲟⲩⲛⲓ ⲗⲁⲃⲛ ⲓⲁⲣ ⲛⲁⲛ ⲉⲧ ⲟⲩⲟⲩⲗⲉ ⲉⲃⲟⲗ ⲗⲁⲃ ⲟⲩⲟⲩⲛⲓ ⲛⲉ</p> <p>ⲉⲧⲃⲉ ⲡⲁⲓ Ⲡ ⲛⲁⲟ ⲛⲁⲟ ⲛⲁⲟ ⲧⲁⲃⲟⲩⲛⲓ ⲡⲓ ⲉⲧ ⲛⲉⲕⲟⲧⲛⲓ ⲛ ⲓ ⲗⲟ ⲟⲩⲧⲉ ⲛ ⲉⲧ ⲛⲟⲟⲩⲧ ⲧⲁⲣⲉ ⲡⲉⲕⲱ ⲟⲩⲟⲩⲛⲓ ⲉⲣⲟ ⲕ</p> <p>ⲁⲟⲩⲟⲩⲧ ⲃⲉ ⲗⲁⲛ ⲟⲩ ⲁⲣⲁⲧⲁⲥⲉ ⲉ ⲧⲉⲧⲛⲓ ⲛⲟⲟⲩⲣⲉ ⲛ ⲁⲩⲟ ⲛ ⲗⲉ ⲗⲁⲥ ⲁⲛⲟⲧ ⲁⲛ ⲁⲗⲗⲁ ⲗⲁⲥ ⲁⲃⲁⲛⲉ</p> <p>ⲉ ⲧⲉⲧⲛⲓ ⲟⲩⲁⲛⲓ ⲛ ⲉ ⲡⲉⲟⲩⲟⲩⲩⲥⲁ ⲛⲉ ⲉⲗⲟⲟⲩⲩⲁ ⲗⲟⲟⲩⲩⲁ</p> <p>ⲉⲧⲃⲉ ⲡⲁⲓ ⲛⲉⲣ ⲟⲩⲁⲃⲉ ⲛⲁ. ⲉ ⲛⲧ ⲁⲗⲗⲁ. ⲉ ⲓ ⲛⲉ ⲗⲉ ⲟⲩ ⲡⲉ ⲡⲓ ⲟⲩⲟⲩⲩⲥⲁ ⲛ ⲡⲓ ⲧⲟⲩⲛⲥ</p> <p>ⲁⲩⲟⲩ ⲛⲉⲣ ⲓⲗⲉ ⲗⲁⲛ ⲟⲩ ⲛⲉⲣ ⲡⲁⲓ ⲉⲧⲃⲟⲩⲛⲓ ⲟⲩⲛⲧⲓⲛⲁ. ⲟⲩⲟⲩⲛⲓ ⲛⲉⲗⲓⲧ Ⲡ ⲁⲗⲗⲁ. ⲛ ⲧⲉⲧⲛⲓ ⲧⲁⲃⲉ ⲉⲃⲟⲗ ⲗⲁⲃ ⲡⲉⲧⲛⲁ.</p> <p>ⲉ ⲧⲉⲧⲛⲓ ⲟⲩⲁⲧⲉ ⲛⲛⲓ ⲛⲉⲧⲛⲓ ⲉⲣⲛⲟⲩ ⲗⲁⲃ ⲗⲁⲃ ⲡⲁⲗⲗⲟⲥ ⲛⲛⲓ ⲗⲁⲃ ⲟⲩⲟⲩⲩⲁ ⲛⲛⲓ ⲗⲁⲃ ⲁⲛ ⲛⲉⲧⲛⲓ ⲗⲓⲧⲉ ⲉ ⲡⲓ ⲧⲟⲩⲛⲥ</p> <p>ⲉ ⲧⲉⲧⲛⲓ ⲟⲩⲡⲓ ⲗⲁⲛⲟⲩ ⲛ ⲟⲩⲟⲩⲩⲥⲁ ⲛⲛⲓ ⲗⲁ. ⲟⲩⲟⲩⲛⲓ ⲛⲛⲓ ⲗⲁⲃ ⲡⲓ ⲣⲁⲛ ⲛ ⲡⲉⲛ ⲧⲟⲩⲛⲥ ⲛⲁⲛ ⲡⲉⲕⲱ ⲛⲧⲛⲁⲛⲟⲩⲧⲉ ⲡⲓ ⲉⲃⲟⲩⲧ</p> <p>ⲉ ⲧⲉⲧⲛⲓ ⲗⲓⲛⲧⲟⲩⲗⲁⲃⲟⲥ ⲛ ⲛⲉⲧⲛⲓ ⲉⲣⲛⲟⲩ ⲗⲁⲃ ⲟⲩ ⲟⲩⲧⲉ ⲛⲛⲛⲟⲥ</p> <p>ⲛⲉⲗⲓⲟⲩⲛⲉ ⲛⲁⲣⲟⲩⲗⲓⲛⲧⲟⲩⲗⲁⲃⲟⲥ ⲛ ⲛⲉⲗⲓ ⲗⲁⲃ ⲛ ⲉ ⲟⲩ ⲛ ⲧⲟⲩⲛⲥ</p> <p>ⲧⲉ ⲡⲓ ⲗⲟⲟⲩⲩⲥⲁ ⲧ ⲁⲃⲉ ⲡⲉ ⲛ ⲧⲉ ⲉⲗⲁⲛⲉ ⲛ ⲉ ⲟⲩ ⲛⲉⲛⲟⲥ ⲉ ⲧ ⲁⲃⲉ ⲡⲉ ⲛ ⲧ ⲉⲕⲗⲁⲛⲥⲁ ⲛⲧⲟⲩⲡⲓ ⲡⲉ ⲡⲓ ⲁⲟⲩⲛⲧⲉ ⲡⲓ ⲡⲓ ⲟⲩⲁⲛⲁ.</p> <p>ⲁⲗⲗⲁ. ⲛ ⲉ ⲉⲧⲉⲣⲉⲧⲉⲕⲗⲁⲛⲥⲁ. ⲗⲓⲛⲧⲟⲩⲗⲁⲃⲟⲥ ⲛⲉⲛⲟⲥ ⲧⲁⲓ ⲗⲁⲃⲟⲩ Ⲡ ⲧⲉ ⲟⲩ ⲉ ⲛⲁⲣⲉ ⲛⲉⲗⲓⲟⲩⲛⲉ ⲗⲁⲃ ⲟⲩ ⲗⲓⲛⲧⲟⲩⲗⲁⲃⲟⲥ ⲛ ⲛⲉⲗⲓ ⲗⲁⲃ ⲗⲁⲃ ⲗⲁⲃ ⲛⲛⲓ</p> <p>ⲛ ⲗⲟⲟⲩⲩⲥⲁ ⲛⲉⲣⲉ ⲛⲉⲧⲛⲓ ⲗⲓⲟⲩⲛⲉ ⲕⲁⲧⲁ. ⲉ ⲟⲩ ⲉ ⲛ ⲧ ⲁ. ⲡⲉⲕⲱ ⲛⲉⲣⲉ ⲧ ⲉⲕⲗⲁⲛⲥⲁ ⲛⲉⲣⲉ ⲁⲩⲟⲩ ⲁ Ⲡ ⲧⲁⲃ. Ⲡ ⲗⲁⲣⲟ ⲥ</p> <p>ⲧⲉ ⲉⲃⲉ ⲧⲣⲟⲟ ⲥ ⲉ ⲁ Ⲡ ⲕⲁⲃⲁⲣⲓⲗⲉ ⲛⲁⲟ ⲥ ⲗⲁⲃ ⲡⲓ ⲟⲩⲁⲧⲉ ⲛⲛⲓ ⲡⲓ ⲧⲟⲩⲛⲥ ⲛ ⲡⲓ ⲛⲟⲟⲩ</p> <p>ⲧⲉⲕⲁⲥ ⲉⲃⲉ ⲧⲁⲗⲉ ⲧ ⲉⲕⲗⲁⲛⲥⲁ. ⲛ Ⲡ ⲉⲣⲁⲧ ⲥ ⲉ ⲧ ⲁⲛⲟⲩ ⲉⲛⲓ ⲁⲗⲗⲁ ⲛ ⲧⲟⲩⲗⲁⲃ ⲛⲉⲗⲓⲧ ⲥ ⲛ ⲧⲟⲩⲛⲓ ⲛ ⲁⲗⲗⲁ ⲛ ⲗⲁⲃ ⲛ ⲗⲁⲃ ⲛⲧⲟⲩⲗⲁⲃ ⲁⲗⲗⲁ. ⲧⲉⲕⲁⲥ ⲉⲟⲩ ⲟⲩⲁⲃⲉ ⲛ ⲥ ⲟⲩⲁⲃⲁ ⲁⲩⲟⲩ ⲛ ⲁⲧⲧⲟⲩⲁⲛ</p>	<p>1012457</p> <p>1012458</p> <p>1012459</p> <p>1012460</p> <p>1012461</p> <p>1012462</p> <p>1012463</p> <p>1012464</p> <p>1012465</p> <p>1012466</p> <p>1012467</p> <p>1012468</p> <p>1012469</p> <p>1012470</p> <p>1012471</p>
---	--

## Besa

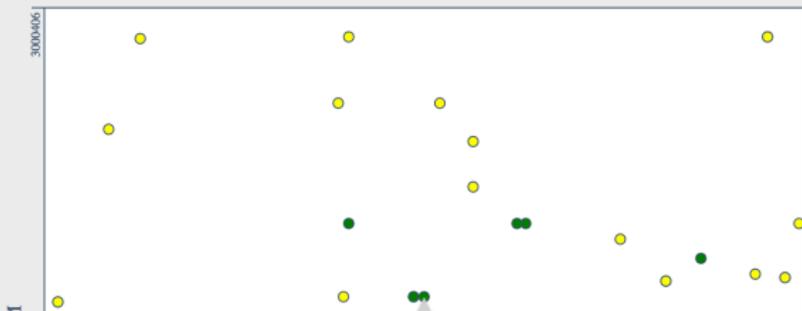
<p>ⲛⲉⲧⲟⲩⲉⲣⲉ ⲓⲁⲣ ⲛⲁⲟ ⲟⲩ ⲛⲧⲟⲩⲉ ⲗⲁⲃ ⲡⲁⲣⲟⲩ ⲛ ⲉⲣⲉⲕⲗⲁⲥⲟⲩⲟⲩ ⲟⲩⲁⲣⲉ ⲗⲁⲃ ⲁⲛ ⲛⲁⲟ ⲟⲩ ⲁⲛⲁⲗⲉ ⲉⲧ ⲟⲩ ⲧⲣⲓⲛⲟ ⲛⲁⲟ ⲟⲩ ⲗⲓⲧⲛⲓ ⲡⲓ ⲟⲩⲟⲩⲛⲓ</p> <p>ⲗⲁⲃⲛ ⲓⲁⲣ ⲛⲁⲛ ⲉⲧ ⲟⲩⲟⲩⲗⲉ ⲉⲃⲟⲗ ⲗⲁⲃ ⲟⲩⲟⲩⲛⲓ ⲛⲉ</p> <p>ⲉⲧⲃⲉ ⲡⲁⲓ ⲡⲁⲗⲟⲩ ⲛ ⲛⲁⲣⲟⲩⲛⲟⲩ ⲛ ⲛⲉⲛ ⲛⲉⲧⲱ ⲁⲟⲩ ⲗⲁⲃ ⲛⲉⲧⲛⲓ ⲛ ⲛⲛⲓ ⲛ ⲛⲉⲟⲩⲉ ⲧⲣⲓⲟⲩ ⲛ ⲡⲓ ⲡⲁⲗⲗⲟⲥ ⲉⲧ Ⲡ ⲛⲟⲩⲧⲉ ⲛⲁⲟ ⲟⲩ ⲉⲗⲟⲩⲛⲓ ⲉⲣⲟ ⲛ ⲛⲛⲓ ⲛ ⲛⲉⲧ ⲁⲧ ⲛⲁⲗⲧⲉ ⲛⲛⲓ ⲛ ⲛⲟⲩⲧⲉ ⲛ ⲉ ⲟⲩ ⲉ ⲉⲩⲁⲩⲧⲣⲉⲛⲛⲉⲩⲉ ⲉ ⲡⲓ ⲡⲉⲟⲩⲟⲩ ⲉⲗⲟⲩⲛⲓ ⲉⲛⲉ ⲛ ⲉⲣⲛⲟⲩ ⲉⲛⲧⲓⲛⲉⲗⲧⲁⲛⲛⲉⲣⲛⲟⲩ ⲁⲛ ⲉⲧⲛⲧⲓⲣⲡⲓ ⲛⲁⲗⲧⲉ</p> <p>ⲁⲩⲟⲩ ⲉ ⲛ ⲉⲛⲁⲣⲓⲕⲉ ⲉⲛⲉ ⲛ ⲉⲣⲛⲟⲩ ⲁⲩⲟⲩ ⲉⲣⲟ ⲛ ⲁⲛ ⲛⲉⲛⲓ ⲛⲁⲟ ⲛ ⲗⲁⲃ ⲉⲣⲉ ⲡⲉ ⲛ ⲛⲉⲩⲉ ⲟⲩⲟⲩⲛⲓ ⲉⲛⲉ ⲛ ⲉⲣⲛⲟⲩ ⲁⲩⲟⲩ ⲉ Ⲡ ⲟⲩⲟⲩⲛⲓ ⲁⲛ ⲟⲩⲟⲩⲩⲥⲁ ⲟⲩⲉⲛⲕⲟⲥⲟⲛⲕⲣⲁⲧⲟⲣ ⲛ ⲡⲓ ⲕⲁⲃⲉ ⲟⲩⲉⲛⲛⲉⲛⲧⲛⲟⲩⲛⲓ ⲛ ⲧ ⲡⲉⲛⲧⲁⲣⲁ. ⲉⲧⲗⲁⲛⲧⲉⲣⲉ ⲁⲩⲟⲩ ⲟⲩⲉⲛⲟⲩⲟⲩⲓ ⲁⲛ ⲗⲓ ⲁⲣⲁ</p> <p>ⲛ ⲉ ⲟⲩ ⲉ ⲛⲧ ⲁ. ⲡⲓ ⲁⲛⲟⲩⲗⲁⲃⲟⲥ ⲧⲟⲟ ⲥ ⲧⲉ ⲗⲓ ⲛⲓ ⲧⲛ ⲛ ⲧ ⲡⲁⲛⲗⲟⲩⲁⲗⲁ. ⲛ ⲡⲓ ⲛⲟⲩⲧⲉ ⲧⲟⲩⲉⲧⲛⲁⲩⲟⲩⲛⲟⲩⲛⲓ ⲉ ⲁⲗⲉ ⲣⲁⲧ ⲧⲛⲧⲓⲧⲓ ⲛⲛⲁⲗⲣⲛ ⲛ ⲕⲟⲧⲛ ⲛ ⲡⲓ ⲡⲁⲗⲗⲟⲥ ⲧⲉ ⲉⲣⲉ ⲡⲉ ⲛ ⲛⲉⲩⲉ ⲟⲩⲟⲩⲛⲓ ⲛ ⲁⲛ ⲁⲛ ⲟⲩⲉⲛⲟⲩⲟⲩⲓ ⲗⲓ ⲁⲣⲁ</p> <p>ⲉⲧⲃⲉ ⲡⲁⲓ ⲟⲩ ⲣⲟⲩⲥ ⲁⲩⲟⲩ ⲛ ⲧⲓ ⲁⲛⲁⲗⲧⲉ ⲛ ⲡⲓ ⲡⲉⲧ ⲛⲁ ⲛⲟⲩ Ⲡ ⲉⲃⲟⲗ ⲧⲉ ⲟⲩⲛⲓ ⲗⲁⲗ ⲉⲩⲗⲁⲛⲧⲉⲣⲛⲉⲣⲛⲟⲩ ⲗⲁⲃ ⲡⲓ ⲟⲩⲁⲛⲁ. ⲉ ⲩ ⲟⲩⲛⲟⲩ ⲁⲗⲉ ⲉⲃⲟⲗ ⲛ ⲛⲉⲗⲓ ⲉⲣⲛⲟⲩ ⲗⲁⲃ ⲡⲓ ⲗⲓⲧⲉ ⲁⲩⲟⲩ ⲧ ⲁⲗⲗⲁⲛⲓ ⲟⲩⲛⲓ ⲗⲁⲗ ⲁⲗⲉ ⲟⲩ ⲉⲩⲗⲁⲛⲧⲉⲣⲛⲉⲣⲛⲟⲩ ⲁⲛ ⲗⲁⲃ ⲡⲓ ⲟⲩⲁⲛⲁ. ⲉⲩⲗⲁⲛⲧⲉⲣⲛⲉⲣⲛⲟⲩ ⲛ ⲛ ⲁⲩⲟⲩ ⲛⲛⲓ ⲗⲁⲃ ⲡⲓ ⲁⲗⲗⲁⲛⲓ ⲛⲉⲛⲟⲥ: ⲛⲛⲓ ⲧ ⲁⲗⲗⲁⲛⲓ ⲛⲉⲣⲉ ⲡⲓ ⲛⲁ. ⲁⲩⲟⲩ ⲡⲓ ⲟⲩⲟⲩⲩⲥⲁ ⲛ ⲁⲩⲟⲩ</p> <p>ⲛⲧⲟⲩⲧⲁ ⲁⲛ ⲉ ⲛⲓⲛⲟⲩ ⲗⲁⲣⲉⲗ ⲁⲩⲟⲩ ⲛ ⲟⲩ ⲛ ⲉⲃⲟⲗ ⲗⲁⲃⲉⲛⲧⲁⲛⲧⲁⲃⲟⲩⲁⲃⲟⲥ ⲟⲩⲟⲩ ⲁⲩⲟⲩ ⲟⲩⲛⲓ ⲗⲁⲗ ⲛⲁ. ⲧⲟⲩⲟⲩ ⲁⲩⲟⲩ ⲛ ⲟⲩ ⲉ ⲧ ⲗⲓⲛⲟⲩ ⲉⲃⲟⲗ ⲛⲁⲛⲟⲩ ⲧⲛ ⲛ ⲉ ⲟⲩ ⲉ ⲛⲧ ⲁ. Ⲡ ⲧⲟⲟ ⲥ ⲧⲉⲛⲁⲣⲟⲩⲟⲩⲟⲩⲟⲩ ⲉⲛⲉ ⲧ ⲗⲓⲛⲟⲩ ⲉⲧ ⲛⲁⲛⲟⲩ ⲟⲩ ⲛ ⲟⲩ ⲉ ⲧ ⲁⲩⲟⲩ ⲛ ⲡⲉⲧⲛⲓ ⲉⲃⲟⲗ ⲉⲧ ⲗⲁⲃ ⲛ ⲡⲉⲣⲉ</p> <p>ⲛ ⲉ ⲟⲩ ⲉ ⲣⲁⲣ ⲉⲛⲧ ⲁ. ⲧⲉⲧⲛⲓ ⲟⲩⲉⲗ ⲧⲓⲛⲧⲛⲓ ⲛⲁⲗⲁ ⲛ ⲉⲧ ⲟⲩⲁⲃⲁ ⲧⲣⲓⲟⲩ ⲧⲉ ⲗⲓⲛⲉⲧⲛⲉⲧⲛⲁⲃⲁⲃⲉⲥ ⲛ ⲉ ⲟⲩ ⲉ ⲛⲧ ⲁ. Ⲡ ⲧⲟⲟ ⲥ ⲧⲉ ⲛⲧⲟⲕ ⲁⲗⲉ ⲁ. ⲕ ⲟⲩⲁⲗ ⲕ ⲛⲁⲗ. ⲧⲁ ⲟⲩⲟⲩ ⲡⲁ. ⲁⲛⲟⲩⲧ ⲡⲓ ⲁⲩⲟⲩⲩⲥⲁ ⲧⲁ. ⲡⲉⲧⲓⲧⲥ ⲧⲁ. ⲛⲉⲧ ⲗⲁⲣⲟⲩ ⲗⲓⲧ ⲧⲁ. ⲁⲗⲗⲁⲛⲓ ⲧⲁ. ⲗⲓⲛⲧⲟⲩⲛⲟⲩⲛⲓ ⲛⲁ. ⲁⲗⲟⲩⲛⲟⲥ ⲛ ⲁⲗⲉ ⲟⲩⲉⲗ ⲧⲓⲛⲧⲛⲓ ⲟⲩ ⲛⲁⲟⲩ ⲟⲩ ⲗⲁⲃ ⲡⲓ ⲗⲁⲣⲉⲗ ⲉⲛⲉ ⲧ ⲛ ⲗⲁⲃ ⲛⲛⲓ ⲧⲛⲓ ⲛⲉⲧⲟⲩⲉⲣⲉ ⲓⲁⲣ ⲛⲁⲟⲩ ⲛ ⲧ ⲛⲉⲧ ⲛⲧⲁⲣⲁⲗⲟⲩ ⲛⲁ. ⲧ ⲧⲁⲕⲟ ⲗⲓⲛⲉⲕⲱ ⲛⲧⲟⲩⲟⲩ ⲟⲩⲛ ⲛ ⲉⲧ ⲛⲉⲣⲉ ⲉ ⲣⲟⲥ ⲉⲣⲟ ⲥ ⲛⲁ. ⲩ</p> <p>ⲉⲩⲧⲉ ⲛⲉⲧⲟⲩⲉⲣⲉ ⲓⲁⲣ ⲛⲁⲟⲩ ⲟⲩ ⲛⲧⲟⲩⲉ ⲗⲁⲃ ⲡⲁⲣⲟⲩ ⲛ ⲉⲣⲉⲕⲗⲁⲥⲟⲩⲟⲩ ⲟⲩⲁⲣⲉ ⲗⲁⲃ ⲁⲛ ⲛⲁⲟ ⲟⲩ ⲁⲛⲁⲗⲉ ⲉⲧ ⲟⲩ ⲧⲣⲓⲛⲟ ⲛⲁⲟ ⲟⲩ ⲗⲓⲧⲛⲓ ⲡⲓ ⲟⲩⲟⲩⲛⲓ</p>	<p>2000003</p> <p>2000004</p> <p>2000005</p> <p>2000006</p> <p>2000007</p> <p>2000008</p> <p>2000009</p> <p>2000010</p>
--	---

# TRAViz by Stefan Jänicke

enter your password in Internet Accounts.

Text 1   Text 2  
 Shenoute XF   Shenoute XM   Set Texts!

duplicates    thinning    circle sizes



## Text Re-use Alignment Visualization

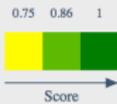
X

Shenoute XM

ΕΡΩΔΝ Π ΝΟΥΤΕ ΖΩΩ Θ ΨΟΧΝΕ Μ ΠΑΙ Ε ΕΙΡΕ Ν ΖΕΝ ΖΒΗΥΕ Ε Υ ΝΑΨΤ ΖΡΑΙ ΝΖΗΤΕ

Shenoute XF

Η ΜΜΟ Ν ΟΥΟΥ ΖΩΩ Θ ΠΕ ΕΤΒΗΝΤΕ ΟΝ ΕΝΤ Α Υ Ψ ΣΒΩ ΝΑ Θ Ε ΣΑΖΩ Θ ΕΒΟΛ ΜΜΟ ΕΤΒΗΝΤΕ ΕΝΤ Α Υ ΜΙΦΕ ΔΥΩ Ε Α Υ ΔΠΕΙΛΕΙ Ε ΠΑΙ Ε ΕΙΡΕ ΖΡΑΙ ΝΖΗΤΕ Ν ΖΕΝ ΖΒΗΥΕ Ε Υ ΝΑΨΤ ΕΤΒΗΝΤΕ ΟΝ Ν ΤΑ Υ ΜΙΦΕ Η ΔΥΑΠΙΛΕΙ ΝΑ Θ Ε ΖΩ Ε Θ ΕΙΝΕ Μ Π ΩΝΖ Ν ΧΝΑΖ ΕΖΟΥΝ ΕΝΕΤΟΥΕΨΠΜΟΥ Η ΕΤΣΩΤΠΠΜΟΥ ΝΖΟΥΟ Ε Π ΩΝΖ ΕΤΒΗΝΤΕ Ν ΤΑ ΟΥ ΣΜΟΥ ΕΙ ΕΧΜ ΠΑΙ ΕΤΒΗΝΤΕ ΟΝ Ν ΤΑ ΟΥ ΣΑΖΟΥ ΕΙ ΕΖΡΑΙ ΕΧΩ Θ ΕΤΒΕ Π ΔΝΑΨ Η ΝΑ Ν ΔΨ ΜΝ ΝΤ ΨΩ ΜΠΨΟΧΕ Μ Π ΝΟΥΤΕ ΕΝΤ Α Θ ΚΑΔ Θ ΝΩ Θ ΕΤΒΕ ΟΥ ΟΥΩΨ Ν ΡΩΜΕ









## Results 3

### Ver. 3 (coarse)

Text 1 (tokens)	Text 2 (tokens)	Text Reuse Candidates
MONB.XM	MONB.XV	133
MONB.XM	MONB.XF	705
MONB.XM	A26	79

### Ver. 6 (fine)

Text 1 (tokens)	Text 2 (tokens)	Text Reuse Candidates
MONB.XM	MONB.XV	13
MONB.XM	MONB.XF	218
MONB.XM	A26	3

# Results 4

## Ver. 3 (coarse)

Text 1 (tokens)	Text 2 (tokens)	Text Reuse Candidates
MONB.XV	MONB.XF	97
MONB.XV	A26	14

## Ver. 6 (fine)

Text 1 (tokens)	Text 2 (tokens)	Text Reuse Candidates
MONB.XV	MONB.XF	14
MONB.XV	A26	0

## Results 5

### Ver. 3 (coarse)

---

Text 1 (tokens)	Text 2 (tokens)	Text Reuse Candidates
MONB.XF	A26	63

---

### Ver. 6 (fine)

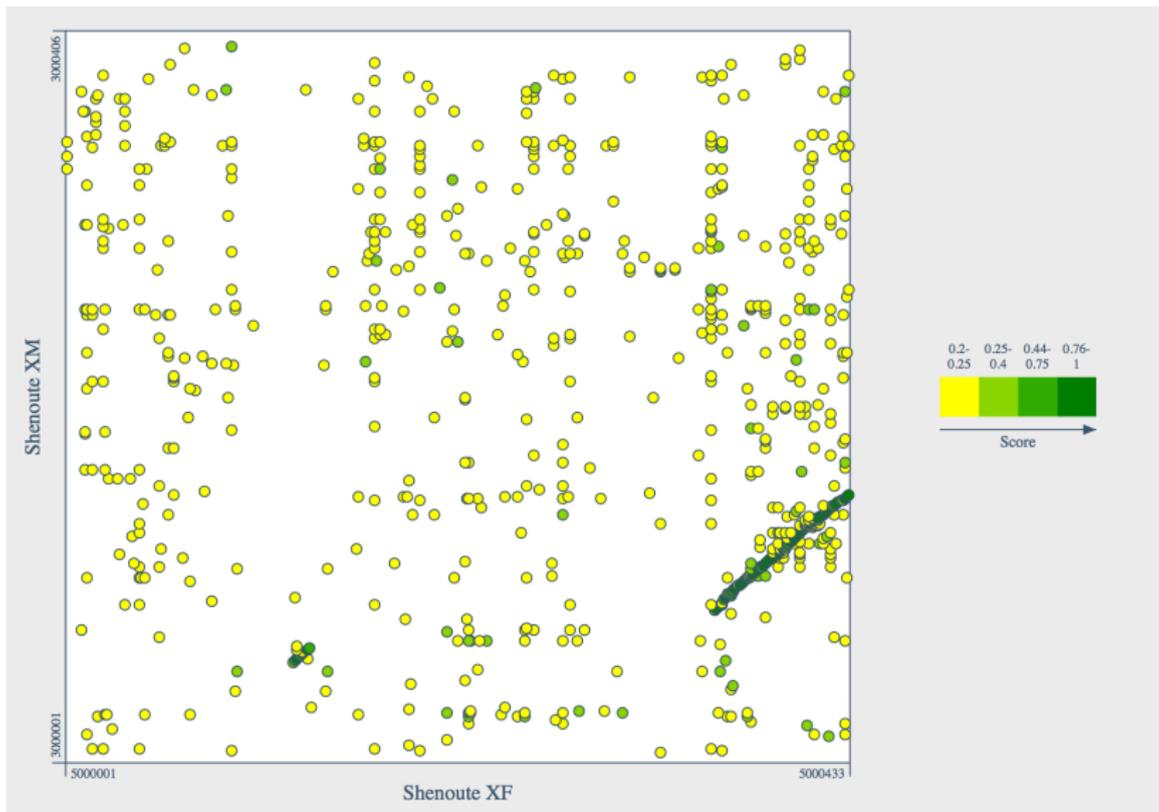
---

Text 1 (tokens)	Text 2 (tokens)	Text Reuse Candidates
MONB.XF	A26	0

---



# Parallel text is easily found







# Parallel text

TABLE 94 (CANON 6)—Continued

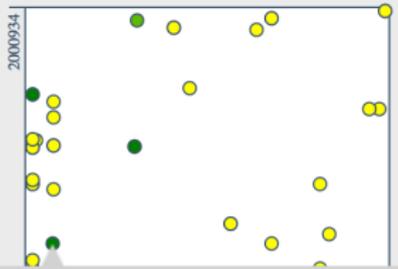
XF	XM	XV	YJ	XL	YK
272:ii.13	285:i.1				
273:i.1	285:i.16				
273:ii.18	286:i.1				
274:i.1	286:i.12				
274:ii.16	287:i.1				
275:i.1	287:i.15				
275:ii.16	288:i.1				
276:i.1	288:i.15				
276:ii.16	289:i.1				
277:i.1	289:i.13				
277:ii.19	290:i.1				
278:i.1	290:i.11				
278:ii.18	291:i.1				
[279:i.1]	291:i.13				
	292:i.1				
	293:i.1				
	294:i.1				
	295:i.1				
	296:i.1				
	297:i.1				
	298:i.1				
	299:i.1				
	300:i.1				
	301:i.1				
	302:i.1				

740

Emmel, Stephen. 2004. *Shenoute's Literary Corpus*. Vol. 2. Peeters. p. 785.

# Coincidence

duplicates  thinning  circle sizes



## Text Re-use Alignment Visualization

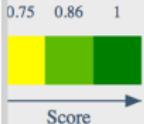


Besa

ΕΨΧΕ Ϛ ΒΟΛΗ Ε Π ΖΗΤ Ν Ν ΕΤ Ο Ν ΣΑΒΕ ΝΑ Υ ΜΑΥΑΑ Υ Ε ΤΜ ΤΡΕ Υ ΖΥΠΟΤΑΣΣΕ Ν Ν ΕΤ ΤΗΩ ΚΑΤΑ Ν ΚΑΝΩΝ Ν ΝΕ  
 ΠΡΟΦΗΤΗΣ ΜΝ Ν ΑΠΟΣΤΟΛΟΣ Ϛ ΟΥΩΝΖ ΖΩΩ Ϛ ΕΒΟΛ Ν ΟΥΟΝ ΝΙΜ ΕΝΤ Α Υ ΠΡΟΚΟΠΤΕ ΖΗΝ Θ ΥΠΟΤΑΓΗ ΧΕ ΜΝ  
 ΛΑΔΥ Ν ΡΩΜΕ Ν ΑΤ ΣΩΤΜ ΝΑΧΙΠΡΟΚΟΠΗ ΕΝΕΖ

Shenoute XF

Η Ν ΣΕ ΧΟΟ Ϛ ΧΕΟΥΟΙ Ν Ν ΕΤ ΧΩ ΜΜΟ Ϛ ΧΕ Ν ΔΝ ΟΥ Π ΠΕΘΟΟΥ ΔΥΩ Π ΠΕΤ ΝΑ ΝΟΥ Ϛ ΖΟΟΥ Ν ΕΤ ΩΠ Μ Π ΚΑΚΕ Ν  
 ΟΥΟΕΙΝ ΔΥΩ ΠΟΥΕΟΙΝ Ν ΚΑΚΕ Ν ΕΤ ΧΩ ΜΜΟ Ϛ ΧΕ Π ΕΤ ΣΑΥΕ ΖΟΛΔ ΔΥΩ Π ΕΤ ΖΟΛΔ Ϛ ΔΨΕ ΟΥΟΙ Ν Ν ΕΤ Ο Ν ΣΑΒΕ  
 ΝΑ Υ ΜΑΥΑΑ Υ ΔΥΩ Ε Υ Ο Ν ΡΜ Ν ΖΗΤ Μ ΠΕΥ ΜΤΟ ΕΒΟΛ



# hidden quotation without quotation markers

## Text Re-use Alignment Visualization

X

Besa

μπρ κτε π ραπ ε γ χολη αγω π καρπος ν τ δικαιοσυνη ν ου ciωε

Shenoute XF

ε γ πωωνε μ π ραπ ε γ χολη αγω π καρπος ν τ δικαιοσυνη ε γ ciωε ε γ ωπ μ π κακε ν ουγειν αγω π ουγειν ν κακε ε γ χω δε ον μμο σ ε π ετ σαωε δε η ρολδ αγω π ετ ρολδ δε η σαωε



© Stefan Jänicke, Leipzig University  
DEV in BMBF-project eTRACES (PN: 01UA1101A)

## Amos 6:12

“But you have turned justice into poison and the fruit of righteousness into wormwood” (ESV)



## One of many unknown text reuses

There are many text reuses between Shenoute and the Bible Amélineau did not find like this example.

### MONB.XF p.1

ΟΥΩΩ Μ- ΠΕ- ΒΟΙ Μ- Π- ΡΕϞ-Ρ-ΝΟΒΕ  
 break ACC- DEF.SG.M- arm.M GEN- DEF.SG.M- person.PCNS-do.PCNS-sin  
 ΜΝ- Π- ΠΟΝΗΡΟΣ Ν- ΔΔΙΜΟΝΙΟΝ  
 with- DEF.SG.M- evil ATTR- daimonion

“Break the arm of the sinner and the demonic evil man”



# The source

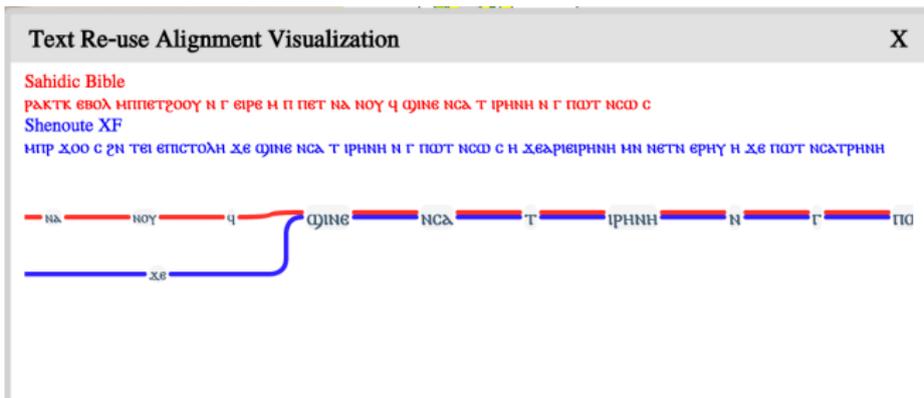
## Psalms 9:36 (crosswire.org number)

ΟΥΩΩ -ϣ Μ- ΠΕ- ΣΒΟΙ Μ- Π-  
 break -3SG.M ACC- DEF.SG.M- arm.M GEN- DEF.SG.M-  
 ΡΕϣ-Ρ-ΝΟΒΕ ΜΝ- Π- ΠΟΝΗΡΟΣ ΣΕ-  
 person.PCNS-do.PCNS-sin with- DEF.SG.M- evil PRS:3PL-  
 ΝΑ- ΩΙΝΕ ΝΣΑ- ΠΕϣ- ΝΟΒΕ [...]

FUT- seek after- POSS.SG.M:3SG.M- sin [...]

“Crush the arm of the sinner and evildoer; his sin shall be sought out”

# One of many unknown text reuses



Sahidic Bible

Shenoute XF

ΡΑΚΤΚ ΕΒΟΛ ΜΠΠΕΤΖΟΥ Ν Γ ΕΙΡΕ Μ Π ΠΕΤ ΝΑ ΝΟΥ Ψ ΩΙΝΕ  
 ΝΣΑ Τ ΙΡΗΝΗ Ν Γ ΠΩΤ ΝΣΩ Σ

1002953

5000134

ΜΠΡ ΧΟΟ Σ ΖΗ ΤΕΙ ΕΠΙΣΤΟΛΗ ΧΕ ΩΙΝΕ ΝΣΑ Τ ΙΡΗΝΗ Ν Γ  
 ΠΩΤ ΝΣΩ Σ Η ΧΕΑΡΙΕΙΡΗΝΗ ΜΝ ΝΕΤΗ ΕΡΗΥ Η ΧΕ ΠΩΤ  
 ΝΣΑΤΡΗΝΗ

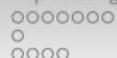
ΜΠΡΧΟΟΣ ΖΗΤΕΙΕΠΙΣΤΟΛΗ ΧΕΩΙΝΕ ΝΣΑΤΡΗΝΗ ΝΓΠΩΤ ΝΣΩΣ ``Do not say in this letter, `seek for peace and pursue it [...]' " (MONB.XF p.184)

## Interesting case

### Shenoute XF p.264

Π-                    ΝΟΥΤΕ            †-                    ρΤΗ-                    -κ                    ε-  
 DEF.SG.M-    god.SG.M    give.SCNS-    heart.SPRO-    -2SG.M    DIR-  
 τὰ-                    ΨΥΧΗ            Ν-                    -Γ-                    COT-  
 POSS.SG.F:1SG-    soul.SG.F    CONJ-    -2SG.M-    redeem.SPRO-  
 -c                    ε-ΒΟΛ                    ρΜ-                    Π-                    CΑΡΟΥ                    Μ-  
 -3SG.SG.F    DIR-outside    LOC-    DEF.SG.M-    curse.SG.M    GEN-  
 ΠΙ-                    ΔΝΑΩ  
 DEM.DIST.SG.M-    oath.M

“God, pay attention to my soul and redeem it from the curse of the oath”



## Gnostic-monastic relationship?

I also played with Gnostic texts.

Pistis Sophia, Book 1, p. 54, l. 11-12

†-	ϷTH-	-κ	ϵ-	Τα-	ΥΥΧΗ
give.SCNS-	heart.SPRO-	-2SG.M	DIR-	POSS.SG.F:1SG-	soul.SG.F
N-	-Γ-	COΤ-	-C	ΝΑΖΜΕ-	-Τ ΕΤΒΕ-
CONJ-	-2SG.M-	redeem.SPRO-	-3SG.F	save.SPRO-	1SG about-
Να-	ΧΑΧΕ				
POSS.PL:1SG-	enemy				

“Pay attention to my soul and redeem it and save me for the sake of my enemies”



# Spelling difference

## Text Re-use Alignment Visualization

X

### Besa

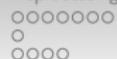
ΠΑΛΙΝ ΟΝ ΜΗΝΝΑΝΕΙΘΒΗΥΕ ΤΗΡ ΟΥ ΝΣΕΘΝΖΕΝΡΩΜΕ ΖΡΑΙ ΝΖΗΤ Ν ΕΥ Ο Ν Ν ΟΕΙΚ Ε ΤΕΥ ΨΥΧΗ ΜΑΥΔΑ Υ ΖΝ ΖΕΝ  
 ΜΕΕΥΕ Ε Υ ΔΑΖΜ Ε Υ ΠΟΡΝΕΥΕ ΖΝ ΠΕΥ ΖΗΤ Ν Θ Ε ΟΝ Ε ΤΕΥ ΣΥΝΗΘΕΙΑ ΤΕ ΤΑΙ ΧΙΝ ΤΕΥ ΑΡΧΗ ΝΘΙ Ν ΕΤ Ο ΝΑ Θ ΗΤ  
 ΖΡΑΙ ΝΖΗΤΕ ΜΠ ΟΥ ΣΕΙ ΖΝ ΤΕΥ ΨΥΧΗ ΨΑΖ ΟΥΝ Ε ΤΕΝΟΥ ΜΠΕΧΠΙΟ Μ Π ΨΑΧΕ Μ ΠΕ ΠΡΟΦΗΤΗΣ ΧΕ ΖΝ ΝΑΙ ΤΗΡ ΟΥ  
 ΜΠ Σ ΚΤΟ Σ ΕΡΟ Ι ΖΜ ΠΕΣ ΖΗΤ ΤΗΡ Ψ ΝΘΙ Τ ΑΘΗΤ ΝΤΕ ΙΟΥΔΑ

### Shenoute XM

ΣΕ ΝΑ ΝΑΥ ΕΡΟ ΝΘΙ Ν ΔΙΚΑΙΟΣ ΕΤ Ρ ΖΟΤΕ ΖΗΤ Ψ ΜΠΕΧΣ Ν ΣΕ ΧΟΟ Σ ΧΕ ΕΙΣ ΟΥ ΣΖΙΜΕ Ε ΜΠ Σ ΚΑ ΠΝΟΥΤΕ ΝΑ Σ Ν  
 ΒΟΗΘΟΣ ΑΛΛΑ Δ Σ ΖΕΛΠΙΖΕ ΕΧΜ Π ΔΨΑΙ Ν ΤΕΣ ΜΝΤ ΧΑΔΙ ΖΗΤ ΜΝΤ ΕΣΜΝΤΡΕΨΚΑΖΤΗΣ ΕΡΟ Σ ΜΑΥΔΑ Σ Ν Θ Ε ΕΝΤ Δ  
 Υ ΧΟΟ Σ ΧΕ ΖΝ ΝΑΙ ΤΗΡ ΟΥ ΜΠ Σ ΚΟΤ Σ ΕΡΟ Ι ΖΜ ΠΕΣ ΖΗΤ ΤΗΡ Ψ ΝΘΙ Τ ΑΘΗΤ ΝΤΕ ΙΟΥΔΑ ΑΛΛΑ ΕΖΡΑΙ ΕΧΝ ΟΥ ΜΝΤ  
 ΝΟΥΧ







## Different spelling by Shenoute

### Shenoutean Canon6 MONB.XM p.469

π-	Ϡαδε	μ-	πε-	προφητης	δε-	ρη-
DEF.SG.M-	saying.M	GEN-	DEF.SG.M-	prophet.M	COMP-	LOC-
ναι	τηρ-	-ου	μη-	-ς-	κτο-	-ς
DEM.PL.SABS	entire-	-3PL	PST:NEG-	-3SG.F-	turn.SPRO-	-3SG.F
ερο-	-ι	ρη-	πες-	ρητ	τηρ-	-ς
DIR.SPRO-	-1SG	LOC-	POSS.SG.M:3SG.F-	heart.M	entire-	-3SG.M
νη-	τ-	αθητ	νηε-	ιουδα		
NOM-	DEF.SG.F-	not_having:mind	GEN-	Judah		

“the saying of the prophet, ‘in all these, the foolish Judah did not return to me with her whole heart’ ” (from Jer 3:10)

# In the future

cohyponym list & synonym list

→ enables to detect semantic text reuse such as allusion and paraphrase

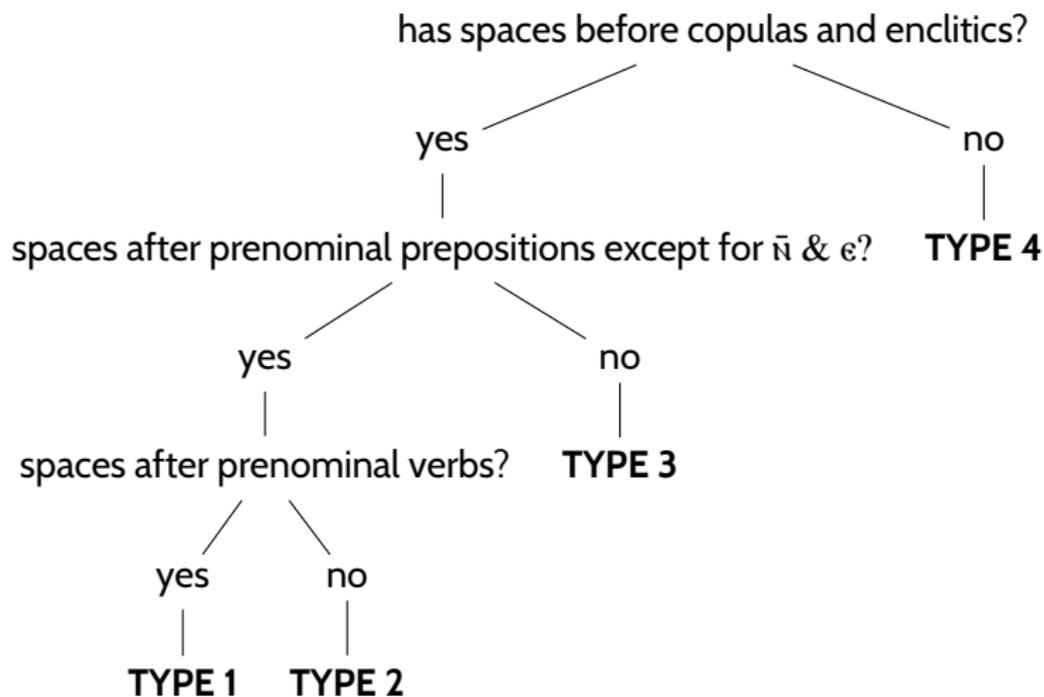
More texts

Big data from White Monastery Library.

correct tokenization or improve tokenizer

now roughly 90% is accurate

# Tokenization is a key



# Typology of spacing

TYPE 1 Quecke (1984) etc.

TYPE 2 Lambdin (1983) etc.

TYPE 3 Layton (2011), Coptic SCRIPTORIUM, Young (1993),  
Amélineau etc.

TYPE 4 Wilmet (1957, 1958, 1959), Kuhn etc.

Walter Till's transcription is a kind of hybrid.

# Thank you!

ⲁⲉⲡⲉⲙⲟⲩ!

## References

Behlmer, H. (2008). Do not believe every word like the fool...! Rhetorical strategies in Shenoute, *Canon 6*. In G. Gabra and H. N. Takla (Eds.), *Christianity and monasticism in upper Egypt, Vol. 1 : Akhmim and Sohag*, pp. 1–12. Cairo: American University in Cairo Press.

Behlmer, H. (2016). Die Bibel im koptischen Mönchtum der Spätantike. In P. Gemeinhardt (Ed.), *Zwischen Exegese und religiöser Praxis: Heilige Texte von der Spätantike bis zum Klassischen Islam*, *Orientalia Lovaniensia Analecta*, pp.143– 175. Tübingen: Mohr Siebeck.

Behlmer, H. (in press). The Use of the Psalms in Shenoute's Tractate *He Who Sits Upon His Throne*. In David Brakke, Stephen J. Davis and Stephen Emmel (Eds.), *Scriptae Coptice. Essays in Honor of Bentley Layton*, *Orientalia Lovaniensia Analecta*, pp. 339–353. Leuven: Peeters.

Lambdin, T. O. (1983). *Introduction to Sahidic Coptic*. Mason: GA: Mercer University Press.

Layton, B. (2011). *A Coptic grammar: With chrestomathy and glossary: Sahidic dialect* (Third ed.). Wiesbaden: Harrassowitz Verlag.

Quecke, H. (1984). *Das Johannesevangelium Säidisch: Text der Handschrift PPalau Rib. Inv.-Nr. 183 mit den Varianten der Handschriften 813 und 814 Chester Beatty Library und der Handschrift M 569*. Roma - Barcelona: Papyrologica Castroctaviana.

Wilmet, M. (1957). *Concordance du Nouveau Testament sahidique, II. Les mots autochtones*, Volume 1. Louvain: Secrétariat du CorpusSCO.

Wilmet, M. (1958). *Concordance du Nouveau Testament sahidique, II. Les mots autochtones*, Volume 2. Louvain: Secrétariat du CorpusSCO.

Wilmet, M. (1959). *Concordance du Nouveau Testament sahidique, II. Les mots autochtones*, Volume 3. Louvain: Secrétariat du CorpusSCO.

Young, D. W. (1993). *Coptic Manuscripts from the White Monastery: Works of Shenute: Textband*. Wien: Verlag Brüder Hollinek.