

A Corpus of Biblical Names in the Greek New Testament

to Study the Additions, Omissions, and Variations
across Different Manuscripts

Christoph Werner¹ Zacharias Shoukry² Soham Al-Suadi² Frank Krüger¹

¹Wismar University of Applied Sciences

²University of Rostock

Hersonissos, 27.05.2024



Introduction

The New Testament and its Manuscripts

Structure of the 27 Books:

- Gospels (Matthew, Mark, Luke, John)
- Acts of the Apostles
- Apostolic Letters (21 letters to churches, persons, general public)
- Book of Revelation

Types and counts of Manuscripts:

- Papyri (136)
- Majuscules (290)
- Minuscules (2,901)
- Lectionaries (2,498)

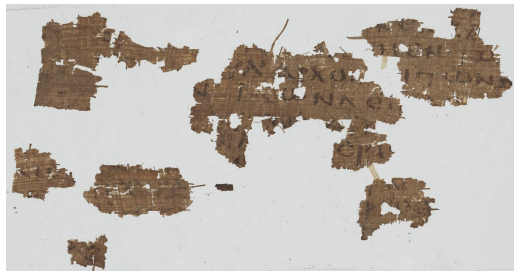


Fig. 1: Papyrus 21 (P21) [1]

Introduction

The New Testament and its Manuscripts

Structure of the 27 Books:

- Gospels (Matthew, Mark, Luke, John)
- Acts of the Apostles
- Apostolic Letters (21 letters to churches, persons, general public)
- Book of Revelation

Types and counts of Manuscripts:

- Papyri (136)
- Majuscules (290)
- Minuscules (2,901)
- Lectionaries (2,498)

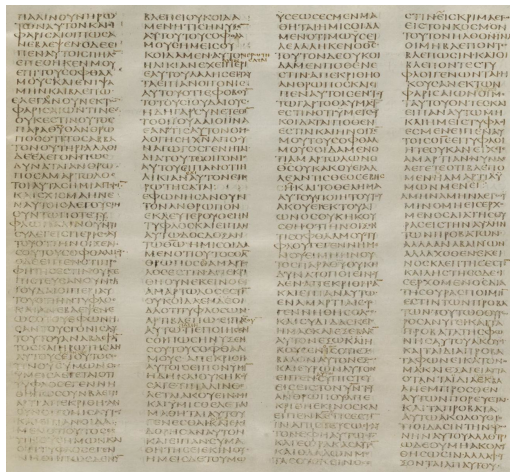


Fig. 2: Detail of Codex Sinaiticus (01) [2]

Introduction

The New Testament and its Manuscripts

Structure of the 27 Books:

- Gospels (Matthew, Mark, Luke, John)
- Acts of the Apostles
- Apostolic Letters (21 letters to churches, persons, general public)
- Book of Revelation

Types and counts of Manuscripts:

- Papyri (136)
- Majuscules (290)
- Minuscules (2,901)
- Lectionaries (2,498)

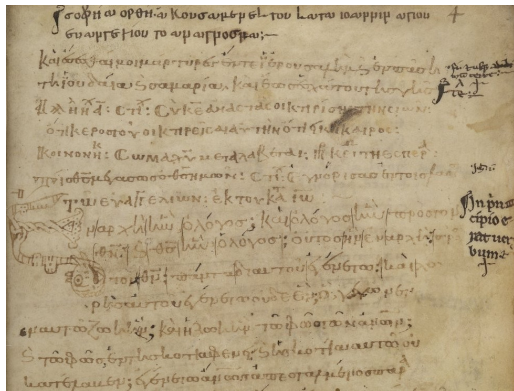


Fig. 3: Detail of Lectionary 60 (L60) [3]

Introduction

Differences between Manuscripts

- Multiple manuscript containing the same verse
- Copying leads to omissions, additions, variations
 - By mistake
 - By bias (allegedly Codex Bezae (05) with anti female tendencies)
- Manuscripts are revision texts

Forms of modification:

1. Leaving out/Adding a name to make it less or more prominent
2. Leaving out/Adding context (e.g. devaluing women by adding 'and their children')
3. Changes in grammar (e.g. 'the news of it/her/him')

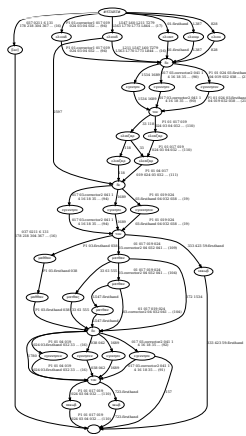


Fig. 4: Collation graph of Matt 1:15

Data Collection and Processing

Transcription Sources

- Institut für neutestamentliche Forschung (INTF) Münster
 - Focus on Gospels, Acts and Book of Revelation
 - Provides API for transcription and manuscript metadata
 - Transcriptions in TEI format (without manuscript metadata)
- International Greek New Testament Project (IGNTP) Birmingham
 - Focus on Epistles to the Romans, Galatians, Ephesians, Philippians, Corinthians
 - Transcriptions in TEI format (with manuscript metadata)

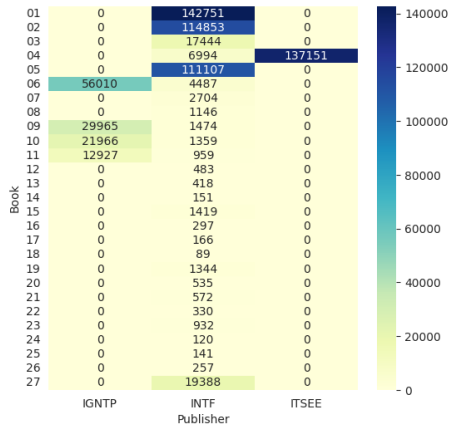


Fig. 5: Transcribed Verses per Book of the New Testament

Data Collection and Processing

TEI Transcription Files

- XML schema provided by the Text Encoding Initiative (TEI)
- TEI is the “encoding scheme” for:
 - Manuscript metadata (title, source description, number of witnesses, publishing information, etc.)

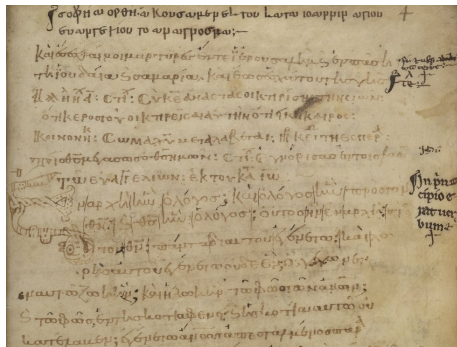


Fig. 6: Detail of Lectionary 60 (L60) [3]

Data Collection and Processing

TEI Transcription Files

- XML schema provided by the Text Encoding Initiative (TEI)
- TEI is the “encoding scheme” for:
 - Manuscript metadata (title, source description, number of witnesses, publishing information, etc.)
 - Text format (number of columns etc.)

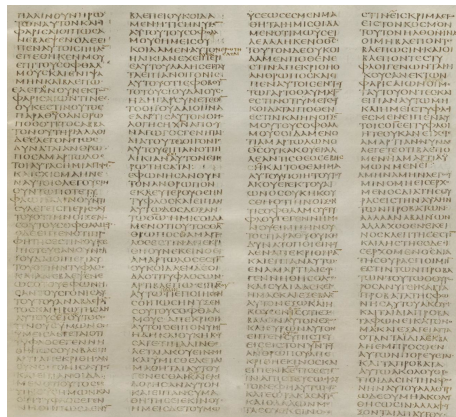


Fig. 7: Detail of Codex Sinaiticus (01) [2]

Data Collection and Processing

TEI Transcription Files

- XML schema provided by the Text Encoding Initiative (TEI)
- TEI is the “encoding scheme” for:
 - Manuscript metadata (title, source description, number of witnesses, publishing information, etc.)
 - Text format (number of columns etc.)
 - Structure (books, chapters, verses, notes)

```
1  ...
2  <div type="book" n="1Cor">
3    <div type="inscriptio">
4      <ab n="1Cor.inscriptio">
5        <w>προς</w>
6        ...
7      </ab>
8    </div>
9    <div type="chapter" n="1Cor.1">
10     <ab n="1Cor.1.1">
11       <w>παυλος</w>
12       ...
13     </ab>
14     <ab n="1Cor.1.2">
15       <w>τη</w>
16       ...
17     </ab>
18     ...
19   </div>
20   <div type="chapter" n="1Cor.2">
21     <ab n="1Cor.2.1">
22       <w>παυλος</w>
23       ...
24     </ab>
25     ...
26   </div>
27   ...
28 </div>
29 ...
```

Fig. 8: Structure of Books / Chapter / Verse

Data Collection and Processing

TEI Transcription Files

- XML schema provided by the Text Encoding Initiative (TEI)
- TEI is the “encoding scheme” for:
 - Manuscript metadata (title, source description, number of witnesses, publishing information, etc.)
 - Text format (number of columns etc.)
 - Structure (books, chapters, verses, notes)
 - Content (letters and their readability)

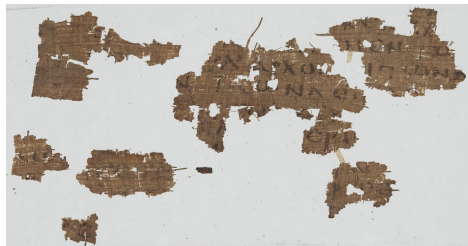


Fig. 9: Fragmentation of Papyrus 21 (P21) [1]

```
1 ...  
2 <ns0:w>αpxo<ns0:unclear>v</ns0:unclear><ns0:supplied  
↳ reason="unspecified">τ</ns0:supplied>ι</ns0:w>  
3 <ns0:w>τωv</ns0:w>  
4 ...
```

Fig. 10: Transcription of the first line in the middle segment of P21

Data Collection and Processing

TEI Transcription Files

- XML schema provided by the Text Encoding Initiative (TEI)
- TEI is the “encoding scheme” for:
 - Manuscript metadata (title, source description, number of witnesses, publishing information, etc.)
 - Text format (number of columns etc.)
 - Structure (books, chapters, verses, notes)
 - Content (letters and their readability)
 - Special groups of letters (abbreviations, nomina sacra)



Fig. 11: Nomina sacra IC XC from the Greek ΙΗΣΟΥΣ ΧΡΙΣΤΟΣ (Jesus Christ) [4]

Data Collection and Processing

Data Dictionary for Verses Collection

Parsing TEI files for:

- Manuscript Identifiers
- Verse Identifiers
- Verse Transcriptions
- Markings of readability
- Manuscript Publisher

Result:

- 667,292 transcriptions
- 7,985 verses
- 956 manuscripts

Column	Description	Example data
ga	Document Identifier	P1
bkv	Verse Identifier	B01K1V1
nkx	Alternative Verse Identifier	Matt.1.1
text	Transcription text	βιβλος γενεσεως ιυ χυ υυ δαυιδ υιου αβρααμ
marks	Readability of letters	cccccc cccccccc cc cc uu uuccc ssss ccccc
publisher	Transcription publisher	The Institut für neutestamentliche Textforschung
source	Download source	ntvmr

Table 1: Data Dictionary for Verses Collection

Data Collection and Processing

List of Names

- FactGrid (database for historians) for initial list of characters
- Manually adding names and known spelling variations from [5, 6]
- Merging characters with identical names

Names vs. Characters

We do use names solely, due to potential identity overlap of some characters.

Mary of Magdala	Female gender
Mary of Nazareth	Female gender
Mary of Rome's Christians	Female gender
Mary sister of Martha and Lazarus	Female gender
Mary, mother of John Mark	Female gender
Mary, wife of Cleopas	Female gender

Fig. 12: Characters named Mary in FactGrid database

label:en	label:el	variations
Mary	Μαρία	μαρια, μαριας, μαριαν

Table 2: Standard spelling variations for Mary

Search for Names

- Utilizing string search
- Search each name variation (857 variations) in each verse (667292 verses)
- Mark occurrence in verse

docID	bkv	occurrence	name	text
30028	B02K16V9	True	μαρια	[...] πρωτον μαρια τη μαγδαλινη [...]
30026	B02K16V9	False	μαρια	[...] πρωτον μαριαι τη μαγδαληνη [...]

Table 3: Potentially missing of Mary

- Iteratively finding and adding unknown variations to list of names

label:en	label:el	variations
Mary	Μαρία	μαρια,μαριαρ,μαριαι,μαρηαμ,μαριας,μαριασ,μαρηα,μαρι,μαριαμ,μαρηαν,μαριαν

Table 4: Found spelling variations for Mary

Analysis of the Occurrences Corpus

Analysis of Additions, Omissions, and Variations of selected Names

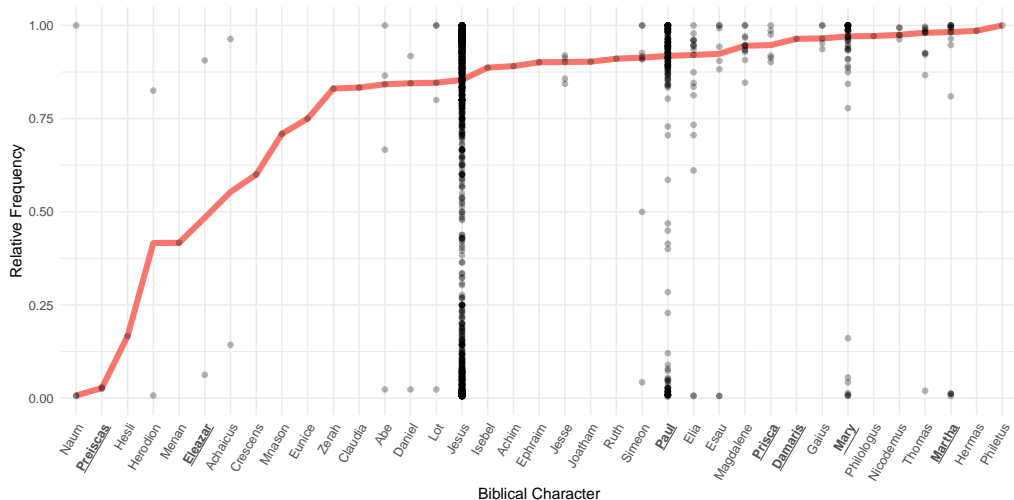


Fig. 13: Relative occurrence frequency of a selected subset of biblical names across all verses.

Analysis of the Occurrences Corpus

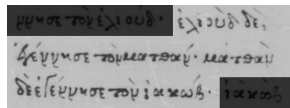
Example: Omission of Eleazar in Matt 1:15

Proper wording of the verse:

greek ελιουδ δε εγεννησε τον ελεαζαρ ελεαζαρ δε εγεννησε τον ματθαν ματθαν δε εγεννησε τον ιακωβ.

english Eliud begat Eleazar, Eleazar begat Mattan, Mattan begat Jacob.

Actual handwriting in Minuscule 2597 and its transcription:



(a) Handwriting of Matt 1:15 in Minuscule 2597

ελιουδ δε
εγεννησε τον ματθαν ματθαν
δε εγεννησε τον ιακωβ

(b) Transcription

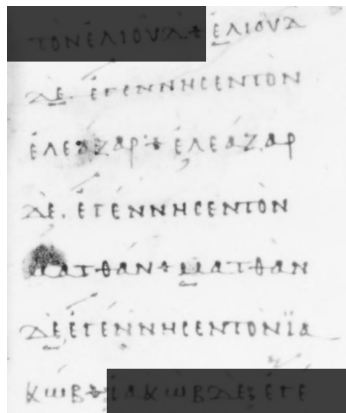
Eliud
begat Mattan Mattan
begat Jacob

(c) Translation

Fig. 14: Omission of Eleazar in the List of Ancestors of Jesus

Analysis of the Occurrences Corpus

Example: Omission of Eleazar in Matt 1:15



(a) Handwriting of Matt 1:15 in Majuscule 045 [7]

ελιουδ
δε εγεννησε τον
ελεαζαρ ελεαζαρ
δε εγεννησε τον
ματθαν ματθαν
δε εγεννησε τον ια
κωβ

(b) Transcription

Eliud
begat
Eleazar Eleazar
begat
Mattan Mattan
begat
Jacob

(c) Translation

Fig. 15: Eleazar in the List of Ancestors of Jesus

Limitations

- Transcribing is largely manual work
- Transcription Guidelines [8, 9] do exist
- Transcribers do not always adhere to those guidelines
- String search is not optimal
 - Computational inefficient
 - Prone to mistakes (e.g. **δια** is the accusative form of Zeus and a preposition)

docID	bkv	occurrence	name	text
30035	B05K14V12	True	ζευσ	[...] βαρναβαν δια τον δε παυλον [...]
30035	B05K14V22	True	ζευσ	[...] πιστει και οτι δια πολλων θλιψεων [...]

Table 5: False positive occurrence of Zeus in Acts 14:22 (B05K14V22)

Conclusion and Future Work

- Known Omissions and Additions found
- Discovery of a not yet discussed addition of Mary
- Improve search by utilizing machine learning methods
 - Part of Speech (POS) tagging
 - Named Entity Recognition (NER)
- Develop an ontology extending:
 - Gender, Sex, and Sexual Orientation Ontology [10]
 - Theographic [11]
- Build a knowledge graph representing the occurrences of names and their variations in the Ancient Greek New Testament

Links



Code Repository on [Github](#)



Data at [Zenodo](#)



Paper at [SemDH 2024](#)

Contact: Christoph Werner (christoph.werner@hs-wismar.de)

References I

- [1] Special Collections and Archives, Trexler Library. Muhlenberg College, P. Oxy. 1227: St. Matthew's gospel, xii., online, accessed 2024-03-11 (2015).
URL https://library.artstor.org/#/asset/SS7730556_7730556_9313349
- [2] Codex Sinaiticus Photos, M_Q80_F7_r, online, accessed 2024-05-13 (2019).
URL <https://www.flickr.com/photos/169807637@N06/40078755793/in/album-72157689563046773/>
- [3] Bibliothèque nationale de France. Département des Manuscrits, Gr. 375, online, accessed 2024-05-13 (2012).
URL <https://gallica.bnf.fr/ark:/12148/btv1b8496551f/f19.item>
- [4] Skylax30, Detail from an icon at the troyan monastery in bulgaria, accessed 2024-05-15 (2018).
URL https://commons.wikimedia.org/wiki/File:Nomina_sacra_IC_XC_2.jpg
- [5] W. Bauer, Griechisch-deutsches Wörterbuch zu den Schriften des Neuen Testaments und der frühchristlichen Literatur, 6th Edition, Walter de Gruyter, Berlin, 2012, frühere Auflage unter dem Titel: Bauer, Walter: Griechisch-deutsches Wörterbuch zu den Schriften des Neuen Testaments und der übrigen urchristlichen Literatur.

References II

- [6] J. P. Louw, E. A. Nida, Greek-English Lexicon of the New Testament, Vol. 1, United Bible Societies, New York, 1988.
- [7] Library of Congress, Collection of Manuscripts – Monastery of Dionysios 55. (old 10). (Greg. 045,). Four Gospels. 8th/9th cent. 259 f., accessed 2024-03-04 (2024).
URL <https://www.loc.gov/resource/amedmonastery.00271050008-ma>
- [8] H. Houghton, C. Smith, IGNTP guidelines for XML transcriptions of New Testament manuscripts (version 1.6) (November 2023).
URL <http://epapers.bham.ac.uk/4301/>
- [9] A. C. Myshrall, R. Kevern, H. Houghton, IGNTP guidelines for the transcription of manuscripts using the Online Transcription Editor (April 2020).
URL <http://epapers.bham.ac.uk/3436/>
- [10] C. A. Kronk, J. W. Dexheimer, Development of the gender, sex, and sexual orientation ontology: Evaluation and workflow, *Journal of the American Medical Informatics Association* 27 (7) (2020) 1110–1115.
doi:10.1093/jamia/ocaa061.

References III

- [11] R. Rouse, Theographic a knowledge graph of the bible, online, accessed 2024-05-13 (2019).
URL <https://github.com/robertrouse/theographic-bible-metadata>