



Normalized Datasets of Hahn's and Zahn's Reconstructions of Marcion's *Gospel*

DATA PAPER

MARK G. BILBY 

]u[ubiquity press

ABSTRACT

These four datasets are the first born-digital, normalized, lexicographically enriched, and peer-reviewed versions of the reconstructions of Marcion's *Gospel* made by August Hahn in 1832 and Theodor Zahn in 1892. Two dataset files were generated for each reconstruction: the first consisting of human-readable Postclassical Greek; the second of lemmatized and morphologically tagged text following the openly licensed BibleWorks Greek Morphology schema. These datasets represent another batch in a series of normalized and enriched datasets of major reconstructions of Marcion's *Gospel* published by *JOHD*.

CORRESPONDING AUTHOR:

Mark G. Bilby

Pollak Library, California State
University, Fullerton, US

mgb8n@virginia.edu

KEYWORDS:

Gospel of Marcion; Marcion
of Sinope; Gospel of Luke;
Synoptic Problem

TO CITE THIS ARTICLE:

Bilby, M. G. (2021). Normalized
Datasets of Hahn's and
Zahn's Reconstructions of
Marcion's *Gospel*. *Journal of
Open Humanities Data*, 7:
31, pp. 1–5. DOI: [https://doi.
org/10.5334/johd.63](https://doi.org/10.5334/johd.63)

(1) OVERVIEW

Repository location: DOI: <https://doi.org/10.7910/DVN/BYNHX6>

Morphology key: Bilby, M.G. (2021a). *Key to BibleWorks Greek Morphology (BGM)* (v1.1). DOI: <https://doi.org/10.5281/zenodo.4950243>

Print sources:

Hahn, A. (1832). *Evangelium Marcionis ex Auctoritate Veterum Monumentorum*. In Ioannis Caroli Thilo, ed., *Codex apocryphus Novi Testamenti* (pp. 401–486). Lipsius: F. C. G. Vogel. ARK: <https://n2t.net/ark:/13960/t23b6zq5d>

Zahn, T. (1888/1892). *Geschichte des neutestamentlichen Kanons* (Vols. 1–2). Erlangen: Andreas Deichert. ARK: <https://n2t.net/ark:/13960/t8cf9s958>

CONTEXT

The history of scholarly reconstructions of the Gospel of Marcion (GMarc) is briefly detailed in our Harnack (Bilby, 2021b) and Roth (Bilby, 2021c) data papers. Eight major published reconstructions have appeared over the last two centuries: Hahn (1832), Zahn (1892), von Harnack (1921/924), Tsutsui (1992), BeDuhn (2013), Roth (2015), Klinghardt (2015/2020, 2021), and Nicolotti (2019). While Ritschl (1846) and Hilgenfeld (1850) undertook thorough prose analyses of GMarc and included many quotations, neither provided a clear, running main text. Gramaglia (2017) has also notably offered a complete Italian translation and critical commentary based on Klinghardt’s (2015) reconstruction. My GMarc reconstruction, embedded within my newly proposed solution to the Synoptic Problem, is in progress (Bilby, 2020-07/2021-12).

Clear relationships exist among these reconstructions because of their shared dependence on common underlying data: over 700 patristic attestations to GMarc by over fifteen ancient witnesses, hundreds of textual variants and thousands of non-variants in manuscripts of Luke, and tens of thousands of close parallel words in other gospels. However, because of quite different methodological assumptions held and approaches taken by their respective editors, these reconstructions reflect widely varying results and reveal editorial idiosyncrasies as well as clear dependencies when cross-compared quantitatively.

Hahn’s pioneering reconstruction (1832) represented the fruit of his extensive earlier study on Marcion and the texts distributed under Marcion’s patronage (Hahn, 1823). As Roth notes, “Hahn’s work was particularly important in that he provided the first attempt to present comprehensively Marcion’s Gospel as reconstructed from the available sources” (2015: 9). While Hahn’s maximalist, continuous approach to reconstruction was challenged by many later editors, he still provided a starting point for all future reconstructions. Zahn sought to correct and pare down Hahn’s overly generous reconstruction, while making use of the *Editio octava critica minor* (EOCM) (Tischendorf, 1869) as a base text. Zahn’s discontinuous approach became the precursor to more radically discontinuous and minimalist reconstructions in subsequent scholarship, namely those of Tsutsui (1992) and Roth (2015).

Given the central place of GMarc in heated debates about the compositional and editorial formation of the earliest gospels (both canonical and non-canonical), quantitative analyses, computational linguistics (CL), and historical corpus linguistics (HCL) have the potential to clarify and transcend the deep subjective and idiosyncratic divides within scholarship and may prove decisive in settling centuries-old questions about the earliest texts that arose out of the Jesus movement.

(2) METHOD

STEPS

Our normalization of GMarc datasets began with the work of Adolf von Harnack (Harnack, 1921/1924) because of the public domain status of the corresponding print edition and its established place as the standard reconstruction of GMarc for most of the last century. Of all major GMarc reconstructions in history, Harnack’s is the most ambiguous and inconsistent in its indications. The challenge of normalizing Harnack’s reconstruction proved crucial to develop and implement a few clear, consistent datatypes and indications that allow for meaningful comparisons by humans and machines of all major GMarc Greek datasets.

- 1) () parentheses for words and phrases to indicate they are necessary to the meaning of the surrounding words or that there was some doubt/confusion about them
- 2) [] square brackets for verses Hahn deemed were removed by Marcion
- 3) *) asterisk followed by right parenthesis to indicate a variant detailed in the footnotes
- 4) #) number followed by right parenthesis to indicate footnotes, which provide references to patristic citations of GMarc and/or Hahn's explanation of the possible inclusion or exclusion of this content from GMarc

To transform Hahn's reconstruction into normalized datasets, indication 1 is kept as is, verses corresponding to indication 2 are omitted, and indication 3 is rendered as empty square brackets. Supplemental version identifiers (here 01H for the first edition of GMarc by Hahn) are added to the beginning of each line to facilitate version identification and content alignment in computational linguistics environments as well as the creation of consistently sorted, interlinear arrangements of past editions of GMarc.

Zahn exercises more caution and nuance in his decisions and indications than did Hahn, resulting in a relatively lesser—yet still unacceptably high—level of Lk2 contamination. Besides restoring clear wording for many verses, Zahn also uses these indications, several of which were later commonplace in Harnack's work:

1. “bare chapter and verse numbers”/“nackten Kapitel- und Verszahlen” without any wording to indicate implicitly present content lacking any documented differences from the text of Luke in Tischendorf's *Editio octava critica maior* (EOCM), which Zahn used as his base text
2. ... ellipses to indicate smaller ranges of implicitly present content aligned with EOCM
3. etc. to indicate trailing ranges of implicitly present content aligned with EOCM
4. (enclosing parentheses) to indicate a lack of attestation for whole verses and specific words
5. parentheses surrounding a self-standing question mark (?) to indicate doubt about the presence of verses or words, whether preceding or intervening, and/or to indicate alternative wording
6. (question mark after parenthesized words?) to indicate uncertain and/or alternative wording
7. “or” (*oder*) or a prefixed em-dash (—) inside parentheses to indicate alternative wording
8. [square brackets] to indicate absent and/or highly doubtful verses and wording, sometimes introduced by “without” [*ohne*] or concluded with a [question mark?]

To transform Zahn's reconstruction into normalized datasets, all clear wording is included, but all implicitly present content corresponding to numbers 1–3 is placed in [brackets]. For content corresponding to indication 4, whole verses are omitted but specific words within verses are retained, since these typically involve contextually required words. For indication 5, preceding words are placed in parentheses, empty parentheses render intervening words and whole verses, and empty square brackets render alternative wording. Uncertain content corresponding to indication 6 is replaced with empty parentheses, but alternative wording is rendered as empty square brackets. For indication 7, the alternative reading is signaled with empty square brackets. Content corresponding to indication 8 is omitted.

QUALITY CONTROL

As detailed in our Harnack data paper, the first dataset for each reconstruction consists of normalized, human-readable Greek, while the second manually applies lemmatization and morphological tagging using the BibleWorks Greek Morphology (BGM) schema, which is lightweight, adaptable, familiar to many scholars, openly licensed for non-commercial use,

and easy to compile, edit, and query in word processor and CL environments. For quality control in the transcription of the respective Greek texts, we created interlinear parallels by verse for all GMarc editions together with canonical Luke, and made second and third passes to check each transcription against the corresponding print edition. Similarly for the lemmatizing and morphological tagging process, we sorted the editions by verse in an interlinear format and made regular use of close or exact parallel tagging already done for the canonical Gospels in BGM and the tagging we had previously done for our Harnack and Roth datasets. As a mediating step for this batch of two longer editions, we wrote and ran an R script that automated the lemmatizing and morphological tagging for about 25% of words, those that either individually or as syntagmata proved lexicographically and syntactically unambiguous. After that, we spent about 100 hours manually tagging the remainder of untagged words, looking them up in the *Thesaurus Linguae Graecae* whenever they varied from words in the canonical gospels or the Harnack and/or Roth reconstructions of GMarc. Finally, we ran granular, segmented cross-checks of word counts within and across datasets, confirming totals of 14442 and 10572 words for the Hahn and Zahn datasets respectively.

These UTF-8 encoded .txt files offer additional starting points for CL research on GMarc and are definitely not the final word. We welcome scholarly feedback and collaboration to correct, improve, and transform our datasets, convert them to other schemata, especially TEI XML enriched with variants, tags, and notes placed in the markup to facilitate deeper analyses of the main text.

(3) DATASET DESCRIPTION

Object name: Normalized Datasets of Hahn's and Zahn's Reconstructions of Marcion's *Gospel*

Format names and versions: UTF-8 encoded .txt

Creation dates: 2020-11-01/2021-10-14

Languages: Postclassical Greek. English

License: CC-BY-NC

Repository name: Journal of Open Humanities Data Dataverse

Publication date: October 26, 2021

Dataset Creators

Mark G. Bilby (California State University, Fullerton) manually created both datasets.

(4) REUSE POTENTIAL

These transformative supplements to the first two major scholarly reconstructions of GMarc—both now in the public domain—represent the third batch of open datasets of GMarc published in the *Journal of Open Humanities Data*. GMarc has suffered considerable degradation and disintegration as a result of its suppression over the last eighteen centuries, but CL, HCL, and open data science methods can restore it to much higher levels of fidelity than currently obtain. Data scientists and humanists alike are invited to use these datasets to identify, disambiguate, and clarify the vocal strata underlying GMarc alongside the other early canonical and non-canonical gospels, whether for the purpose of uncovering the earliest textual history of the Jesus movement or to explore how editorial voices and stages in ancient texts can be scientifically delineated and sequenced.

ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Dataset.** These four UTF-8 encoded .txt files are the first born-digital, normalized, lexicographically enriched, and peer-reviewed versions of the reconstructions of Marcion's Gospel made by August Hahn in 1832 and Theodor Zahn in 1892. Two dataset files were generated for each reconstruction: the first consisting of human-readable Postclassical Greek; the second of lemmatized and morphologically tagged text following the openly licensed BibleWorks Greek Morphology schema. DOI: <https://doi.org/10.5334/johd.63.s1>

The author has no competing interests to declare.

AUTHOR AFFILIATIONS

Mark G. Bilby  orcid.org/0000-0003-0100-6634
Pollak Library, California State University, Fullerton, US

REFERENCES

- BeDuhn, J. (2013). *The First New Testament*. Salem: Polebridge.
- Bilby, M. G. (2020-07/2021-11). *The First Gospel, the Gospel of the Poor*. LODLIB. DOI: <https://doi.org/10.5281/zenodo.3927056>
- Bilby, M. G. (2021a). Key to BibleWorks Greek Morphology (v1.1). DOI: <https://doi.org/10.5281/zenodo.4950243>
- Bilby, M. G. (2021b). Normalized Datasets of Harnack's Reconstruction of Marcion's Gospel. *Journal of Open Humanities Data*, 7(24), 1–7. DOI: <https://doi.org/10.5334/johd.47>
- Bilby, M. G. (2021c). Normalized Datasets of Roth's Reconstruction of Marcion's Gospel. *Journal of Open Humanities Data*, 7(27), 1–6. DOI: <https://doi.org/10.5334/johd.57>
- Gramaglia, P. A. (2017). *Marcione e il Vangelo (di Luca)*. Turin: Accademia. ARK: <https://n2t.net/ark:/13960/t6vx6m34h>
- Hahn, A. (1823). *Das Evangelium Marcions in seiner ursprünglichen Gestalt, nebst dem vollständigsten Beweise dargestellt, da es nicht selbstständig, sondern ein verstümmeltes und verfälschtes Lukas-Evangelium war, den Freunden des Neuen Testaments und den Kritikern insbesondere, namentlich Herrn Hofrath, Ritter und Professor Dr. Eichhorn zur strengen Prüfung vorgelegt*. Königsberg: Universitäts Buchhandlung.
- Hahn, A. (1832). Evangelium Marcionis ex Auctoritate Veterum Monumentorum. In Ioannis Caroli Thilo, ed., *Codex apocryphus Novi Testamenti* (pp. 401–486). Leipzig: F. C. G. Vogel. ARK: <https://n2t.net/ark:/13960/t23b6zq5d>
- Harnack, A. von (1921/1924). *Marcion: Das Evangelium vom fremden Gott* (2nd ed.). Leipzig: J. C. Hinrichs. <https://commons.ptsem.edu/id/marciondasevange00harn>
- Klinghardt, M. (2015/2020). *Das älteste Evangelium und die Entstehung der kanonischen Evangelien*. (2nd ed., Vols. 1–2). Tübingen: Narr Francke Attempto Verlag.
- Klinghardt, M. (2021). *The Oldest Gospel* (Vols. 1–2). Leuven: Peeters.
- Nicolotti, A. (2019). Il Vangelo di Marcione. In C. Gianotto and A. Nicolotti (Eds.), *Il Vangelo di Marcione* (pp. 1–233). Turin: Einaudi.
- Roth, D. T. (2015). *The Text of Marcion's Gospel*. Leiden: Brill. DOI: <https://doi.org/10.1163/9789004282377>
- Tischendorf, C. (1869). *Novum Testamentum Graece* (Vol. 1). Leipzig: Giesecke & Devrient.
- Tsutsui, K. (1992). "Das Evangelium Marcions: Ein neuer Versuch der Textrekonstruktion." *Annual of the Japanese Biblical Institute*, 18, 67–132.
- Zahn, T. (1888/1892). *Geschichte des neutestamentlichen Kanons* (Vols. 1–2). Erlangen: Andreas Deichert. ARK: <https://n2t.net/ark:/13960/t8cf9s958>

TO CITE THIS ARTICLE:

Bilby, M. G. (2021). Normalized Datasets of Hahn's and Zahn's Reconstructions of Marcion's Gospel. *Journal of Open Humanities Data*, 7: 31, pp. 1–5. DOI: <https://doi.org/10.5334/johd.63>

Published: 29 December 2021

COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Journal of Open Humanities Data is a peer-reviewed open access journal published by Ubiquity Press.