



Normalized Datasets of Klinghardt's and Nicolotti's Reconstructions of Marcion's Gospel

DATA PAPER

MARK G. BILBY 

]u[ubiquity press

ABSTRACT

These are the first normalized, peer-reviewed, and morphologically enriched datasets to be published based on the reconstructions of Marcion's *Gospel* by Matthias Klinghardt and Andrea Nicolotti. Two 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. Among several recent critical reconstructions of Marcion's *Gospel*, those of Klinghardt and Nicolotti stand out as ambitiously maximalist and fully continuous, with considerably higher word counts than found in other recent reconstructions. These are the fourth and final batch in a series of normalized and enriched Greek 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 Klinghardt's and
Nicolotti's Reconstructions
of Marcion's Gospel. *Journal
of Open Humanities Data*, 7:
32, pp. 1–6. DOI: [https://doi.
org/10.5334/johd.70](https://doi.org/10.5334/johd.70)

(1) OVERVIEW

Repository locations: DOI: <https://doi.org/10.7910/DVN/JGL7LB> (Nicolotti's reconstruction); DOI: <https://doi.org/10.7910/DVN/BVEOEX> (Klinghardt's reconstruction)

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

Print sources:

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.

CONTEXT

The history of scholarly reconstructions of the *Gospel of Marcion* (GMarc) is outlined more thoroughly in the introductory section of Klinghardt's (2015/2020, 2021) first volume and in Gianotto's (2019) portion of the introduction to *Il Vangelo di Marcione*. More succinct overviews may be found in our previous *JOHD* data papers for the Harnack (Bilby, 2021b), Roth (Bilby, 2021c), and Hahn and Zahn (Bilby, 2021d) datasets. Over the past 200 years, eight major published reconstructions have appeared, six of which include a Greek text: Hahn (1832, Greek), Zahn (1892, Greek), von Harnack (1921/1924, Greek), Tsutsui (1992, Latin), BeDuhn (2013, English), Roth (2015, Greek), Klinghardt (2015/2020, 2021, Greek/German and Greek/English), and Nicolotti (2019, Greek/Italian).

Klinghardt's work was first released in 2015 but has recently been expanded in a second German edition (2020) and English translation (2021) that include corrections and revisions in response to some of the criticisms made in the interim between the editions. Much of the content of an entire issue of the *Zeitschrift für antikes Christentum* (ZAC) journal consisted of such criticisms along with a concluding response by Klinghardt (2017). BeDuhn (2017) deemed Klinghardt's 2015 work decisive in proving the greater antiquity of GMarc to canonical Luke. While lauding the thorough scraping of textual variants from manuscripts of Luke to fill out a more robust text for GMarc, BeDuhn nevertheless critiqued Klinghardt for not making serious use of the full range of relevant patristic sources in the reconstruction, for disregarding various non-canonical sources in the attempt to resolve the synoptic problem, for dismissing rather than modifying the Q / Two Source Hypothesis, and for dating the formation of the entire New Testament canon to the mid-second century. Bauer (2017) challenged Klinghardt's claim that GMarc was translated into Latin before the mid-2nd century, the overreaching correlation of GMarc with so-called Western text traditions, and the overreliance on variants peculiar to medieval Latin manuscripts such as Codex Palatinus. Roth (2017) focused his criticism on a meticulous examination of the sources and reconstruction of one representative verse, pointing out several problematic word choices and even self-contradictory evaluations of the patristic attestations. A quantitative approach surfaced in the review by Schmid (2017), who statistically refuted Klinghardt's position that the textual transmission of GMarc shows comparable rates of variation as found in canonical text traditions. Schmid also found the unfiltered use of idiosyncratic features in patristic attestations and late Lukan manuscripts highly problematic and indeed the entire reconstructive proposal and model unconvincing. The most involved critical response to Klinghardt's work thus far is the book-length, annotated Italian translation by Gramaglia (2017), whose detailed analysis and footnotes routinely provide counts of lemmata and syntagmata in GMarc and/or Luke, alternatively confirming or challenging Klinghardt's reconstruction on philological grounds. Gramaglia also made a counter-argument throughout that GMarc and canonical Luke are successive recensions by the same author reflective of two different passes at appropriating and editing material from the Q sayings gospel.

Nicolotti (2019) has made the most recent major attempt in print at reconstructing GMarc. Both in the goal of restoring a fully continuous text and in the frequent use of Codex Bezae to fill in the gaps between the patristic attestations to GMarc, Nicolotti's reconstruction proves similar to Klinghardt's in many respects. Even so, as our normalized datasets help to clarify, Nicolotti restored considerably fewer passages, verses, and words than did Klinghardt, and often at a lower level of certainty. Scholarly reviews of Nicolotti's reconstruction have varied widely, reinforcing two distinct sides of the scholarly debate. The French Canadian scholar Paul-Hubert Poirier found in Nicolotti's work confirmation of the Schwegler and/or Semler hypothesis that GMarc is an earlier and simpler version of the Gospel of Luke rather than a later evisceration of an earlier, longer canonical text: "Marcion a repris à son compte un écrit préexistant... qui se trouvait anticiper ce que

sera le Luc canonique” (2019: 319). Poirier also validated Nicolotti’s continuous and maximalist approach to restoration: “essentiellement pour permettre une lecture suivie et intelligible du texte marcionite, il a eu sans aucun doute tout à fait raison de décider ainsi” (318). He ultimately gave highest praise to Nicolotti’s work in comparison with other recent critical reconstructions: “la plus efficace de l’évangile marcionite... l’une des meilleures—sinon la meilleure—reconstructions de l’*instrumentum* de Marcion à avoir été publiées ces dernières années” (320). Several Italian reviews (Girolami, 2020; Mantelli, 2020; Ronzani, 2020) have consistently taken the opposite point of view, repeating the early-orthodox defense of the greater antiquity of canonical Luke and maintaining Marcion’s evisceration of canonical Luke as both historical and scholarly consensus. These Italian reviews cast doubt on the usefulness of Codex Bezae to fill in the gaps between patristic attestations to GMarc, and they favorably echo Gramaglia’s (2017) conclusion that any maximalist, continuous reconstruction of GMarc represents too hypothetical and tendentious a philological undertaking. Several inaccuracies or infelicities in Nicolotti’s Italian translation are detailed (Girolami, 2020: 568; Mantelli, 2020: 606–607). Yet these reviews also recognize that Nicolotti’s effort to reconstruct the text of GMarc makes “un ulteriore utile contributo alla discussione su un tema assai controverso” (Girolami, 2020: 568) and reflects a “lavoro minuzioso” (Ronzani, 2020: 401), especially in the production of an apparatus that is “molto puntuale ed esaustivo e rispecchia l’acribia con cui il lavoro è stato condotto” (Mantelli, 2020: 607).

Together with prior major reconstructions of GMarc, those of Klinghardt and Nicolotti yield important data, both to assess the history of scholarship and to transition to the use of data science methods to confirm, assess, and restore an ancient historical text. These reconstructions all draw on a broader array of data: hundreds of patristic attestations to GMarc, hundreds of textual variants and thousands of non-variants in manuscripts of Luke, and tens of thousands of close parallel words in other gospels both canonical and non-canonical. Because of widely divergent editorial assumptions and methods, the reconstructions vary considerably and sometimes drastically in their use of these underlying data. This is evident, for example, in a simple overview of counts of the number of verses attested and/or restored, as detailed in our *First Gospel* LODLIB (2020-07/2021-12): 823.5 in Hahn, 619.5 in Zahn, 580 in Harnack, 617 in Tsutsui, 541.5 in BeDuhn, 480.5 in Roth, 799 in Klinghardt, and 772.5 in Nicolotti. Data science, computational linguistics (CL), and historical corpus linguistics (HCL) have the potential to bring scientific objectivity and interdisciplinary collaboration to the study of GMarc, clarifying its place and significance in the history of the formation of the earliest canonical and non-canonical gospels. Transcribing, normalizing, and enriching the data from major GMarc reconstructions represents a crucial step in such research.

(2) METHOD

CHALLENGES AND RESOLUTIONS

Our published data paper for the Harnack datasets (Bilby, 2021b) sets forth the normalization standards we developed to transform Harnack’s—the most ambiguous and convoluted of all GMarc reconstructions—into a singular, clean, and consistent dataset whose words can be tokenized and thus quantitatively analyzed and compared by humans and machines. Compared to Harnack’s indications, Klinghardt’s are far more clearly defined, meticulous, and consistently implemented.

1. **bold underlined font** for **content confirmed verbatim**
2. **bold font** for **content attested but not quoted verbatim**
3. normal font for unattested content most likely in GMarc
4. {} braces for {content attested for GMarc without counterparts in canonical Luke}
5. † daggers for †content inconsistently attested for GMarc by the heresiologists†
6. (¿?) parenthetical question marks around italics for (¿content open to decision?)
7. [] square brackets with embedded regular font to indicate [variant readings]
8. [] square brackets with embedded subscript for [content unattested and likely missing]
9. [[]] double square brackets with subscript for [[content in canonical Luke missing from GMarc]]
10. ↑↓ vertical arrows for ↑content located differently in GMarc than in Luke↓

Our normalized datasets render in normal font words corresponding to indications 1–3. For indication 4, the braces are ignored as comparative rather than intrinsic in significance. For indication

5, Klinghardt typically decides on a more likely reconstruction, which we render unenclosed if Klinghardt uses bold underlined font (indication 1), yet within parentheses if Klinghardt uses bold and/or normal font (indications 2–3). In all cases, the variants signaled by indication 5 are rendered as trailing empty square brackets. For indication 6, we render empty parentheses when the indication applies to the whole verse or longer segments of words, but occasionally include the word(s) within parentheses when they are necessitated by context.¹ For indication 7, we render variants as empty square brackets. Content falling into categories 8 and 9 is omitted altogether. Indication 10 is disregarded as comparative rather than intrinsic, but the normalized datasets do preserve differences in the editorial order of verses and words. In the interest of normalized segmentation, we also take the liberty of splitting apart conflated verses, e.g., 11.53–54.

We have also found additional indications used that are not part of Klinghardt’s key or formal explanations. For example, at 17.2 italic font is used without surrounding parentheses, while at 21.32 parentheses are used around normal text. For this first iteration of the datasets, we have rendered 17.2 as normal text and kept the parentheses in 21.32. It should finally be noted that Klinghardt uses several additional indications specific to his extensive tripartite apparatus (section A for testimonials, section B for text-critical references, section C for appraisals), but these are irrelevant to the compilation of a usable dataset for the main running text.

Nicolotti’s edition makes use of a much simpler indication schema than found in the recent editions of Roth and Klinghardt, making the work easier to transform into normalized data.

1. **grassetto/bold font** for **content that is secure or very likely present** in GMarc in this form or a similar one, because quoted by some ancient author
2. tondo/normal font for certainly or probably present in GMarc in a more or less similar form, whether on the basis of allusions or quotations by ancient authors, necessary for the narrative to make sense, or possibly present in the translator’s opinion, following the text of D
3. **grassetto e corsivo / bold and italic font** for **likely choices made between contradictory options**
4. *corsivo/italic font* for *uncertain parts of the text* owing to contradictory or incomplete attestations or because modern editors have doubted their presence
5. *[italicized square brackets and font]* for *[variants within uncertain parts]*
6. ⌋ ⌋ begin and end low tone for ⌋ uncertain word order⌋

To normalize this text, we rendered content corresponding to indications 1–3 in normal font. For indications 3 and 5, we added trailing empty square brackets [] to represent the variant. Except for words or phrases necessitated by the immediate context, indication 4 content was replaced with empty parentheses (). Indication 6 was disregarded.

QUALITY AND VERSION CONTROL

Just as described in our Harnack (Bilby, 2021b), Roth (Bilby, 2021c), and Hahn and Zahn (Bilby, 2021d) data papers, 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 the tagging of the longer editions (Hahn, Zahn, Nicolotti, and Klinghardt), 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

¹ E.g., Klinghardt’s reconstruction of GMarc 5.18 reads ἐπὶ (ἰκράβαττου?) whereas the parallel portion in Luke 5.18 reads ἐπὶ κλίνης. The restored preposition requires a noun so we render it as ἐπὶ (κράβαττου).

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 a total of 12850 and 10870 words in the Klinghardt datasets and Nicolotti datasets respectively.

These UTF-8 encoded .txt files offer a starting point for CL research on GMarc, not the final word. We welcome scholarly feedback and collaboration to correct, improve, and enrich our datasets, converting them to other schemata, especially TEI XML that can allow for lower confidence readings, variants, notes, syntactical tags, and other tags to be placed in the markup while maintaining the visualized flow of the running main text. This can provide deeper user engagement, alternative analytical scenarios, more sophisticated analysis, and meaningful correlation with broader linguistic corpora.

(3) DATASET DESCRIPTION

Object name: Normalized Datasets of Klinghardt's and Nicolotti's Reconstructions of Marcion's Gospel

Format names and versions: UTF-8 encoded .txt

Creation dates: 2020-11-01/2021-10-16

Languages: Postclassical Greek, English

Rights: Permission to publish these datasets as transformational uses was generously provided by Andrea Nicolotti and Narr Francke Attempto Verlag GmbH.

License: CC-BY-NC

Repository name: Journal of Open Humanities Data Dataverse

Publication date: 2021-12-07

DATASET CREATORS

Mark G. Bilby (California State University, Fullerton) manually created both datasets based on the critical editions of Matthias Klinghardt (Technische Universität Dresden) and Andrea Nicolotti (Università degli Studi di Torino)

(4) REUSE POTENTIAL

These datasets are crucial transformative supplements to the two most recent major reconstructions of GMarc. They are by no means substitutions for the published books in which the reconstructions are found. We highly encourage readers to consult those books firsthand for the full benefit of their rich indications, apparatus, and analysis, and also to use them to evaluate our datasets critically and to suggest corrections and improvements. These datasets round out a new corpus of Postclassical Greek reconstructions of GMarc that contains 57241 tokens altogether.

As a non-canonical text suppressed for some 1800 years, GMarc has suffered much decay and disintegration, but CL, HCL, and open data science has enormous potential to restore this text to a much higher level of reliability and fidelity than currently obtains, doing so by means of scientific data restoration methods for identifying and disambiguating underlying voices while clarifying relationships with the numerous historical-vocal strata embedded in early canonical and non-canonical Gospels. These normalized datasets anticipate and resource GMarc becoming a major focus among data scientists and humanities scholars alike.

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 datasets based on the reconstructions of Marcion's Gospel made by Matthias Klinghardt and Andrea Nicolotti. Two 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.70.s1>

The author has no competing interests to declare.

AUTHOR AFFILIATION

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

REFERENCES

- Bauer, T. J.** (2017). Das Evangelium des Markion und die *Vetus Latina*. *Zeitschrift für antikes Christentum*, 21(1), 73–89. DOI: <https://doi.org/10.1515/zac-2017-0005>
- BeDuhn, J.** (2013). *The First New Testament*. Salem, OR: Polebridge.
- BeDuhn, J.** (2017). New Studies of Marcion's *Evangelion*. *Zeitschrift für antikes Christentum*, 21(1), 8–24. DOI: <https://doi.org/10.1515/zac-2017-0001>
- Bilby, M. G.** (2021a). Key to BibleWorks Greek Morphology (v1.1). *Zenodo*. 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>
- Bilby, M. G.** (2021d). Normalized Datasets of Hahn's and Zahn's Reconstructions of Marcion's Gospel. *Journal of Open Humanities Data*, 7(##), 1–5. DOI: <https://doi.org/10.5334/johd.63>
- Bilby, M. G.** (2020–07/2021–12). *The First Gospel, the Gospel of the Poor*. LODLIB. DOI: <https://doi.org/10.5281/zenodo.3927056>
- Gianotto, C.** (2019). Marcione e il suo “Nuovo Testamento.” In C. Gianotto and A. Nicolotti (Eds.), *Il Vangelo di Marcione* (pp. vii–lxviii). Turin: Einaudi.
- Girolami, M.** (2020). Review of C. Gianotto and A. Nicolotti, *Il Vangelo di Marcione*. *Studia Patavina*, 67, 565–568.
- Gramaglia, P. A.** (2017). *Marcione e il Vangelo (di Luca)*. Turin: Accademia. ARK: <https://n2t.net/ark:/13960/t6vx6m34h>
- Hahn, A.** (1832). *Evangelium Marcionis ex Auctoritate Veterum Monumentorum*. In I. C. Thilo (ed.), *Codex apocryphus Novi Testamenti* (pp. 401–486). Lipsius: F. C. G. Vogel. ARK: <https://n2t.net/ark:/13960/t23b6zq5d>
- Harnack, A. von.** (1924). *Marcion: Das Evangelium vom fremden Gott* (2nd ed.). Leipzig: J. C. Hinrichs. <https://commons.ptsem.edu/id/marciondasevange00harn>
- Klinghardt, M.** (2017). Das marcionitische *Evangelium* und die Textgeschichte des Neuen Testaments: Eine Antwort an Thomas Johann Bauer und Ulrich B. Schmid. *Zeitschrift für antikes Christentum*, 21(1), 110–120. DOI: <https://doi.org/10.1515/zac-2017-0007>
- Klinghardt, M.** (2015/2020). *Das älteste Evangelium und die Entstehung der kanonischen Evangelien* (2nd ed., Vols. 1–2). Texte und Arbeiten zum neutestamentlichen Zeitalter 60. 1–2. Tübingen: Narr Francke Attempto Verlag.
- Klinghardt, M.** (2021). *The Oldest Gospel* (Vols. 1–2). Leuven: Peeters.
- Mantelli, S.** (2020). Review of C. Gianotto and A. Nicolotti, *Il Vangelo di Marcione*. *Augustinianum*, 60(2), 601–608. DOI: <https://doi.org/10.5840/agstm202060234>
- Nicolotti, A.** (2019). *Il Vangelo di Marcione*. In C. Gianotto and A. Nicolotti (Eds.), *Il Vangelo di Marcione* (pp. 1–233). Turin: Einaudi.
- Poirier, P.-H.** (2019). Review of C. Gianotto and A. Nicolotti, *Il Vangelo di Marcione*. *Laval théologique et philosophique*, 75(2), 314–317. DOI: <https://doi.org/10.7202/1070841ar>
- Ronzani, R.** (2020). Review of C. Gianotto and A. Nicolotti, *Il Vangelo di Marcione*. *Chiesa e Storia*, 10, 394–403.
- Roth, D. T.** (2015). *The Text of Marcion's Gospel*. Leiden: Brill. DOI: <https://doi.org/10.1163/9789004282377>
- Roth, D. T.** (2017). Marcion's *Gospel* and the History of Early Christianity: The Devil is in the (Reconstructed) Details. *Zeitschrift für antikes Christentum*, 21(1), 25–40. DOI: <https://doi.org/10.1515/zac-2017-0002>
- Schmid, U.** (2017). Das marcionitische *Evangelium* und die (Text-)Überlieferung der Evangelien. *Zeitschrift für antikes Christentum*, 21(1), 90–109. DOI: <https://doi.org/10.1515/zac-2017-0006>
- 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 Klinghardt's and Nicolotti's Reconstructions of Marcion's Gospel. *Journal of Open Humanities Data*, 7: 32, pp. 1–6. DOI: <https://doi.org/10.5334/johd.70>

Published: 30 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.