



# Normalized Datasets of Roth's Reconstruction of Marcion's *Gospel*

DATA PAPER

MARK G. BILBY 

]u[ubiquity press

## ABSTRACT

These two datasets are the first normalized, peer-reviewed, and lexicographically enriched transformations of Dieter Roth's 2015 reconstruction of Marcion's *Gospel* to be published. The first dataset consists of human-readable Postclassical Greek, while the second lemmatizes and morphologically tags the text according to the openly licensed BibleWorks Greek Morphology schema. Among several recent critical reconstructions of Marcion's *Gospel*, Roth's has gained the most acceptance among scholars. These datasets represent the second batch in a series of normalized and enriched datasets of major reconstructions of Marcion's *Gospel* published by *JOHD*, Harnack's public domain work being the basis of the first.

## CORRESPONDING AUTHOR:

**Mark G. Bilby**

Pollak Library, California State  
University, Fullerton, Fullerton,  
US

[mgb8n@virginia.edu](mailto:mgb8n@virginia.edu)

---

## KEYWORDS:

Gospel of Marcion; Marcion of  
Sinope; Gospel of Luke

## TO CITE THIS ARTICLE:

Bilby, M. G. (2021).  
Normalized Datasets of  
Roth's Reconstruction of  
Marcion's *Gospel*. *Journal of  
Open Humanities Data*, 7:  
27, pp. 1–6. DOI: [https://doi.  
org/10.5334/johd.57](https://doi.org/10.5334/johd.57)

## (1) OVERVIEW

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

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

Print source: Roth, D. T. (2015). *The Text of Marcion's Gospel*. Leiden: Brill. DOI: <https://doi.org/10.1163/9789004282377>

## CONTEXT

Here we briefly recap the history of scholarly reconstructions of the *Gospel of Marcion* (GMarc) as outlined more thoroughly in our Harnack data paper (Bilby, 2021b). Eight major published reconstructions have appeared over the last 200 years: Hahn (1832), Zahn (1892), Harnack (1921, 1924), Tsutsui (1992), BeDuhn (2013), Roth (2015), Klinghardt (2015/2020, 2021), and Nicolotti (2019). Among recent reconstructions, Roth's has gained the most acceptance among scholars, as seen in reviews by Lieu (2015), Guignard (2016), Hixson (2016), Gathercole (2017), and Holmes (2017). Even highly critical reviews, such as BeDuhn's (2017), find much to praise in Roth's rigorous caution, which has lent the work a greater sense of reliability than other recent reconstructions. By comparison, the ambitious restoration by Klinghardt has been extensively critiqued in the translation by Gramaglia (2017) and in critical reviews by Bauer (2017), Schmid (2017), and Roth (2017). The recent reconstruction of Nicolotti (2019), though clearly influenced by Klinghardt's work, restores far less content. For a recent, pragmatic example of the scholarly preference for Roth's work, see Smith (2019), who uses it as the sole basis for a statistical analysis of GMarc.

The eight major scholarly reconstructions contain a plethora of interrelated yet varied data for the computational linguistic analysis of GMarc. These reconstructions are derived from much larger and more varied 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. Owing to highly divergent *a priori* assumptions and methods, the reconstructions vary widely in their respective use of these underlying data. This is evident in the total word counts of our normalized datasets: 14,442 for Hahn, 10,572 for Zahn, 4,338 for Harnack, 3,296 for Tsutsui (using a Latin to Greek conversion ratio of 1.082 based on the Vulgate having 18,003 words compared to 19,482 Greek words in canonical Luke), 7,419 for BeDuhn (using an English to Greek conversion ratio of 0.77 words based on a sampling of BeDuhn's translation habits in comparison with the New Revised Standard Version's translators), 4,169 for Roth, 12,850 for Klinghardt, and 10,870 for Nicolotti.

The methodological chasms separating these formations gauge deep disagreements on foundational matters. Should GMarc be reconstructed as a fully continuous text or is a radically discontinuous text the best scholars may hope to achieve? Can content that is not explicitly attested in patristic citations of GMarc be restored or not, and if so, on what grounds? To what degree may we lean on the variants in Codex Bezae and other so-called Western manuscript readings of Luke as a reliable basis to recover wording for GMarc? Is the target text a later, deviant evisceration of the primal, apostolic, canonized Gospel of Luke or is the canonical version a later expansion of a text anachronistically caricatured as heretical? Do these two texts reflect different authors (the Schwegler hypothesis), the same author (the early-orthodox hypothesis), or—as Gramaglia's (2017) critical translation of Klinghardt's edition posits—two recensions by the same author across two different time periods?

Despite Humanities scholars turning increasingly from theology- to text-based approaches to GMarc, many of the same debates of the last 175 years linger, with scholars running in their respective circles whilst accusing the other side of circular reasoning. As detailed more thoroughly in our Harnack data paper (Bilby, 2021a), past statistical and stylometric studies have not settled the foundational questions decisively, falling short of scientific rigor and method. Lacking careful and critical engagement with the text of GMarc, Sanday's (1876) stylometric claims made in the defense of the early-orthodox hypothesis have held sway for nearly 150 years, only passingly defended by scholars such as Rowe (2003), Wolter (2008), Hays (2008), and Roth (2015). While (the American) John Knox (1942) and his former student Tyson (2006) both roundly challenged Sanday's consensus, their conclusions—based primarily

on their use of Harnack's reconstruction—have not yet won broad acceptance, with notable exceptions in the sympathetic receptions of BeDuhn (2013), Vinzent (2014), Matthews (2018), and Klinghardt (2015/2020, 2021). More recently, Smith (2019) has compiled statistics based on verse counts from Roth's reconstruction.

Our iterative *First Gospel* LODLIB (Bilby, 2020-07/2021-10) builds on the previous statistical and stylometric studies of Knox, Tyson, and Smith by running verse and word counts and separating out Single, Double, and Triple traditions. Our work differs substantially from earlier studies by integrating computational linguistics (CL) and historical corpus linguistics (HCL) methods (signals analysis, cluster analysis, binomial distributions, data visualization, etc.) and open science principles (scientifically testable hypotheses, open data, version control, etc.) in the interest of transcending the subjective impasse in GMarc studies decisively. This has led us to compile, normalize, and enrich datasets based on all major GMarc reconstructions so that we and others may analyze and correlate these data and draw scientifically sound conclusions.

## (2) METHOD

### CHALLENGES AND RESOLUTIONS

Of all major reconstructions of GMarc, Harnack's is the most ambiguous and convoluted in its variety and inconsistent application of indications. Our data paper for the Harnack datasets (Bilby, 2021b) sets forth the normalization standards we developed to render clean, consistent data that can be scientifically compared with other GMarc datasets by humans and machines. While Roth's reconstruction is far more carefully and consistently organized than Harnack's, its variety and ambiguity of indications resemble those of Harnack in many respects. Roth details and deploys an elevenfold (!) continuum of confidence for words:

1. **bold font** for **secure**
2. **bold italics** for **very likely**
3. normal font for probable
4. *italics* for *possible*
5. () parentheses for (precise wording not attested)
6. [] square brackets for [likely present]
7. [] square brackets for [may have been present]
8. [] square brackets for [likely not present]
9. [] square brackets for [may not have been present]
10. [] square brackets for [possibly not present]
11. [] square brackets for [readings with ambiguous options]

In conjunction with the above, two more indications for words or groups of words are used.

12. ... ellipses for lacunae or unrestorable content necessary for the syntax or narrative
13. {} braces for {uncertain word order}

For verses (individually or as ranges), Roth also makes use of several more labels:

14. "attested"
15. "attested but no insight into wording can be gained"
16. "attestation uncertain"
17. "unattested" or "not attested"
18. explicitly "attested as not present"
19. implicitly (*tacitus*) "attested as not present"

To create normalized datasets from this profoundly ambiguous text, content corresponding to indications 1–3 is rendered in normal font. Parentheses enclose content corresponding

to indications 4–7. Empty parentheses substitute for indications 12, 15, and 16. To break out indication 11, we wrap the first listed reading within parentheses and substitute empty brackets in place of the variant(s). Indication 13 is simply ignored; the datasets replicate the word order as presented. The datasets omit all content corresponding to indications 8–10 and 17–19. These decisions together employ binary decision-making to produce a clear, consistent, and tokenizable running script akin to a unitary vocal performance or recording, where likely/performed content is included, but unlikely/unperformed/alternate content is not.

The editor explains these intricate typographical and tagging conventions as follows:

since perhaps the most pronounced weaknesses of all previous reconstructions, including Harnack's, is the lack of distinction between various levels of certainty for attested readings, the following reconstruction clearly reveals the attempt to indicate what level of confidence can be assigned to any particular reading for Marcion's text. Therefore, even when the wording of this reconstruction agrees with that of Harnack's, the ability to see an assessment of the relative confidence that one can place in a specific reading seeks to provide significantly more helpful insight into Marcion's Gospel. (Roth, 2015: 410)

The descriptions provided for indications 1–3 elaborate on this proposed method, according to which alignments between patristic citation habits and manuscripts of Luke merit higher levels of confidence, but disparities or lack of corroboration between them require lower levels. As BeDuhn (2017) noted, this method—though pretending to objective neutrality—inherently biases toward the canonized form of the textual tradition and confirms by way of circular reasoning the *a priori* assumption of the early orthodox hypothesis. Scholars who closely scrutinize the indications across myriad decisions and indecisions may find themselves wondering to what extent they achieve an objective standard for data restoration and to what extent they obfuscate the data by focusing on the editor's ambiguous array of skeptical sentiments. Whatever the case may be, to move from quasi-scientific confidence to actual scientific confidence, scholarship on GMarc must shift to hypothesis-driven, verifiable research and statistically significant findings based on normalized data.

## QUALITY AND VERSION CONTROL

As detailed in our Harnack data paper, the first dataset consists of human-readable Greek, and the second manually lemmatizes and morphologically tags the text using the BibleWorks Greek Morphology (BGM) schema. The BGM arose out of the previous work of the Computer Assisted Tools for Septuagint/Scriptural Studies (CATSS) project at the University of Pennsylvania under Robert Kraft (1999) to apply morphological tagging to the Septuagint. Thereafter this tagging was extended to the New Testament and other early Christian texts by Michael Bushnell of BibleWorks, Jean-Noel Aletti of the Pontifical Biblical Institute, and Andrzej Gieniusz. The BGM schema is lightweight, adaptable, familiar to many scholars, licensed for non-commercial use, and easy to compile, edit, and query in word processors and CL environments. For quality control, we ran segmented cross-checks of word counts across datasets, confirming 4,169 words total in each.

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 transform our datasets, convert them to other schemata, and create TEI XML files enriched with unlikely and variant wording and notes placed in the markup so as not to disrupt the visualized flow of the running main text while also allowing for deeper analysis and alternative scenarios.

## (3) DATASET DESCRIPTION

*Object name:* Normalized Datasets of Roth's Reconstruction of Marcion's Gospel

*Format names and versions:* UTF-8 encoded .txt

*Creation dates:* 2020-11-01/2021-09-09

## DATASET CREATORS

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

*Languages:* Postclassical Greek. English

*License:* CC-BY-NC-ND

*Repository name:* Journal of Open Humanities Data Dataverse

*Publication date:* [to be decided]

## (4) REUSE POTENTIAL

These datasets are by no means substitutions for Roth's monograph in general or his reconstruction in particular, which we highly encourage readers to consult firsthand. These datasets do, however, provide important transformative supplements to today's most widely accepted reconstruction of GMarc. They build on the previous publication of normalized datasets of Harnack's reconstruction in *JOHD* and anticipate the future publication of datasets based on other GMarc reconstructions. As a non-canonical text suppressed for some 1,800 years, GMarc has suffered much decay and disintegration, but CL, HCL, and open data science have enormous potential to restore this text to a much higher level of fidelity than currently obtains, doing so by means of scientific data restoration methods, including the identification and disambiguation of underlying voices and clarification of interdependent relationships with the voices embedded in other early canonical and non-canonical Gospels. These normalized datasets anticipate GMarc becoming a major focus among data scientists and humanities scholars alike.

## ADDITIONAL FILES

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

- **Dataset.** These two UTF-8 encoded .txt dataset files are the first normalized, peer-reviewed, and lexicographically enriched transformations of Dieter Roth's 2015 reconstruction of Marcion's Gospel to be published. The first dataset consists of human-readable Postclassical Greek, while the second lemmatizes and morphologically tags the text according to the openly licensed BibleWorks Greek Morphology schema. DOI: <https://doi.org/10.5334/johd.57.s1>
- **Key to BibleWorks Greek Morphology (BGM) (v1.1).** The BibleWorks Greek Morphology (BGM) schema is, together with its datasets, openly licensed for non-commercial distribution. The schema provides a lightweight, compact means of adding Part of Speech (PoS) tags subsequent to lemmatized words. Each element of the schema occupies a set location within a given sequence. This morphological key elaborates the schema and numbers the respective positions for the sake of clarity. Each option is represented by a single alphanumeric abbreviation dependent on its precursors and position within the sequence. DOI: <https://doi.org/10.5334/johd.57.s2>

## COMPETING INTERESTS

The author has no competing interests to declare.

## AUTHOR AFFILIATION

Mark G. Bilby  [orcid.org/0000-0003-0100-6634](https://orcid.org/0000-0003-0100-6634)

Pollak Library, California State University, Fullerton, 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: 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). 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.** (2020-07/2021-10). *The First Gospel, the Gospel of the Poor*. LODLIB. DOI: <https://doi.org/10.5281/zenodo.3927056>
- Gathercole, S.** (2017). Review of Dieter T. Roth, *The Text of Marcion's Gospel*. *Journal of Ecclesiastical History*, 68(1), 130–132. DOI: <https://doi.org/10.1017/S0022046916002025>
- Gramaglia, P. A.** (2017). *Marcione e il Vangelo (di Luca)*. Turin: Accademia. ARK: <https://n2t.net/ark:/13960/t6vx6m34h>
- Guignard, C.** (2016). Review of Dieter T. Roth, *The Text of Marcion's Gospel*. *Journal of Theological Studies*, 67(2), 775–778. DOI: <https://doi.org/10.1093/jts/flw170>
- 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** (1924). *Marcion: Das Evangelium vom fremden Gott* (2nd ed.). Leipzig: J.C. Hinrichs. <https://commons.ptsem.edu/id/marciondasevange00harn>
- Hays, C. M.** (2008). Marcion vs. Luke: A Response to the Plädoyer of Matthias Klinghardt. *Zeitschrift für die neutestamentliche Wissenschaft*, 99, 213–232. DOI: <https://doi.org/10.1515/ZNTW.2008.017>
- Hixson, E.** (2016). Review of Dieter T. Roth, *The Text of Marcion's Gospel*. *Journal of the Evangelical Theological Society*, 59(1), 204–207.
- Holmes, M. W.** (2017). Review of Dieter T. Roth, *The Text of Marcion's Gospel*. *Bulletin for Biblical Research*, 27(2), 278–279.
- Klinghardt, M.** (2015/2020). *Das älteste Evangelium und die Entstehung der kanonischen Evangelien*. (2nd ed., Vols. 1–2). Tübingen: Francke.
- Klinghardt, M.** (2021). *The Oldest Gospel* (Vols. 1–2). Leuven: Peeters.
- Knox, J.** (1942). *Marcion and the New Testament*. Chicago: University of Chicago Press.
- Kraft, R.** (1999). *Computer Assisted Tools for Septuagint/Scriptural Study*. Retrieved from <https://ccat.sas.upenn.edu/rak/catss.html> (last accessed: 28 October 2021).
- Lieu, J.** (2015). *Marcion and the Ideology of Texts*. Retrieved from <https://marginalia.lareviewofbooks.org/marcion-and-the-ideology-of-texts-by-judith-lieu> (last accessed: 28 October 2021).
- Matthews, S.** (2018). Does Dating Luke-Acts into the Second Century Affect the Q Hypothesis? In M. Müller & H. Omerzu (Eds.), *Gospel interpretation and the Q-hypothesis* (pp. 243–263). LNTS 573. London: Bloomsbury T&T Clark. DOI: <https://doi.org/10.5040/9780567670069.ch-013>
- Nicolotti, A.** (2019). *Il Vangelo di Marcione*. In C. Gianotto & 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>
- 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>
- Rowe, C. K.** (2003). The God of Israel and Jesus Christ: Luke, Marcion, and the Unity of the Canon. *Nova et Vetera*, 1, 359–380.
- Sanday, W.** (1876). *The Gospels in the Second Century*. London: Macmillan. ARK: <https://n2t.net/ark:/13960/t5p84js50>
- 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>
- Smith, D. A.** (2019). Marcion's Gospel and the Synoptics: Proposals and Problems. In J. Schröter, T. Nicklas & J. Verheyden (Eds.), *Gospels and Gospel Traditions in the Second Century: Experiments in Reception* (pp. 129–174). BZNW 235. Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9783110542349-008>
- Tsutsui, K.** (1992). Das *Evangelium Marcions*: Ein neuer Versuch der Textrekonstruktion. *Annual of the Japanese Biblical Institute*, 18, 67–132.
- Tyson, J. B.** (2006). *Marcion and Luke-Acts: A Defining Struggle*. Columbia: University of South Carolina Press.
- Vinzent, M.** (2014). *Marcion and the Dating of the Synoptic Gospels*. Leuven: Peeters.
- Wolter, M.** (2008). *Das Lukasevangelium*. Tübingen: Mohr Siebeck.
- Zahn, T.** (1888/1892). *Geschichte des neutestamentlichen Kanons* (Vols. 1–2). Erlangen: Andreas Deichert.

TO CITE THIS ARTICLE:

Bilby, M. G. (2021). Normalized Datasets of Roth's Reconstruction of Marcion's *Gospel*. *Journal of Open Humanities Data*, 7: 27, pp. 1–6. DOI: <https://doi.org/10.5334/johd.57>

Published: 12 November 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.