Normalized Datasets of Harnack's Reconstruction of Marcion's *Gospel*

MARK G. BILBY iD

]u[ubiquity press

ABSTRACT

These two datasets are the first born-digital, normalized, peer-reviewed datasets of Harnack's classic reconstruction of Marcion's *Gospel*. The first consists of human-readable postclassical Greek, the second of lemmatized and morphologically tagged text following the openly licensed BibleWorks Greek Morphology schema. The recent deluge of critical editions of Marcion's *Gospel* makes Harnack's public domain work even more relevant as scholars turn from theology- to text-based approaches to restore Marcion's *Gospel* and account for its place in the editorial history of early canonical and non-canonical Gospels. These datasets resource Marcion's *Gospel* becoming a major topic of interest in Computational Linguistics research.

CORRESPONDING AUTHOR:
Mark G. Bilby

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

mgb8n@virginia.edu

(1) OVERVIEW

Repository location: *https://doi.org/10.7910/DVN/5TEA5A*.

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

Print source: Harnack, A. von (1924). *Marcion: Das Evangelium vom fremden Gott* (2nd ed.). Leipzig: J.C. Hinrichs. (Harnack 1924). *https://commons.ptsem.edu/id/marciondasevange00harn.*

CONTEXT

As a text deemed heretical and suppressed for some eighteen centuries, Marcion's *Gospel* (GMarc) does not exist in any known manuscripts. Scholars however concur that GMarc is a version (whether earlier or later) of the canonical Gospel of Luke. Its text is attested over 700 times by more than fifteen ancient Christian writers. Thus it is not lost to history nor necessarily relegated to obscurity. Reconstructions of GMarc involve the painstaking and judicious use of a wide variety of evidence: patristic polemical quotations, paraphrases, and allusions; variants in the manuscripts of Luke; and close parallels in the canonical Gospels of Mark, Matthew, Luke, and John, as well as the non-canonical Gospels of Thomas and Peter.

As a contributor to Thilo's *Codex apocryphus Novi Testamenti* (Hahn, 1832), August Hahn became the first scholar to produce a fully continuous Greek text of GMarc based on his earlier study (Hahn, 1823) and compilation of attestations. In 1892, in his *Geschichte des neutestamentlichen Kanons* (Zahn, 1888–1892, 2.2: 455–529), Theodor Zahn produced an often discontinuous Greek text, pared down to correct and rival that of Hahn while making use of Tischendorf's *Editio octava critica maior* (1869, hereafter EOCM) of Luke as a base text. While other 19th century scholars such as Ritschl (1846) and Hilgenfeld (1850) undertook thorough analyses of GMarc and included many quotations, they did not offset a clear, running main text.

Thus Harnack's critical study and reconstruction of GMarc in 1921, followed by a second edition in 1924, became history's third major reconstruction and the scholarly standard for the better part of a century, translated into English several decades later (Harnack, 1990). Adopting Zahn's discontinuous approach to restoration and many of his indications, Harnack nevertheless pares down Zahn's text and attempts to correct many of his editorial decisions. More recently, Tsutsui's (1992) and Roth's (2015) reconstructions of GMarc both follow this discontinuous approach. The former is unique in rendering the main text in Latin instead of Greek, while the latter valuably compiles attestations, comparative citations, and critical notes in well-organized chapters and subsections. Reconstructions by Klinghardt (2015/2020; 2021) and Nicolotti in Gianotto and Nicolotti (2019) return back to Hahn's maximalist, continuous approach, while providing for the first time in history a robust, consistent critical apparatus immediately after and/or below the restored text. The English reconstruction by BeDuhn (2013) finds a middle ground between a continuous and discontinuous approach. Differing approaches notwithstanding, all recent reconstructions of GMarc make use of newer critical editions of patristic sources (e.g., Tertullian, Epiphanius, *Adamantius Dialogue*) yet still frequently cite Harnack's text and follow his decisions. As Roth (2015: 2n2) stated, "Harnack is still invaluable for Marcion studies, and some truth remains in Helmut Koester's [1982] statement, 'All further research is based on Harnack's work.'"

While recent scholarship on GMarc has shifted from theology-based to text-based approaches, stylometric and/or statistical studies on GMarc have been infrequent and lacking in scientific method and rigor. Sanday's (1876) stylometric claims about GMarc (made apart from serious engagement with its actual text) held sway for nearly 150 years. Knox (1942) and Tyson (2006) both critiqued Sanday's work and undertook their own statistical and stylometric analyses, but neither proved scientifically rigorous nor generally compelling (Cadbury, 1943; Filson, 1944; Spencer, 2008; Johnson, 2008; Roth, 2008). A recent chapter by Daniel A. Smith (2019) compiles verse count statistics based on Roth's edition. My iterative *First Gospel* LODLIB (Bilby, 2021a) challenges the unscientific *status quo* in GMarc studies and reconstructions by integrating Computational Linguistics (CL), Historical Corpus Linguistics (HCL), and Open Data Science methods more generally.

Now in the public domain, Harnack's text of GMarc is an ideal starting point for CL and HCL. Recent works may require copyright permission that is not always granted. Most original language texts in the Tufts Perseus project[1] are public domain, including its New Testament (Westcott and Hort, 1885). The PROIEL Treebank (Haug and Jøhndal, 2008),[2] a standard resource for HCL, uses Tischendorf's EOCM. Stripping out parenthetical comments, marginalia, apparatus, and footnotes from the main text, these datasets are not mere print analogues or substitutions. Their data are transformative supplements, normalized to be read, analyzed, and connected to other data by humans and machines.

While specialists may view public domain texts as outdated and largely irrelevant for current scholarship, from a statistical and stylometric point of view the differences between older and newer editions are typically minimal. GMarc is an oddity in this regard, with unusually wide variations between reconstructions, stemming from the different *a priori* assumptions and methodologies of modern editors. Even so, clear relationships exist between texts. As we show in our *First Gospel* LODLIB, in total word count and numerous other metrics, the reconstructions of GMarc by Harnack and Roth are closely correlated, while those of Hahn, Klinghardt, and Nicolotti have many notable similarities among them.

(2) METHOD
CHALLENGES AND RESOLUTIONS

While the purpose of open Humanities datasets is not confined to the replication of a print reading experience, several challenges still present themselves to anyone attempting to build normalized datasets based on Harnack's reconstruction of GMarc. The Greek text in the main body—found within the appendix whose pages are labeled with asterisks—is discontinuous and frequently interrupted by brief observations, summaries, and other kinds of descriptive indications in German, both inside and outside of parentheses. The term "unattested" / *unbezeugt* occurs frequently, indicating Lukan content not corroborated by patristic citations to GMarc. Sometimes unattested content is noted implicitly, as in the comments on GMarc 4.33–35 (184–85*), where 4.33 is skipped and "only" / *nur* some specific words in 4.34–35 are noted as attested for the passage. At other times Harnack clearly indicates words or verses as "missing" / *fehlte*, "erased" / *getilgt*, or "stricken" / *gestrichen*. Elsewhere Harnack states that "nothing is known" / *ist nichts bekannt*, as in 4.36–39 (185*). Whether absent or unattested or unknown, all such material is simply omitted from the datasets without any corresponding indication.

The term "allusion" / *Anspielung* also occurs frequently in Harnack's main text, indicating attested passages and verses deemed too unclear to restore some or all specific wording. Footnote references also appear regularly in the main text, usually but not always abbreviated "*s. u.*" (*siehe unten*), sometimes paired with references to allusions and usually in places where Harnack refrained from restoring clear wording. The corresponding footnotes are an inconsistent hodgepodge: smatterings of quotations and translations of attestations to GMarc, brief lists of some related manuscript variants in Luke, and sometimes additional secondary analysis. For example, in the running text at 7.22 (197*), Harnack notes "*s. u.*" and in the corresponding footnote on the previous page (196*) quotes a German translation of Eznik of Kolb's Armenian attestation to this verse, given without accompanying analysis or evaluation of how it might be used in a reconstruction. We render unrestored allusions and footnote references as empty parentheses (). We use this same indication when individual verse numbers appear without any subsequent content, as at 5.29 (189*), and also when individual verse numbers are skipped within passages generally attested as present, particularly when intervening content is necessitated by context and Zahn indicated the verse as implicitly present, as at 7.25 (197*).

Ellipses in the main text are used ambiguously, usually to note content implied or necessitated by the context of surrounding attested content, but sometimes to convey alignments between GMarc and Luke. When κτλ or "usw." / "etc." appears on its own or in front of ellipses, as at 18.20 (226*), it clearly indicates the subsequent alignment of GMarc and Luke. But the three ellipses across 18.10 and 14 (225*) likely indicate intervening unclear content, in keeping with

1   G.R. Crane, ed., Perseus Digital Library, *https://catalog.perseus.org*, last accessed 2021-09-20.

2   D.T.T. Haug and M.L. Jøhndal, ed., The PROEIL Treebank, *https://proiel.github.io/*, last accessed 2021-09-20.

Harnack's description of 18.9–14 as merely "allusions" / *Anspielungen*. For all ellipses we make educated judgments to resolve ambiguities, rendering empty parentheses where we think Harnack conveyed unrestorable allusions, but rendering EOCM wording [inside square brackets] where we think Harnack conveyed alignment with Luke. Occasionally Harnack uses — (em-dash) to indicate alignments with Luke, an indication Zahn had previously used. An em-dash appears, for example, between 8.8 and 8.16, likely indicating (as Zahn previously did) that GMarc 8.9–15 aligns with Luke.

Parentheses are Harnack's most ambiguous indication, with several possible meanings. 1) Clearly restored wording, often adjoined to brief summaries, as at 4.16 (185–86*), "Das Auftreten in Nazareth (ἐλθὼν δὲ εἰς Ναζαρέθ ὅπου ἦν κατὰ τὸ εἰωθὸς ἐν τ ἡμέρ τῶν σαββάτων εἰς τὴν συναγωγὴν)." 2) More likely readings than the first readings given, as at 9.34, "ἐκ τοῦ οὐρανοῦ (ἐκ τῆς νεφέλης wahrscheinlicher)." 3) Less confident readings, as at 8.25 (199*): "τίς (ἄρα)." 4) Apparent readings that follow from clearly attested wording, as at 16.17 (220*), "εὐκοπώτερον (δέ ἐστιν)." 5) Variant readings in general, as at 16.16 (220*), "ἐξ (ἀπ') οὗ ἡ βασιλεία," or introduced by "or" / *oder*, as at 9.8, "εἷς τίς τῶν ἀρχαίων προφητῶν (oder προφήτης τῶν ἀρχαίων)." 6) Highly doubtful variant readings, concluded with an internal question mark, as at 4.31 (183*): "(ἀπὸ τοῦ οὐρανοῦ?)." 7) Somewhat doubtful readings, with a standalone question mark or "uncertain" / *unsicher* note following a word, as at 10.4 (205*): "μήτε ῥάβδον (?)." This typology does not include a variety of other brief descriptions and summations of content. We resolve these ambiguities as follows: 1) by rendering clearly restored wording without parentheses; 2) by following the more likely reading, putting it in parentheses and replacing the less likely reading with empty square brackets; 3) by keeping less confident readings in parentheses; 4) by keeping implicit readings in parentheses; 5) by replacing variants with empty square brackets; 6) by rendering seriously doubted readings as empty parentheses; and 7) by wrapping a word in parentheses when a standalone question mark or "uncertain" note follows it.

Both inside and outside of parentheses and/or quotation marks in the main text, Harnack also sometimes uses German or Latin to indicate the presence of Greek words or expressions ambiguously. For example, 5.33 (189*) has "(Christi Jünger)" instead of clarifying whether this meant the Lukan phrase "but those who are yours" / οἱ δὲ σοὶ, the Markan (2.18) "but those who are your disciples" / οἱ δὲ σοὶ μαθηταὶ or the Matthean (9.14) "but your disciples" / οἱ δὲ μαθηταί σου. In this instance, following from Harnack's view that GMarc was a later, abridged version of Luke, we default to the corresponding EOCM wording of Luke, placed in square brackets. As another example, the Latin "*coetus*" inside of quotation marks is the only word given for 5.17 (189*), corresponding to Tertullian's paraphrase "amidst a throng" / *in coetu* (*Marc.* 4.10.1). Luke 5.17 describes "Pharisees and lawyers" but lacks a specific term corresponding to Tertullian's "throng", hence Harnack's vagueness. In this case we treat the term as an allusion, rendered as empty parentheses.

To complicate matters further, Harnack also used angled brackets to indicate variants, as at 5.14 (189*): "<τοῦτο>." He also used square brackets inside of parentheses to indicate variants within doubted readings, as at 4.31 (184*), "(πόλιν τῆς Γαλιλαίας [Ἰουδαίας]?)." We omit variant wording but indicate these omissions as empty brackets.

In conclusion, for all content that is not clearly restored in the main text, we use two simple indication symbols to cover four different batches of material: 1) (apparent and lower confidence restorations within parentheses); 2) empty parentheses () for footnotes, allusions, seriously doubted content, and most individual verse numbers skipped and all bare verse numbers lacking subsequent wording; 3) [alignments with EOCM Luke in square brackets]; and 4) empty square brackets [] for variants both inside and outside of parentheses. This allows for explicitly restored (batch 1) and implicitly restored [batch 3] words to be counted and analyzed in certain scenarios, while avoiding the problems inherent in any attempt to apply statistical analysis to the unclear content in batches 2 and 4. In this way we distill Harnack's noisy text of GMarc into clear, scientifically useful data. These normalized indications and data types are also applied to our other forthcoming GMarc datasets in the interest of consistency and meaningful comparison.

QUALITY AND VERSION CONTROL

To enrich the Greek text and provide for deeper CL analysis, we also manually created a second dataset with lemmatization and morphological tagging for all words. The BibleWorks Greek Morphology (BGM) schema (Bilby, 2021b) was an ideal choice for this work, given its open license for non-commercial use, its familiarity to many scholars, its lightweight schema that is easy to create and to query in word processors and advanced CL environments, and the previous collaboration of scholars to apply BGM tagging to the canonical Gospels, which often have close or exact parallels to the text of GMarc. For ambiguous options (e.g., conjunctive vs. adverbial καί), BGM also allows for multiple tags separated by a forward slash.

Practicing rapid, agile, and iterative development, we offer these UTF-8 encoded txt file datasets as a starting point for CL research on GMarc. For quality control, we ran extensive cross-checks between our Harnack and Zahn datasets, performed segmented word counts and made them openly accessible in our *First Gospel* LODLIB, confirming a total of 4338 words in both Harnack datasets. We welcome feedback from scholars and will gladly release corrected versions in the future. We also welcome collaborations to convert these datasets to schemata such as MorphGNT (Tauber, 2017) and especially enriched TEI XML with variants and notes placed in the markup so as not to interrupt the visualized flow of the main text.

(3) DATASET DESCRIPTION

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

Format names and versions: UTF-8 encoded txt

Creation dates: 2020-11-01/2021-09-10

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: 2021-09-10

(4) REUSE POTENTIAL

A recent surge of scholarly interest and new yet highly divergent reconstructions and/or translations of GMarc (BeDuhn, 2013; Roth, 2015; Klinghardt, 2015/2020; 2021; Gramaglia, 2017; Nicolotti, 2019)—most of which have hundreds of citations to Harnack's edition—makes this classic work even more relevant today, especially amidst intense scholarly debates about the place of GMarc in the Synoptic Problem and the history of the formation and transmission of the earliest Gospels. Numerous comparable open datasets exist for the canonical Gospels, but as a non-canonical text, GMarc has suffered neglect in New Testament, Classics, and Computational Linguistics studies. We release these normalized open datasets of GMarc as the first of a forthcoming batch of datasets of prior reconstructions. We hope they also spur the authors of recent editions to release normalized open datasets.

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 born-digital, normalized, peer-reviewed versions of Harnack's classic 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.47.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.47.s2*

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR AFFILIATION

Mark G. Bilby (iD) *orcid.org/0000-0003-0100-6634*

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

REFERENCES

BeDuhn, J. (2013). *The First New Testament*. Salem: Polebridge.

Bilby, M. G. (2021a). *The First Gospel, the Gospel of the Poor: A New Reconstruction of Q and Resolution of the Synoptic Problem based on Marcion's Early Luke*. LODLIB. DOI: *https://doi.org/10.5281/zenodo.3927056*

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

Cadbury, H. J. (1943). Review of J. Knox, *Marcion and the New Testament*. *Journal of Biblical Literature, 62*(2), 123–27. DOI: *https://doi.org/10.2307/3262435*

Filson, F. V. (1944). Review of J. Knox, *Marcion and the New Testament*. *The Jewish Quarterly Review, 34*(4), 385–86. DOI: *https://doi.org/10.2307/1452019*

Gianotto, C., & **Nicolotti, A.** (2019). *Il Vangelo di Marcione*. Turin: Einaudi.

Gramaglia, P. A. (2017). *Marcione e il Vangelo (di Luca)*. Turin: Accademia. *https://archive.org/details/GramagliaMarcioneEIlVangeloDiLuca*

Hahn, A. (1823). *Das Evangelium Marcions in seiner ursprünglichen Gestalt, nebst dem vollständigsten Beweise dargestellt, dass 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* (Lipsius: F.C.G. Vogel), 401–86. *https://archive.org/details/codexapocryphusn00thil*

Harnack, A. von. (1924). *Marcion: Das Evangelium vom fremden Gott* (2nd ed.). Leipzig: J.C. Hinrichs. *https://commons.ptsem.edu/id/marciondasevange00harn*

Harnack, A. von. (1990). *Marcion: The gospel of the alien god* (J. E. Steely and L. D. Bierma, trans.). Durham: Labyrinth Press. *https://archive.org/details/marciongospeloft00harn*

Haug, D. T., & **Jøhndal, M.** (2008). Creating a parallel treebank of the old Indo-European Bible translations. In *Proceedings of the second workshop on language technology for cultural heritage data (LaTeCH 2008)* (pp. 27–34).

Hilgenfeld, A. (1850). *Kritische Untersuchungen über die Evangelien Justins, der Clementinischen Homilien und Marcions: Ein Beitrag zur Geschichte der ältesten Evangelien-Literatur*. Halle: C.A. Schwetschke.

Johnson, L. K. (2008). Review of J.B. Tyson, *Marcion and Luke Acts*. *Journal of Early Christian Studies, 16*(1), 104–6. DOI: *https://doi.org/10.1353/earl.2008.0006*

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.

Perseus Digital Library. G. R. Crane (Ed.). Tufts University. *https://www.perseus.tufts.edu*

Ritschl, A. (1846). *Das Evangelium Marcions und das kanonische Evangelium des Lucas*. Tübingen: Osiander'sche Buchhandlung.

Roth, D. T. (2008). Review of J.B. Tyson, *Marcion and Luke-Acts*. *Review of Biblical Literature*.

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

Sanday, W. (1876). *The gospels in the second century*. London: Macmillan.

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* (129–74). BZNW 235. Berlin: De Gruyter. DOI: *https://doi.org/10.1515/9783110542349-008*

Spencer, F. S. (2008). Review of R.I. Pervo, *Dating Acts* and J.B. Tyson, *Marcion and Luke-Acts*. *Interpretation, 62*(2), 190–93. DOI: *https://doi.org/10.1177/002096430806200212*

Tauber, J. K. (2017). *MorphGNT: SBLGNT Edition (v6.12)*. *https://doi.org/10.5281/zenodo.376200*

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.

Tyson, J. B. (2006). *Marcion and Luke-Acts: A Defining Struggle*. Columbia: University of South Carolina Press.

Westcott, B. F., & **Hort, F. J. A.** (1885). *The New Testament in the Original Greek*. Cambridge: MacMillan and Co.

Zahn, T. (1888–1892). *Geschichte des neutestamentlichen Kanons* (Vols. 1–2). Erlangen: Andreas Deichert.